General Internal Medicine
Boston University School of Medicine
2006 Publications

1. **Alford DP**, Compton P, **Samet JH**. Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. Ann Intern Med 2006;144:127-134.
2. Amodeo M, Ellis M, **Samet JH**. Introducing evidence-based practices into substance abuse treatment using organization development methods. Am J Drug Alc Abuse. 2006;32:555-561.
3. **Battaglia TA**, **Ash A**, Prout MN, **Freund KM**. Cancer prevention trials and primary care physicians: Factors associated with recommending trial enrollment. Cancer Detection and Prevention. 2006;30:34-37
4. Bokhour BG, **Berlowitz DR**, Long JA, **Kressin NR**. How do providers assess antihypertensive medication adherence in medical encounters? J Gen Intern Med. 2006;21:577-583.
5. **Carr PL**, Edelin KC**.** Increasing the number of medical students from under-represented minorities. BMJ 2006;332:(7544):740.
6. **Chuang CH**, **Liebschutz JM**, **Horton NJ**, **Samet JH**. Association of violence victimization with inconsistent condom use in HIV-infected persons. AIDS and Behavior. 2006;10:201207.
7. Cook RL, McGinnis KA, Kraemer KL, Gordon AJ, Conigliaro J, Maisto SA, **Samet JH**, Crystal S, Rimland D, Bryant KJ, Braithwaite RS, Justice AC. Intoxication before intercourse and risky sexual behavior in male veterans with and without Human Immunodeficiency Virus infection. Med Care. 2006;44:31-36.
8. **Crosby S**, Norredam M, **Paasche-Orlow MK**, Piwowarczyk L, Heeren T, Grodin M. Prevalence of torture survivors among foreign-born patients presenting to an urban ambulatory care practice. J Gen Intern Med. 2006;21:764-768.
9. Dickert N, Kass N, **Paasche-Orlow MK**, Taylor H. Exclusion particular information from consent forms. Accountability in Res. 12:33-45.
10. Fang MC, Go AS, **Hylek EM**, Chang Y, **Henault LE**, Jensvold NG, Singer DE. Age and the risk of warfarin-associated hemorrhage: The Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. J Am Ger Soc 2006;54:1231-6.
11. Fowler FJ, Barry MJ, Walker-Corkery B, Caubet JF, Bates DW, **Lee JM**, Hauser A, McNaughton-Collins M. The impact of a suspicious prostate biopsy on patient's psychological, socio-behavioral, and medical care outcomes. J Gen Intern Med. 2006;21:715-721.
12. Frayne SM, Parker VA, Christiansen CL, Loveland S, **Seaver MR**, Kazis LE, Skinner KM. Health status among 28,000 women veterans: The VA women's health program evaluation project. J Gen Intern Med. 2006;21:Suppl3:S40-46.
13. Garcia DA, **Hylek EM**. Antithrombotic therapy in atrial fibrillation. In: Jacobs LG: Clinics in geriatric medicine: Thromboembolic disease and antithrombotic agents in the elderly. W.B. Saunders Co. 2006;22:155-66.
14. Garcia DA, Regan S, Crowther M, **Hylek EM**. The risk of hemorrhage among patients with warfarin-associated coagulopathy. J Am Coll Cardiol 2006;47:804-8.

15. **Greenwald JL**, Hall J, Skolnik PR. Approaching the CDC's guidelines on the HIV testing of inpatients: Physician-referral versus nonreferral-based testing. AIDS Patient Care & STDs. 2006;5(20):1-7.
16. **Greenwald JL**. Routine rapid HIV testing in hospitals: Another opportunity for hospitalists to improve care. Journal of Hospital Medicine 2006;1(2):106-112.
17. **Greenwald JL**, Burnstein GR, **Pincus J**, Branson B. A Rapid review of rapid HIV antibody tests. Infectious Disease Reports 2006;8:125-131.
18. **Greenwald JL**, **Rich CA**, Bessega S, Posner MA, Maeda JL, Skolnik PR**.** Evaluation of the Centers for Disease Control and prevention's recommendations regarding routine testing for human immunodeficiency virus by an inpatient service: Who are we missing? Mayo Clin Proc 2006;81(4):452-458.
19. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, **Hylek EM**, Phillips B, Raskob G, Zelman-Lewis S, Schunemann H. Grading strength of recommendations and quality of evidence in clinical guidelines. Report from an American College of Chest Physicians task force. Chest. 2006;129:174-81.
20. **Hylek EM**, **Henault LE**, Evans-Molina C, Shea C, Regan S. Translating the results of randomized trials into clinical practice: the challenge of warfarin candidacy among hospitalized elderly patients with atrial fibrillation [letter]. Stroke. 2006;37:2868.
21. **Hylek EM**, D'Antonio J, Evans-Molina C, Shea D**, Henault LE**, Regan S. Translating the results of randomized trials into clinical practice. The challenge of warfarin candidacy among hospitalized elderly patients with atrial fibrillations. Stroke. 2006;37:1075-1080.
22. Jagsi R, Phil D, Guancial EA, Cooper Worobey C, **Henault LE**, Chang Y, Starr R, Tarbell NJ, **Hylek EM**. The "Gender Gap" in authorship of academic medical literature – a 35 – year perspective**.** N Engl J Med. 2006;355:281-7.
23. Kertesz S, Larson MJ, **Cheng DM**, Tucker JA, Winter M, Mullins A, **Saitz R, Samet JH.** Need and non-need factors associated with addiction treatment utilization in a cohort of homeless and housed urban poor. Med Care. 2006;44:225-33.
24. **Kim TW**, **Alford DP**, Malabanan A, Holick MF, **Samet JH**. Low bone density in patients receiving methadone maintenance treatment. Drug Alc Depend. 2006;85:258-62.
25. **Kim TW**, Kertesz SG, **Horton NJ**, Tibbetts N, **Samet JH**. Episodic homeless and health care utilization in a prospective cohort of HIV-infected persons with alcohol problems. BMC Health Serv Res. 2006;6:1-10.
26. Kripalani S, **Paasche-Orlow MK**, Parker RM, Saha S. Advancing the field of health literacy. J Gen Intern Med. 2006;21:804-5.
27. Krupitsky EM, Zvartau EE, Lioznov DA, Tsoy MV, Egorova VY, Belyaeva TV, Antonova TV, Brazhenko NA, Zagdyn ZM, Verbitskaya EV, Zorina Y, Karandashova GF, Slavina TY, Grinenko AY, **Samet JH**, Woody GE. Co-morbidity of infectious and addictive diseases in St. Petersburg and the Leningrad Region, Russia. Eur Addict Res. 2006;12:12-19.
28. **Larson MJ**, **Saitz R**, **Horton NJ**, Lloyd-Travaglini C, **Samet JH**. Emergency department and hospital utilization among alcohol and drug-dependant detoxification patients without primary medical care. The American Journal of Drug and Alcohol Abuse. 2006;32:435-52.
29. Libman H, **Saitz R**, Nunes D, **Cheng DM**, Richardson JM, Vidaver J, Alperen JK, **Samet JH**. Hepatitis C infection is associated with more depressive symptoms in HIV-infected adults with a history of alcohol problems. Am J Gastro. 2006;101:1804-10.

30. **Liebschutz JM**, Darko GO, Finley EP, Cawse JM, Bharel M, **Orlander JD**. In the minority: black physicians in residency and their experiences. J Nat'l Med Assoc. 2006;98:1441-48.
31. Lincoln A, **Paasche-Orlow MK**, **Cheng DM**, Lloyd-Travaglini C, Caruso C, **Saitz R**, **Samet JH**. Impact of health literacy on depressive symptoms and mental health-related quality of life among adult with addiction. J Gen Intern Med. 2006;21:818-22.
32. **Migneault J**, **Farzanfar R**, **Wright J**, **Friedman RH**. How to write health dialog for a talking computer. Journal of Biomedical Informatics. 2006;39(5):468-81.
33. Nunes DP, **Saitz R**, Libman H, **Cheng DM**, Vidaver J, **Samet JH.** Barriers to treatment of Hepatitis C in HIV/HCV-coinfected adults with alcohol problems. Alc Clin Exp Res. 2006;30:1520-26.
34. **Paasche-Orlow MK**, **Cheng DM**, **Palepu A**, **Meli S**, Faber V, **Samet JH**. Health literacy, antiretroviral adherence, and HIV-RNA suppression: a longitudinal perspective. J Gen Internal Med. 2006;21:835-40.
35. **Paasche-Orlow MK**, Schillinger D, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. J Gen Intern Med. 2006;21:884-7.
36. Palepu A, **Cheng DM**, **Kim TW,** Nunes D, Vidaver J, **Alperen JK**, **Saitz R**, **Samet JH**. Substance abuse treatment and receipt of liver specialty care among persons coinfected with HIV/HCV who have alcohol problems. J Subst Abuse Treat. 2006;31:411-17.
37. **Raj A**, **Cheng DM**, Levison R, **Meli S**, **Samet JH**. Sex trade, sexual risk and non-disclosure of HIV serostatus: findings from HIV-infected persons with a history of alcohol problems**.** AIDS and Behar. 2006;10:149-57.
38. **Raj A**, Livramento KN, Santana MC, Gupta J, Silverman JG. Victims of initimate partner violence more likely to report abuse from in-laws. Violence Against Women. 2006;12:936-49.
39. **Raj A**, Santana MC, La Marche A, Amaro H, Cranston K, Silverman JG. Perpetration of initmate partner violence associated with sexual risk behaviors among young adult men. American Journal of Public Health. 2006;96:1873-78.
40. **Ramani S**. Twelve tips to promote excellence in medical teaching. Med Teacher. 2006;28:19-23.
41. **Ramani S**, Gruppen L, Kachur EK. Twelve tips to developing effective mentors. Medical Teacher. 2006;28(5):404-8.
42. **Rose AJ**, Stein MR, Arnsten JH, **Saitz R**. Teaching internal medicine resident physicians about alcoholics anonymous: a pilot study of an educational intervention. Sub Abuse. 2006;27(3):5-11.
43. **Rubin A**, **Migneault JP**, Marks L, Goldstein E, **Ludena K**, **Friedman RH**. Automated telephone screening for problem drinking. J. Stud. Alcohol. 2006; 67: 454-57.
44. **Saitz R**, Svikis D, D'Onofrio G, Kraemer KL, Perl H. Challenges applying alcohol brief intervention in diverse practice settings: populations, outcomes and costs. Alc Clin Exp Res. 2006;30:332-338.
45. **Saitz R**, **Freedner N**, Palfai TP, **Horton NJ**, **Samet JH**. The severity of unhealthy alcohol use in hospitalized medical patients: the spectrum is narrow. J Gen Intern Med. 2006; 21:381-85.
46. **Samet JH**, Galanter M, **Bridden C**, Lewis D. Association for medical education and research in substance abuse. Addiction. 2006;101:10-15.

47. Santana MC, **Raj A**, Decker MR, La Marche A, Silverman JG. Masculine gender roles associated with increased sexual risk and intimate partner violence perpetration among young adult men. Journal of Urban Health: Bulletin of the New York Academy of Medicine. 2006;83:575-85.

48. **Sherman BJ**, Savage CR, Eddy KT, Blais MA, Deckersbach T, Jackson SC, Franko DL, Rauch SL, Herzog DB. Strategic memory in adults with anorexia nervosa: are there similarities to obsessive compulsive spectrum disorders? Int J Eat Disord. 2006;6:468-76.

49. Shwartz M, **Ash AS**, Pekoz EA. Risk adjustment and risk-adjusted provider profiles. International Journal of Healthcare Technology and Management. 2006;7(1/2):15-41.

50. **Smith KL**, **Horton NJ**, **Saitz R**, **Samet JH**. Use of mini-mental state examination in recruitment for substance abuse research studies. Drug Alc Depend. 2006; 82:231-237.

51. Velicer WF, **Friedman RH**, Fava JL, Gulliver SB, Keller S, Sun X, Ramelson H, Prochaska JO. Evaluating nicotine replacement therapy and stage-based therapies in a population-based effectiveness trial. Journal of Consulting and Clinical Psychology. 2006;74:1162-72.

52. Williams EC, Kivlahan DR, **Saitz R**, Merrill JO, Achtmeyer CE, McCormick KA, Bradley KA. Readiness to change in primary care patients who screen positive for alcohol misuse. Ann Fam Med. 2006;4:213-20.

53. Yano EM, Bastian LA, Frayne SM, Howell AL, Lipson RL, McGlynn G, Schnurr PP, **Seaver MR**, Spungen Am, Fihn SD. Toward a VA women's health research agenda: setting evidence-based priorities to improve the health and health care of women veterans. J Gen intern Med. 2006;(3):S93-101.

Annals of Internal Medicine | PERSPECTIVE

Acute Pain Management for Patients Receiving Maintenance Methadone or Buprenorphine Therapy

Daniel P. Alford, MD, MPH; Peggy Compton, RN, PhD; and Jeffrey H. Samet, MD, MA, MPH

More patients with opioid addiction are receiving opioid agonist therapy (OAT) with methadone and buprenorphine. As a result, physicians will more frequently encounter patients receiving OAT who develop acutely painful conditions, requiring effective treatment strategies. Undertreatment of acute pain is suboptimal medical treatment, and patients receiving long-term OAT are at particular risk. This paper acknowledges the complex interplay among addictive disease, OAT, and acute pain management and describes 4 common misconceptions resulting in suboptimal treatment of acute pain. Clinical recommendations for providing analgesia for patients with acute pain who are receiving OAT are presented. Although challenging, acute pain in patients receiving this type of therapy can effectively be managed.

Ann Intern Med. 2006;144:127-134. www.annals.org
For author affiliations, see end of text.

The treatment of opioid dependence, both on heroin and prescription narcotics, with opioid agonist therapy (OAT) (that is, methadone or buprenorphine) is effective: It decreases opioid and other drug abuse, increases treatment retention, decreases criminal activity, improves individual functioning, and decreases HIV seroconversion (1–5). Because of the increasing use of these medications for prolonged periods in primary care, a practice called *office-based opioid treatment*, nonaddiction specialists will be treating more of these affected patients in clinical practice, including those with episodes of acute pain (6–11).

Adequate treatment of acute painful conditions is an essential dimension of quality medical care (12–17). Inadequate treatment is common among a wide spectrum of patients (18–23). Nonopioid analgesics (for example, nonsteroidal anti-inflammatory drugs and acetaminophen) are recommended for treating acute pain; however, moderate to severe acute pain will often require opioid analgesics (24). Physicians may not prescribe effective opioid analgesia across all patient populations because of fears of cognitive, respiratory, and psychomotor side effects; iatrogenic drug addiction; and prescription drug diversion (25, 26). This tendency of health care professionals to undermedicate patients with opioid analgesics has been termed *opiophobia* (27). Such fears are exaggerated when treating patients with a known history of a substance use disorder. The provision of opioid analgesics to a patient with opioid dependence receiving OAT can be particularly challenging (28, 29).

We highlight the issues associated with the management of acute pain in patients receiving OAT and describe theoretical and empirical findings that suggest unique requirements for opioid analgesia for such patients. In addition, we identify common misconceptions of health care providers that underlie inadequate pain management and provide practical recommendations for the analgesic management of acute pain in this special clinical population. To help illustrate these issues, we present the following clinical vignette from our experience.

A 29-year-old woman reported severe right arm pain after fracturing her olecranon process. She had a history of injection heroin use and received methadone, 90 mg/d, in a methadone maintenance program. In the emergency department, she seemed uncomfortable and received one 2-mg dose of intramuscular morphine sulfate over 6 hours. While hospitalized, she continued to report severe pain despite receiving her daily methadone dose and intramuscular ketorolac. She was told that her usual methadone dose should help control her pain. She was labeled as "drug-seeking" because of her constant requests for additional pain medications.

Pain and Opioid Dependence

The clinical conditions of pain and opioid dependence are not unrelated phenomena (30–32). Forty-one years ago, Martin and Inglis (33) observed that opioid-addicted patients abuse opioids to treat "an abnormally low tolerance for painful stimuli." Opioids, whether administered with analgesic or addictive intent, activate opiate receptors in the locus coeruleus and amygdala, which provide both analgesia and reward (34, 35).

The presence of one condition seems to influence the expression of the other. Clinical examples of this include how the presence of acute pain seems to decrease the euphorogenic (pleasurable) qualities of the opioid (36) and how the presence of addictive disease seems to worsen the experience of pain. With respect to the latter, Savage and Schofferman (37) found a decade ago that persons with addiction and pain have a "syndrome of pain facilitation." Their pain experience is worsened by subtle withdrawal syndromes, intoxication, withdrawal-related sympathetic nervous system arousal, sleep disturbances, and affective

See also:

Web-Only
CME quiz
Conversion of tables into slides

Table 1. **Pain and Addiction Terminology***

Term	Definition
Physical dependence	Normal physiologic adaptation defined as the development of withdrawal or abstinence syndrome with abrupt dose reduction or administration of an antagonist
Tolerance	Normal neurobiological event characterized by the need to increase the dose over time to obtain the original effect
Cross-tolerance	Normal neurobiological event of tolerance to effects of medication within the same class
Substance (opioid) dependence (addiction)	Chronic neurobiological disorder defined as a pattern of maladaptive behaviors, including loss of control over use, craving and preoccupation with nontherapeutic use, and continued use despite harm resulting from use with or without physical dependence or tolerance
Pseudoaddiction	Behavioral changes in patients that seem similar to those in patients with opioid dependence or addiction but are secondary to inadequate pain control
Drug-seeking behaviors	Directed or concerted efforts on the part of the patient to obtain opioid medication or to ensure an adequate medication supply; may be an appropriate response to inadequately treated pain
Therapeutic dependence	Patients with adequate pain relief may demonstrate drug-seeking behaviors because they fear not only the reemergence of pain but perhaps also the emergence of withdrawal symptoms
Pseudo-opioid resistance	Adequate pain relief continue to report persistent severe pain to prevent reduction in current opioid analgesic dose
Opioid-induced hyperalgesia	A neuroplastic change in pain perception resulting in an increase in pain sensitivity to painful stimuli, thereby decreasing the analgesic effects of opioids

* Adapted from references 40–44.

changes, all consequences of addictive disease (37). Supporting a negative effect of addiction on pain tolerance, patients who abuse stimulants and those who abuse opioids have been shown to be less tolerant of pain than their peers in remission (38, 39).

The clinical approach is complicated by the confusing and often misunderstood terminology used in pain management and addiction medicine (40–43). As detailed in **Table 1**, physical dependence and tolerance are typical and predictable physiologic consequences of opioid exposure. These terms in and of themselves do not indicate maladaptive behaviors and do not meet the diagnostic criteria of substance dependence (41) outlined in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision, without, for example, loss of control or continued use despite harm. *Drug-seeking* is another commonly ill-defined term that may indicate the presence of addiction but, as will be described, can also reflect *pain relief–seeking* because of unrelieved pain or anxiety about pain management (44).

Common Misconceptions

Four common misconceptions of health providers result in the undertreatment of acute pain in patients receiving OAT: 1) The maintenance opioid agonist (methadone or buprenorphine) provides analgesia; 2) use of opioids for analgesia may result in addiction relapse; 3) the additive effects of opioid analgesics and OAT may cause respiratory and central nervous system (CNS) depression; and 4) the pain complaint may be a manipulation to obtain opioid medications, or drug-seeking, because of opioid addiction.

Misconception 1: The Maintenance Opioid Agonist (Methadone or Buprenorphine) Provides Analgesia

There are pharmacokinetic and pharmacodynamic explanations for why patients do not receive adequate analgesia from maintenance opioids prescribed for addiction treatment. Not only do the analgesic and addiction treatment profiles of these opioids differ, but the neuroplastic changes associated with long-term opioid exposure (that is, tolerance and hyperalgesia) may effectively diminish their analgesic effectiveness (45).

Analgesic Properties of Maintenance Opioids

Patients receiving maintenance therapy with opioids for addiction treatment do not derive sustained analgesia from it. Methadone and buprenorphine, potent analgesics, have a duration of action for analgesia (4 to 8 hours) that is substantially shorter than their suppression of opioid withdrawal (24 to 48 hours) (46–50). Because most patients receiving OAT are given a dose every 24 to 48 hours, the period of even partial pain relief with these medications is small.

Opioid Tolerance

Tolerance is one factor that explains why these patients derive little pain relief from maintenance opioids. Tolerance, the need for increasing doses of a medication to achieve its initial effects, develops with continuous opioid use but differentially affects specific opioid properties. For example, tolerance readily develops to the respiratory and CNS depressive effects of opioids but not to their constipating effects (51, 52). Analgesic tolerance develops for different medications within the opioid class, a phenomenon called *cross-tolerance* (53, 54). Doverty and colleagues (55) found that patients receiving maintenance methadone therapy were cross-tolerant to the analgesic effects of morphine and that pain relief, when obtained, did not last as long as expected. Therefore, cross-tolerance between the opioids used for maintenance therapy and other opioids used for analgesia may explain why patients receiving OAT often require higher and more frequent doses of opioid analgesics to achieve adequate pain control.

Opioid-Induced Hyperalgesia

An alternative explanation for the lack of analgesia derived from maintenance opioids may be the presence of opioid-induced hyperalgesia. This is the result of neuroplastic changes in pain perception that yield an increase in pain sensitivity. The outcome is that opioids have less potent analgesic effects (45, 56–58). Empirical evidence supports increased sensitivity to experimental pain in patients receiving OAT (33, 38, 55, 59–62), such that patients receiving maintenance methadone therapy tolerate cold-pressor pain only half as long as do matched controls (55, 59). Accumulating evidence suggests that maintenance with buprenorphine therapy has similar and statistically significant effects on pain tolerance, although to a lesser degree than methadone (63). The pain intolerance of patients receiving methadone and buprenorphine maintenance therapy can be conceptualized as a latent hyperalgesia secondary to long-term opioid exposure.

The presence of hyperalgesia with ongoing opioid use has resulted in reexamination of the previously described phenomenon of opioid analgesic tolerance. Both hyperalgesia and opioid tolerance involve neuroplastic changes associated with excitatory amino acid (*N*-methyl-D-aspartate) and opioid receptors (64–70). The hyperalgesic processes precipitated by opioid administration serve to counteract opioid analgesia (56, 71–73); thus, it is possible that what seems to be opioid analgesic tolerance may in fact be an expression of an opioid-induced increased sensitivity to pain.

Therefore, despite the benefits of OAT for the opioid-dependent person, the accompanying hyperalgesia (or analgesic tolerance) counteracts the analgesic effects of opioids and complicates pain management. At clinically effective doses for the treatment of opioid dependence, patients do not experience analgesia to experimental pain but demonstrate the hyperalgesic effects of OAT. Thus, from a theoretical and experimental basis, it is clear that the perception of pain is not decreased in OAT patients.

Misconception 2: Use of Opioids for Analgesia May Result in Addiction Relapse

A common concern of physicians is that the use of opioids for analgesia in patients receiving OAT May result in relapse to active drug use. However, there is no evidence that exposure to opioid analgesics in the presence of acute pain increases rates of relapse in such patients. A small retrospective study (74) of patients enrolled in maintenance methadone programs who received opioid analgesics after surgery did not find a difference in relapse indicators compared with matched patients receiving maintenance methadone therapy. Similarly, no evidence of relapse was seen in 6 patients receiving methadone maintenance therapy who were treated with opioid analgesics for cancer-related pain (75). In fact, relapse prevention theories would suggest that the stress associated with unrelieved pain is more likely to be a trigger for relapse than adequate analgesia. In a study by Karasz and colleagues (76), patients receiving methadone maintenance therapy stated that pain played a substantial role in their initiating and continuing drug use.

Misconception 3: The Additive Effects of Opioid Analgesics and OAT May Cause Respiratory and CNS Depression

Physicians' concerns that opioid analgesics will cause severe respiratory or CNS depression in patients receiving OAT is a theoretical risk, which has never been clinically demonstrated. As previously noted, tolerance to the respiratory and CNS depressant effects of opioids occurs rapidly and reliably (50–52). Similarly, patients with worsening cancer-related pain who require dose escalations typically do not exhibit respiratory and CNS depressant effects when additional opioids are administered (75, 77–79). It has been suggested that acute pain serves as a natural antagonist to opioid-associated respiratory and CNS depression (15, 43). This purported effect is supported by the observation that a patient with chronic pain who was treated with opioids developed signs of respiratory depression after a successful nerve block procedure (80). Therefore, the concern about severe drug toxicity with analgesic opioid treatment is not supported by clinical or empirical experience.

Misconception 4: Reporting Pain May Be a Manipulation To Obtain Opioid Medications, or Drug-Seeking, because of Opioid Addiction

Physicians' concerns about being manipulated by drug-seeking patients is substantial, difficult to quantify, and emotion-laden. It is a powerful motive underlying physicians' reservations about prescribing opioid analgesia for acute pain to patients receiving OAT for opioid dependence. Pain is always subjective, making assessment of its presence and severity difficult. A careful clinical assessment for objective evidence of pain will decrease the chance of being manipulated by a drug-seeking patient and will support the use of opioid analgesics in patients with a history of opioid dependence. Reports of acute pain with objective findings are less likely to be manipulative gestures than are reports of chronic pain with vague presentations. Furthermore, patients receiving OAT typically receive treatment doses that block most euphoric effects of coadministered opioids, theoretically decreasing the likelihood of opioid analgesic abuse (81, 82).

Not uncommonly, patients dependent on opioids are perceived by health care providers to be demanding when hospitalized with acute pain. This scenario develops in part because of the patients' distrust of the medical community, concern about being stigmatized, and fears that their pain will be undertreated or that their OAT may be altered or discontinued (76, 83). Patient anxiety related to these concerns, which can be profound and well-founded, can complicate provision of adequate pain relief.

Requests for opioid analgesia from patients receiving

OAT may be labeled as drug-seeking behaviors, which are defined as concerted efforts on the part of the patient to obtain opioid medication, including engaging in illegal activities (44). It is important to keep in mind that there may be appropriate reasons for a patient to seek medication. The distinction between appropriate drug-seeking and addiction is harder to discern when the patient requests a drug with known abuse potential, such as opioid analgesics, regardless of the apparent validity of the complaint. In the case of unrelieved pain, drug-seeking behaviors arise when a patient cannot obtain tolerable relief with the prescribed dose of analgesic and seeks alternate sources or increased doses, a phenomenon referred to as *pseudoaddiction* (84). Alternately, patients receiving good pain relief may exhibit drug-seeking behaviors because they fear not only the reemergence of pain but perhaps also the emergence of withdrawal symptoms. Rather than indicating addictive disease, such behaviors, termed *therapeutic dependence* (85), are actually efforts to maintain a tolerable level of comfort. Other patients with adequate pain control may continue to report persistent severe pain to prevent reduction in current effective doses of opioid analgesics, a behavior termed *pseudo-opioid resistance* (86).

Table 2. **Recommendations for Treating Acute Pain in Patients Receiving Opioid Agonist Therapy***

Addiction treatment issues
Reassure patient that addiction history will not prevent adequate pain management.
Continue the usual dose (or equivalent) of OAT.
Methadone or buprenorphine maintenance doses should be verified by the patient's methadone maintenance clinic or prescribing physician.
Notify the addiction treatment program or prescribing physician regarding the patient's admission and discharge from the hospital and confirm the time and amount of last maintenance opioid dose.
Inform the addiction treatment maintenance program or prescribing physician of any medications, such as opioids and benzodiazepines, given to the patient during hospitalization because they may show up on routine urine drug screening.
Pain management issues
Relieve patient anxiety by discussing the plan for pain management in a nonjudgmental manner.
Use conventional analgesics, including opioids, to aggressively treat the painful condition.
Opioid cross-tolerance and patient's increased pain sensitivity will often necessitate higher opioid analgesic doses administered at shorter intervals.
Write continuous scheduled dosing orders rather than as-needed orders.
Avoid using mixed agonist and antagonist opioids because they may precipitate an acute withdrawal syndrome.
If the patient is receiving methadone maintenance therapy and requires opioid analgesics
Continue methadone maintenance dose.
Use short-acting opioid analgesics.
If the patient is receiving buprenorphine maintenance therapy and requires opioid analgesics, 4 options are available and should be chosen on the basis of the anticipated duration of pain, treatment setting, and response to the chosen option
Continue buprenorphine maintenance therapy and titrate short-acting opioid analgesics (for pain of short duration only).
Divide buprenorphine dose to every 6–8 hours.
Discontinue buprenorphine maintenance therapy and use opioid analgesics. Convert back to buprenorphine therapy when acute pain no longer requires opioid analgesics.
If the patient is hospitalized, discontinue buprenorphine therapy, treat opioid dependence with methadone at 20–40 mg, and use short-acting opioid analgesics to treat pain. Have naloxone available at the bedside. Discontinue methadone therapy and convert back to buprenorphine therapy before hospital discharge (for inpatients only).

* These recommendations are applicable only for patients receiving OAT who require opioid analgesics. OAT = opioid agonist therapy.

RECOMMENDATIONS FOR TREATING ACUTE PAIN

General Recommendations

The appropriate treatment of acute pain in patients receiving OAT includes uninterrupted therapy to address the patient's baseline opioid requirement for their addiction treatment and aggressive pain management (**Table 2**). As with all patients who have acute pain, nonpharmacologic and nonopioid analgesic pain-relieving interventions should be aggressively implemented. However, patients with moderate to severe acute pain will often require opioid analgesics (24). The literature suggests that undertreating acute pain may lead to decreased responsiveness to opioid analgesics, thus making subsequent pain control more difficult (54, 87). To decrease the total amount of opioid provided to these patients, multimodal analgesia (for example, nonsteroidal anti-inflammatory drugs and acetaminophen) (88) and adjuvant analgesics that enhance opioid effects (for example, tricyclic antidepressants) (89) may be coadministered (90). Continuing the usual dose of OAT, after the important step of verification with the patient's provider or program, avoids worsening pain symptoms due to the increased pain sensitivity associated with opioid withdrawal (77, 91, 92). Thus, daily opioid treatment requirements must be met before attempting to achieve analgesia. To decrease anxiety, patients should be reassured that the treatment for their opioid addiction will continue and that their pain will be aggressively treated. When the increased pain sensitivity and cross-tolerance with OAT are considered, adequate pain control will generally necessitate higher doses of opioid analgesic administered at shorter intervals. Analgesic dosing should be continuous or scheduled, rather than as needed. Allowing pain to reemerge before administering the next dose causes unnecessary suffering and anxiety and increases tension between the patient and the treatment team.

Empirical data on the use of patient-controlled analgesia in patients with substance dependence are limited. Paige and colleagues (93) reported that although women receiving maintenance therapy with methadone had higher pain scores after cesarean section surgery, there was no statistically significant difference in use of opioid analgesics compared with controls. Boyle (94) reported on the successful use of postoperative patient-controlled analgesia in a patient who actively used heroin. Clinical experience supports consideration of patient-controlled analgesia in patients receiving OAT; increased control over analgesia minimizes patient anxiety about pain management.

The pharmacologic properties of opioids must be con-

sidered when selecting an opioid analgesic for the patient receiving OAT. Although opioids bind to multiple subtypes of opioid receptors in the CNS, binding to the μ receptor subtype is primarily responsible for the analgesic effect (95). Mixed agonist and antagonist opioid analgesics, such as pentazocine (Talwin, Sanofi-Synthelabo Inc., New York, New York), nalbuphine (Nubain, Bristol-Myers Squibb Holdings Pharma, Ltd., Manati, Puerto Rico), and butorphanol (Stadol, Bristol-Myers Squibb Co., Princeton, New Jersey), must be avoided because they probably will displace the maintenance opioid from the μ receptor, thus precipitating acute opioid withdrawal in these patients (28). Combination products of opioid analgesics containing fixed doses of acetaminophen and an opioid (for example, Percocet, Wilmington Laboratories, L.L.C., Wilmington, Delaware, and Vicodin, Knoll Pharmaceuticals, Mount Olive, New Jersey) should be limited to patients not requiring large doses to avoid acetaminophen-induced hepatic toxicity. Alternatively, each medication could be prescribed individually at appropriate doses to achieve the desired analgesic effect and to avoid hepatic damage.

Recommendations for Patients Receiving Maintenance Methadone Therapy

Acute pain management for the patient receiving maintenance methadone therapy should follow the aforementioned general recommendations, which include using opioid analgesics, when indicated, in addition to the patient's daily methadone maintenance dose (**Table 2**). If the patient is hospitalized, in addition to dose verification, the methadone maintenance program should be notified at the time of hospital admission and discharge, in part to make program clinical staff aware of any controlled substances that were given to the patient and may be detectable by drug testing. If the patient is not receiving oral intake, the methadone dose can be given parenterally. Intramuscular or subcutaneous methadone dosing should be given as half to two thirds the maintenance dose divided into 2 to 4 equal doses (48, 96, 97).

Recommendations for Patients Receiving Maintenance Buprenorphine Therapy

Clinical experience treating acute pain in patients receiving maintenance therapy with buprenorphine is limited. Pain treatment with opioids is complicated by the high affinity of buprenorphine for the μ receptor. This high affinity risks displacement of, or competition with, full opioid agonist analgesics when buprenorphine is administered concurrently or sequentially. There are several possible approaches for treating acute pain that requires opioid analgesia in the patient receiving buprenorphine therapy (**Table 2**). With such limited clinical experience, the following treatment approaches are based on available literature, pharmacologic principles, and published recommendations. The most effective approach will be elucidated with increased clinical experience. In all cases, because of highly variable rates of buprenorphine dissociation from the μ receptor, naloxone should be available and level of consciousness and respiration should be frequently monitored. Treatment options are as follows.

1. Continue buprenorphine maintenance therapy and titrate a short-acting opioid analgesic to effect (90, 98). Because higher doses of full opioid agonist analgesics may be required to compete with buprenorphine at the μ receptor, caution should be taken if the patient's buprenorphine therapy is abruptly discontinued. Increased sensitivity to the full agonist with respect to sedation and respiratory depression could occur.

2. Divide the daily dose of buprenorphine and administer it every 6 to 8 hours to take advantage of its analgesic properties. For example, for buprenorphine at 32 mg daily, the split dose would be 8 mg every 6 hours. The available literature suggests that acute pain can be effectively managed with as little as 0.4 mg of buprenorphine given sublingually every 8 hours in patients who are opioid naive (47, 99, 100). However, these low doses may not provide effective analgesia in patients with opioid tolerance who are receiving OAT. Therefore, in addition to divided dosing of buprenorphine, effective analgesia may require the use of additional opioid agonist analgesics (for example, morphine).

3. Discontinue buprenorphine therapy and treat the patient with full scheduled opioid agonist analgesics by titrating to effect to avoid withdrawal and then to achieve analgesia (for example, sustained-release and immediate-release morphine) (90, 98, 101). With resolution of the acute pain, discontinue the full opioid agonist analgesic and resume maintenance therapy with buprenorphine, using an induction protocol (98, 102).

4. If the patient is hospitalized with acute pain, his or her baseline opioid requirement can be managed and opioid withdrawal can be prevented by converting buprenorphine to methadone at 30 to 40 mg/d. At this dose, methadone will prevent acute withdrawal in most patients (97) and, unlike buprenorphine, binds less tightly to the μ receptor. Thus, responses to additional opioid agonist analgesics will be as expected (that is, increasing dose will provide increasing analgesia). If opioid withdrawal persists, subsequent daily methadone doses can be increased in 5- to 10-mg increments (103). This method allows titration of the opioid analgesic for pain control in the absence of opioid withdrawal. When the acute pain resolves, discontinue the therapy with the full opioid agonist analgesic and methadone and resume maintenance therapy with buprenorphine, using an induction protocol (98, 102). If the patient is discharged while full opioid agonist analgesics are still required, then discontinue methadone therapy and treat the patient as stated in the third buprenorphine approach.

If buprenorphine therapy needs to be restarted (buprenorphine induction) after acute pain management (that is, the third and fourth approaches), it is important to keep in mind that buprenorphine can precipitate opioid with-

drawal. Thus, a patient receiving a full opioid agonist regularly should be in mild opioid withdrawal before restarting buprenorphine therapy (98, 102).

CONCLUSION

Addiction elicits neurophysiologic, behavioral, and social responses that worsen the pain experience and complicate provision of adequate analgesia. These complexities are heightened for patients with opioid dependency who are receiving OAT, for whom the neural responses of tolerance or hyperalgesia may alter the pain experience. As a consequence, opioid analgesics are less effective; higher doses administered at shortened intervals are required. Opioid agonist therapy provides little, if any, analgesia for acute pain. Fears that opioid analgesia will cause addiction relapse or respiratory and CNS depression are unfounded. Furthermore, clinicians should not allow concerns about being manipulated to cloud good clinical assessment or judgment about the patient's need for pain medications. Reassurance regarding uninterrupted OAT and aggressive pain management will mitigate anxiety and facilitate successful treatment of pain in patients receiving OAT.

From Boston University Medical Center, Boston, Massachusetts, and University of California, Los Angeles, School of Nursing, Los Angeles, California.

Acknowledgments: The authors thank Jessica Richardson for editorial assistance.

Grant Support: Drs. Alford and Samet were supported by a grant from the National Institute on Drug Abuse (R25-DA-13582).

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Daniel P. Alford, MD, MPH, Clinical Addiction Research and Education (CARE) Unit, Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118.

Current author addresses are available at www.annals.org.

References

1. **Institute of Medicine.** Federal Regulation of Methadone Treatment. Washington, DC: National Academy of Medicine; 1995:21-6.
2. **Ball JC, Ross A.** The Effectiveness of Methadone Maintenance Treatment. New York: Springer-Verlag; 1991.
3. **Metzger DS, Woody GE, McLellan AT, O'Brien CP, Druley P, Navaline H, et al.** Human immunodeficiency virus seroconversion among intravenous drug users in- and out-of-treatment: an 18-month prospective follow-up. J Acquir Immune Defic Syndr. 1993;6:1049-56. [PMID: 8340896]
4. Effective medical treatment of opiate addiction. National Consensus Development Panel on Effective Medical Treatment of Opiate Addiction. JAMA. 1998; 280:1936-43. [PMID: 9851480]
5. **O'Connor PG, Fiellin DA.** Pharmacologic treatment of heroin-dependent patients. Ann Intern Med. 2000;133:40-54. [PMID: 10877739]
6. **Cooper JR.** Including narcotic addiction treatment in an office-based practice. JAMA. 1995;273:1619-20. [PMID: 7745777]
7. **Fiellin DA, O'Connor PG.** New federal initiatives to enhance the medical treatment of opioid dependence. Ann Intern Med. 2002;137:688-92. [PMID: 12379070]
8. **Krantz MJ, Mehler PS.** Treating opioid dependence. Growing implications for primary care. Arch Intern Med. 2004;164:277-88. [PMID: 14769623]
9. **Clark HW.** Office-based practice and opioid-use disorders. N Engl J Med. 2003;349:928-30. [PMID: 12954740]
10. **Fiellin DA, O'Connor PG.** Clinical practice. Office-based treatment of opioid-dependent patients. N Engl J Med. 2002;347:817-23. [PMID: 12226153]
11. **Merrill JO.** Policy progress for physician treatment of opiate addiction. J Gen Intern Med. 2002;17:361-8. [PMID: 12047733]
12. Quality improvement guidelines for the treatment of acute pain and cancer pain. American Pain Society Quality of Care Committee. JAMA. 1995;274: 1874-80. [PMID: 7500539]
13. **Federation of State Medical Boards of the United States, Inc.** Model Guidelines for the Use of Controlled Substances for the Treatment of Pain. Euless, TX: Federation of State Medical Boards of the United States; 2003.
14. **Gordon DB, Dahl JL, Miaskowski C, McCarberg B, Todd KH, Paice JA, et al.** American Pain Society recommendations for improving the quality of acute and cancer pain management: American Pain Society Quality of Care Task Force. Arch Intern Med. 2005;165:1574-80. [PMID: 16043674]
15. The use of opioids for the treatment of chronic pain. A consensus statement from the American Academy of Pain Medicine and the American Pain Society. Clin J Pain. 1997;13:6-8. [PMID: 9084947]
16. **Joint Commission on Accreditation of Healthcare Organizations.** Comprehensive Accreditation Manual for Hospitals: The Official Handbook. Oakbrook Terrace, IL; Joint Commission on Accreditation of Healthcare Organizations; 1999.
17. **Joint Commission on Accreditation of Healthcare Organizations.** Joint Commission Focuses on Pain Management. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 1999.
18. **Marks RM, Sachar EJ.** Undertreatment of medical inpatients with narcotic analgesics. Ann Intern Med. 1973;78:173-81. [PMID: 4683747]
19. **Rupp T, Delaney KA.** Inadequate analgesia in emergency medicine. Ann Emerg Med. 2004;43:494-503. [PMID: 15039693]
20. **Cleeland CS, Gonin R, Hatfield AK, Edmonson JH, Blum RH, Stewart JA, et al.** Pain and its treatment in outpatients with metastatic cancer. N Engl J Med. 1994;330:592-6. [PMID: 7508092]
21. **Breitbart W, Rosenfeld BD, Passik SD, McDonald MV, Thaler H, Portenoy RK.** The undertreatment of pain in ambulatory AIDS patients. Pain. 1996; 65:243-9. [PMID: 8826513]
22. **Hill CS Jr.** When will adequate pain treatment be the norm? [Editorial] JAMA. 1995;274:1881-2. [PMID: 7500540]
23. **Melzack R.** The tragedy of needless pain. Sci Am. 1990;262:27-33. [PMID: 2296714]
24. **Carr DB, Goudas LC.** Acute pain. Lancet. 1999;353:2051-8. [PMID: 10376632]
25. **Savage SR.** Opioid use in the management of chronic pain. Med Clin North Am. 1999;83:761-86. [PMID: 10386124]
26. **Lander J.** Fallacies and phobias about addiction and pain. Br J Addict. 1990; 85:803-9. [PMID: 1974156]
27. **Morgan JP.** American opiophobia: customary underutilization of opioid analgesics. Adv Alcohol Subst Abuse. 1985;5:163-73. [PMID: 2870626]
28. **Scimeca MM, Savage SR, Portenoy R, Lowinson J.** Treatment of pain in methadone-maintained patients. Mt Sinai J Med. 2000;67:412-22. [PMID: 11064492]
29. **Hoffman M, Provatas A, Lyver A, Kanner R.** Pain management in the opioid-addicted patient with cancer. Cancer. 1991;68:1121-2. [PMID: 1913484]
30. **Portenoy RK, Dole V, Joseph H, Lowinson J, Rice C, Segal S, et al.** Pain management and chemical dependency. Evolving perspectives. JAMA. 1997;278: 592-3. [PMID: 9268282]
31. **Savage SR.** Addiction in the treatment of pain: significance, recognition, and management. J Pain Symptom Manage. 1993;8:265-78. [PMID: 7525743]
32. **Schnoll SH, Weaver MF.** Addiction and pain. Am J Addict. 2003;12 Suppl 2:S27-35. [PMID: 12857661]
33. **Martin JE, Inglis J.** Pain tolerance and narcotic addiction. Br J Soc Clin Psychol. 1965;4:224-9. [PMID: 5872680]
34. **Compton P, Gebhart G.** The neurophysiology and pain and addiction. In: Graham AW, Schultz TK, Mayo-Smith MF, Ries RK, Wilford BB, eds. Principles of Addiction Medicine. 3rd ed. Annapolis, MD: American Society of Addiction Medicine; 2003:1385-404.
35. **Gardner EL.** Brain reward mechanism. In: Lowinson JH, Ruiz P, Millman

RB, Langrod JG, eds. Substance Abuse: A Comprehensive Textbook. 3rd ed. Baltimore: Williams and Wilkins; 1997:51-84.
36. **Zacny JP, McKay MA, Toledano AY, Marks S, Young CJ, Klock PA, et al.** The effects of a cold-water immersion stressor on the reinforcing and subjective effects of fentanyl in healthy volunteers. Drug Alcohol Depend. 1996;42:133-42. [PMID: 8889412]
37. **Savage SR, Schofferman J.** Pharmacological therapies of pain in drug and alcohol addictions. In: Miller N, Gold M, eds. Pharmacological Therapies for Drug and Alcohol Addictions. New York: Dekker; 1995:373-409.
38. **Compton MA.** Cold-pressor pain tolerance in opiate and cocaine abusers: correlates of drug type and use status. J Pain Symptom Manage. 1994;9:462-73. [PMID: 7822886]
39. **Compton P.** Pain tolerance in opioid addicts on and off naltrexone pharmacotherapy: a pilot study. J Pain Symptom Manage. 1998;16:21-8. [PMID: 9707654]
40. **Prater CD, Zylstra RG, Miller KE.** Successful pain management for the recovering addicted patient. Prim Care Companion J Clin Psychiatry. 2002;4: 125-131. [PMID: 15014719]
41. **American Psychiatric Association.** Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Text revision (DSM-IV-TR). Washington, DC: American Psychiatric Association; 2000.
42. **Savage SR, Joranson DE, Covington EC, Schnoll SH, Heit HA, Gilson AM.** Definitions related to the medical use of opioids: evolution towards universal agreement. J Pain Symptom Manage. 2003;26:655-67. [PMID: 12850648]
43. **Eriator I.** Narcotic analgesics for chronic pain management. Current Review of Pain. 1998;2:193-200.
44. **Vukmir RB.** Drug seeking behavior. Am J Drug Alcohol Abuse. 2004;30: 551-75. [PMID: 15540493]
45. **White JM.** Pleasure into pain: the consequences of long-term opioid use. Addict Behav. 2004;29:1311-24. [PMID: 15345267]
46. Buprenorphine: an alternative to methadone. Med Lett Drugs Ther. 2003; 45:13-5. [PMID: 12592214]
47. **Johnson RE, Fudala PJ, Payne R.** Buprenorphine: considerations for pain management. J Pain Symptom Manage. 2005;29:297-326. [PMID: 15781180]
48. **Fishman SM, Wilsey B, Mahajan G, Molina P.** Methadone reincarnated: novel clinical applications with related concerns. Pain Med. 2002;3:339-48. [PMID: 15099239]
49. **Walsh SL, Eissenberg T.** The clinical pharmacology of buprenorphine: extrapolating from the laboratory to the clinic. Drug Alcohol Depend. 2003;70: S13-27. [PMID: 12738347]
50. **Gutstein HB, Akil H.** Opioid analgesics. In: Hardman JG, Goodman Gilman A, Limbird LE, eds. The Pharmacological Basis of Therapeutics. 10th ed. New York: McGraw-Hill; 2001:569-620.
51. **Inturrisi CE.** Clinical pharmacology of opioids for pain. Clin J Pain. 2002; 18:S3-13. [PMID: 12479250]
52. **McNicol E, Horowicz-Mehler N, Fisk RA, Bennett K, Gialeli-Goudas M, Chew PW, et al.** Management of opioid side effects in cancer-related and chronic noncancer pain: a systematic review. J Pain. 2003;4:231-56. [PMID: 14622694]
53. **Houtsmuller EJ, Walsh SL, Schuh KJ, Johnson RE, Stitzer ML, Bigelow GE.** Dose-response analysis of opioid cross-tolerance and withdrawal suppression during LAAM maintenance. J Pharmacol Exp Ther. 1998;285:387-96. [PMID: 9580575]
54. **Collett BJ.** Opioid tolerance: the clinical perspective. Br J Anaesth. 1998;81: 58-68. [PMID: 9771273]
55. **Doverty M, Somogyi AA, White JM, Bochner F, Beare CH, Menelaou A, et al.** Methadone maintenance patients are cross-tolerant to the antinociceptive effects of morphine. Pain. 2001;93:155-63. [PMID: 11427327]
56. **Laulin JP, Célèrier E, Larcher A, Le Moal M, Simonnet G.** Opiate tolerance to daily heroin administration: an apparent phenomenon associated with enhanced pain sensitivity. Neuroscience. 1999;89:631-6. [PMID: 10199599]
57. **Mao J.** Opioid-induced abnormal pain sensitivity: implications in clinical opioid therapy. Pain. 2002;100:213-7. [PMID: 12467992]
58. **Streltzer J.** Pain management in the opioid-dependent patient. Curr Psychiatry Rep. 2001;3:489-96. [PMID: 11707163]
59. **Compton P, Charuvastra VC, Kintaudi K, Ling W.** Pain responses in methadone-maintained opioid abusers. J Pain Symptom Manage. 2000;20:237-45. [PMID: 11027904]
60. **Ho A, Dole VP.** Pain perception in drug-free and in methadone-maintained human ex-addicts. Proc Soc Exp Biol Med. 1979;162:392-5. [PMID: 515020]
61. **Doverty M, White JM, Somogyi AA, Bochner F, Ali R, Ling W.** Hyperalgesic responses in methadone maintenance patients. Pain. 2001;90:91-6. [PMID: 11166974]
62. **Schall U, Katta T, Pries E, Klöppel A, Gastpar M.** Pain perception of intravenous heroin users on maintenance therapy with levomethadone. Pharmacopsychiatry. 1996;29:176-9. [PMID: 8895942]
63. **Compton P, Charuvastra VC, Ling W.** Pain intolerance in opioid-maintained former opiate addicts: effect of long-acting maintenance agent. Drug Alcohol Depend. 2001;63:139-46. [PMID: 11376918]
64. **Devillers JP, Boisserie F, Laulin JP, Larcher A, Simonnet G.** Simultaneous activation of spinal antiopioid system (neuropeptide FF) and pain facilitatory circuitry by stimulation of opioid receptors in rats. Brain Res. 1995;700:173-81. [PMID: 8624708]
65. **Dunbar SA, Pulai IJ.** Repetitive opioid abstinence causes progressive hyperalgesia sensitive to N-methyl-D-aspartate receptor blockade in the rat. J Pharmacol Exp Ther. 1998;284:678-86. [PMID: 9454814]
66. **Grilly DM, Gowans GC.** Acute morphine dependence: effects observed in shock and light discrimination tasks. Psychopharmacology (Berl). 1986;88:500-4. [PMID: 3085139]
67. **Devulder J, Bohyn P, Castille F, De Laat M, Rolly G.** A case of uncommon withdrawal symptoms after a short period of spinal morphine administration. Pain. 1996;64:589-91. [PMID: 8783325]
68. **Devulder J.** Hyperalgesia induced by high-dose intrathecal sufentanil in neuropathic pain. J Neurosurg Anesthesiol. 1997;9:146-8. [PMID: 9100184]
69. **Laulin JP, Larcher A, Célèrier E, Le Moal M, Simonnet G.** Long-lasting increased pain sensitivity in rat following exposure to heroin for the first time. Eur J Neurosci. 1998;10:782-5. [PMID: 9749743]
70. **Laulin JP, Maurette P, Corcuff JB, Rivat C, Chauvin M, Simonnet G.** The role of ketamine in preventing fentanyl-induced hyperalgesia and subsequent acute morphine tolerance. Anesth Analg. 2002;94:1263-9. [PMID: 11973202]
71. **Célèrier E, Laulin JP, Corcuff JB, Le Moal M, Simonnet G.** Progressive enhancement of delayed hyperalgesia induced by repeated heroin administration: a sensitization process. J Neurosci. 2001;21:4074-80. [PMID: 11356895]
72. **Célèrier E, Laulin J, Larcher A, Le Moal M, Simonnet G.** Evidence for opiate-activated NMDA processes masking opiate analgesia in rats. Brain Res. 1999;847:18-25. [PMID: 10564731]
73. **Célèrier E, Rivat C, Jun Y, Laulin JP, Larcher A, Reynier P, et al.** Long-lasting hyperalgesia induced by fentanyl in rats: preventive effect of ketamine. Anesthesiology. 2000;92:465-72. [PMID: 10691234]
74. **Kantor TG, Cantor R, Tom E.** A study of hospitalized surgical patients on methadone maintenance. Drug Alcohol Depend. 1980;6:163-73. [PMID: 6107237]
75. **Manfredi PL, Gonzales GR, Cheville AL, Kornick C, Payne R.** Methadone analgesia in cancer pain patients on chronic methadone maintenance therapy. J Pain Symptom Manage. 2001;21:169-74. [PMID: 11226767]
76. **Karasz A, Zallman L, Berg K, Gourevitch M, Selwyn P, Arnsten JH, et al.** The experience of chronic severe pain in patients undergoing methadone maintenance treatment. J Pain Symptom Manage. 2004;28:517-25. [PMID: 15504628]
77. **Jasinski DR.** Tolerance and dependence to opiates. Acta Anaesthesiol Scand. 1997;41:184-6. [PMID: 9061104]
78. **Bruera E, Macmillan K, Hanson J, MacDonald RN.** The cognitive effects of the administration of narcotic analgesics in patients with cancer pain. Pain. 1989; 39:13-6. [PMID: 2812850]
79. **Walsh T, Baxter R, Bowerman K, Leber B.** High-dose morphine and respiratory function in chronic cancer pain [Abstract]. Pain. 1981;S1:39.
80. **Hanks GW, Twycross RG, Lloyd JW.** Unexpected complication of successful nerve block. Morphine induced respiratory depression precipitated by removal of severe pain. Anaesthesia. 1981;36:37-9. [PMID: 6894066]
81. **Dole VP, Nyswander ME, Kreek MJ.** Narcotic blockade. Arch Intern Med. 1966;118:304-9. [PMID: 4162686]
82. **Jones BE, Prada JA.** Drug-seeking behavior during methadone maintenance. Psychopharmacologia. 1975;41:7-10. [PMID: 47638]
83. **Merrill JO, Rhodes LA, Deyo RA, Marlatt GA, Bradley KA.** Mutual mistrust in the medical care of drug users: the keys to the "narc" cabinet. J Gen Intern Med. 2002;17:327-33. [PMID: 12047728]
84. **Weissman DE, Haddox JD.** Opioid pseudoaddiction—an iatrogenic syndrome. Pain. 1989;36:363-6. [PMID: 2710565]
85. **Portenoy RK, Foley KM.** Chronic use of opioid analgesics in non-malignant pain: report of 38 cases. Pain. 1986;25:171-86. [PMID: 2873550]
86. **Evers GC.** Pseudo-opioid-resistant pain. Support Care Cancer. 1997;5:457-

60. [PMID: 9406359]
87. **Mao J, Price DD, Mayer DJ.** Mechanisms of hyperalgesia and morphine tolerance: a current view of their possible interactions. Pain. 1995;62:259-74. [PMID: 8657426]
88. **Kehlet H, Dahl JB.** The value of "multimodal" or "balanced analgesia" in postoperative pain treatment. Anesth Analg. 1993;77:1048-56. [PMID: 8105724]
89. **Botney M, Fields HL.** Amitriptyline potentiates morphine analgesia by a direct action on the central nervous system. Ann Neurol. 1983;13:160-4. [PMID: 6219612]
90. **Mitra S, Sinatra RS.** Perioperative management of acute pain in the opioid-dependent patient. Anesthesiology. 2004;101:212-27. [PMID: 15220793]
91. **O'Brien CP.** Drug addiction and drug abuse. In: Hardman JG, Goodman Gilman A, Limbird LE, eds. The Pharmacological Basis of Therapeutics. 10th ed. New York: McGraw-Hill; 2001:621-42.
92. **Tilson HA, Rech RH, Stolman S.** Hyperalgesia during withdrawal as a means of measuring the degree of dependence in morphine dependent rats. Psychopharmacologia. 1973;28:287-300. [PMID: 4734848]
93. **Paige D, Proble L, Watrous G, Kaalaas-Sittig J, Compton P.** PCA use in cocaine using patients: a pilot study. American Journal of Pain Management. 1994;4:101-5.
94. **Boyle RK.** Intra- and postoperative anaesthetic management of an opioid addict undergoing caesarean section. Anaesth Intensive Care. 1991;19:276-9. [PMID: 2069256]
95. **Pasternak GW.** Multiple opiate receptors: déjà vu all over again. Neuropharmacology. 2004;47 Suppl 1:312-23. [PMID: 15464147]
96. **O'Connor PG, Samet JH, Stein MD.** Management of hospitalized intravenous drug users: role of the internist. Am J Med. 1994;96:551-8. [PMID: 8017454]
97. **Fultz JM, Senay EC.** Guidelines for the management of hospitalized narcotics addicts. Ann Intern Med. 1975;82:815-8. [PMID: 1138596]
98. **Center for Substance Abuse Treatment.** Clinical Guideline for the Use of Buprenorphine in the Treatment of Opioid Addiction. Treatment Improvement Protocol (TIP) Series 40. DHHS publication no. (SMA) 04-3939. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2004.
99. **Edge WG, Cooper GM, Morgan M.** Analgesic effects of sublingual buprenorphine. Anaesthesia. 1979;34:463-7. [PMID: 474948]
100. **Moa G, Zetterström H.** Sublingual buprenorphine as postoperative analgesic: a double-blind comparison with pethidine. Acta Anaesthesiol Scand. 1990; 34:68-71. [PMID: 2309545]
101. **McCance-Katz EF.** Office-based buprenorphine treatment for opioid-dependent patients. Harv Rev Psychiatry. 2004;12:321-38. [PMID: 15764468]
102. **Johnson RE, Strain EC, Amass L.** Buprenorphine: how to use it right. Drug Alcohol Depend. 2003;70:S59-77. [PMID: 12738351]
103. **Center for Substance Abuse Treatment.** State Methadone Treatment Guidelines. Treatment Improvement Protocol (TIP) Series 1. DHHS publication no. (SMA) 93-1991. Rockville, MD: Substance Abuse and Mental Health Services Administration; 1993.

Current Author Addresses: Drs. Alford and Samet: Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118.

Dr. Compton: University of California, Los Angeles, School of Nursing, Factor Building 4-246, Box 956918, Los Angeles, CA 90095-6918.

The American Journal of Drug and Alcohol Abuse, 32: 555–560, 2006

ISSN: 0095-2990 print/1097-9891 online
DOI: 10.1080/00952990600920250

informa
healthcare

Introducing Evidence-Based Practices into Substance Abuse Treatment using Organization Development Methods

Maryann Amodeo, M.S.W., Ph.D.,[1] Michael A. Ellis, M.S.W.,[1] and Jeffrey H. Samet, M.D., M.A., M.P.H.[2]

[1]Boston University School of Social Work, Boston, Massachusetts, USA
[2]Boston University School of Medicine, Boston, Massachusetts, USA

Abstract: *Background*: Dissemination of evidence-based practices (EBPs) in addiction settings is a national priority. We tested Organization Development (OD) methods for dissemination. *Methods*: Using OD in two addiction treatment programs we developed an organization-specific treatment plan using employee work teams with the goals of changes in organizational policies and procedures and improvement in practitioner skills. *Results*: OD was effectively applied, but EBPs were premature for these addiction programs because they first needed to address more fundamental aspects of client-clinician interaction and agency treatment philosophy. *Conclusion*: The OD approach in addiction treatment is complementary to other technology transfer efforts by being: (a) ''organization-centered,'' engaging practitioners at all levels; (b) ''needs-focused,'' addressing concerns of the particular organization; (c) flexible in its responsiveness to readiness for change; and (d) relatively affordable. However, before absorbing EBPs, substance abuse treatment organizations must develop strengths in delivering fundamental aspects of care.

Keywords: Evidence-based practices, institutional change, organization development, substance abuse treatment

INTRODUCTION

During the 1990s, deficiencies were identified in addiction treatment agencies including: inadequately trained counselors with large caseloads,

Address correspondence to Maryann Amodeo, Boston University School of Social Work, 264 Bay State Road, Boston, Massachusetts 02215, USA. E-mail: mamodeo@bu.edu

inconsistent beliefs about medications and confrontational counseling, treatment plans without behavioral outcomes, and treatment modalities without a clear rationale (1–3). Treatment providers saw the lack of integration of research in practice as resulting from the presentation of research findings in a way that was not understandable, and studies that did not advance treatment knowledge. To facilitate the transfer of research-validated treatment technology to the field, management science was employed to assist treatment organizations adopt evidence-based practices (EBPs) (4, 5). Adoption of EBPs has been a significant challenge in all areas of health care for centuries; acceptance of the practice of consuming citrus fruits to eliminate scurvy took more than a century (6, 7). We used Organization Development (OD) to help addiction programs implement EBPs and illustrate implementation challenges and accomplishments with two specific case studies.

METHODS

Examples of EBPs include the Center for Substance Abuse Treatment's (CSAT) Treatment Improvement Protocols (TIPs). Brown and Flynn (8) criticized such approaches for their over-reliance on print media. McLellan emphasized the need for ''threat reduction'' methods to engage resistant staff, but noted that lack of skilled personnel was a barrier to innovative methods (9). One model used on-site clinical instruction, mentoring, and supervision to help workers use an EBP approach to family treatment of substance abuse. Another model (11) used computer-assisted collection of patient information on the Addiction Severity Index and increased treatment completion rates from 45 to 70%. Since staff training often results in only temporary behavior change, in the absence of organizational supports (1, 12, 13), the OD approach was used to accomplish essential change tasks (8): Prepare the organization for change through motivational methods, assure sufficient resources for implementation, and engage in outcome assessments.

The OD approach (14) engages employee teams in problem solving for organizational assessment, diagnosis, and treatment. OD is distinguished by the use of a *change agent as catalyst* to gather organizational data, feed it back to the organization, and facilitate cross-functional agency teams in diagnosing problems and planning and executing remedial action (15, 16). The consultants were available to each agency for about 18 months, for about 2 1/2 person-days per week. Employee ''opinion leaders'' were selected by the consultants and top management to participate in teams that chose the intervention targets, designed the work plans, and implemented the changes (14). Desired outcomes included: increased use of EBPs, knowledge of their benefits, and mechanisms to ensure institutionalized change.

The consultants familiarized administrators and staff with a range of NIDA-endorsed EBPs using EBP manuals and protocols and tools for client assessment and treatment planning. They recommended ways to modify client intake and assignment procedures, and improve staff and management communication.

TWO CASE STUDIES: DIVERGENT EXPERIENCES

Organization X

This organization is a free-standing, multi-function addiction program that serves a racially- and ethnically-diverse, inner-city client population. A needs assessment used focus groups with administrators, staff and clients with a goal of introducing EBPs or EBP-related program elements. Support was expressed for Motivational Interviewing as an EBP (18, 19). However, we discovered two organizational problems: no common standards of client care, and not all employees saw themselves as part of the therapeutic environment. Specific problems included confidentiality violations; breaches in staff-client boundaries; and lack of respect for client self-determination. The agency's Consumer Advisory Committee was helpful in focusing on these dysfunctional staff-client interactions. Further, staff members took offence when clients relapsed or left the program, emphasized client deficits instead of strengths; and did not include clients in treatment planning.

The staff team recognized that implementing EBPs would require that the agency develop standards for treating all clients as respected partners in treatment. Administrators were surprised and initially alarmed at these conclusions, but then empowered the change team to write and implement standards of care, adding them to new-employee orientation and clinical staff training, and evaluating compliance with the standards. While such direct feedback to administrators can cause a "stonewalling" of change, it was positive here and provided quicker consequences for employees who failed to function professionally in client interactions.

Clinical staff also learned Motivational Interviewing (18, 19) through training sessions, "training of trainers" to continue the training throughout the agency, and a Standards of Client Care manual emphasizing a motivational approach to client care.

Organization Y

Organization Y is also a well-established addiction treatment agency serving a racially- and ethnically-diverse, inner city client population with a continuum of care. The consultants initially forged a strong alliance with administrators, in contrast to Organization X where they relied on an

employee change team. The process in organization Y was more "top-down" for interpreting needs assessment findings, prioritizing needs, and determining the EBPs to be adopted. Training targeted counselor and supervisor skills. To reduce initial counselor and supervisor resistance to EBPs, clinicians were told that they would not need to abandon their current treatment approaches but that EBPs could be helpful in addressing chronic or nagging problems in client care; EBPs could make the treatment easier and more interesting for the clinician, and more understandable, engaging and effective for the client. After several hours of training, staff were asked to apply their newly acquired skills to selected clients on their caseloads. The staff reviewed treatment plans completed prior to the EBP training and revised those documents in keeping with new EBPs.

Administrators made policy and procedural changes to reinforce the use of EBPs. These included sanctions for client non-compliance with treatment, and monitoring counselor record keeping for client's stage of readiness for change and its use in treatment planning. However, administrators were distressed at troublesome staff behaviors including giving and receiving gifts, personal relationships with clients, and inappropriate use of counselor authority (both excessive use of authority and failure to exercise authority when needed). When administrators confronted staff, some aggressively defended their behavior, stating that the multiply diagnosed, multi-stressed, and ethnically diverse clients on their caseloads benefited from this non-traditional approach.

The extent to which counselors and supervisors used their newly acquired EBPs remained unclear. Some believed that their prior counseling methods (however poorly defined) were equally effective as EBPs and that clinicians should be free to utilize whatever methods they chose; this frustrated administrators, but they lacked effective sanctions (e.g., staff performance evaluation and dismissal mechanisms).

Two outcomes were (a) using new methods for organizing treatment groups based on clinical rather than administrative factors, and (b) developing a treatment philosophy based on NIDA principles, incorporating the goals of abstinence and harm reduction, and providing guidelines for intensifying treatment for poorly performing clients.

DISCUSSION

Two variants of the OD approach were used in addiction agencies and facilitated the introduction of EBPs in both settings. The approach was "systems-centered," engaging key players in the change effort and allowing the addiction program to remain in the driver's seat, specifying the desired target areas, content, and processes of change. The OD approach

is also ''needs-centered,'' emphasizing what the organization's members think needs improvement, at its level of readiness for change. Thus, this approach incorporates for organizations some of the same principles of ''motivational enhancement'' (19) that have been developed for individuals. The ''open-agenda,'' with the organization choosing the focus of change, provides flexibility that is crucial in facilitating a permissive environment for organizations to uncover issues of concern to staff, administrators, and clients. The OD approach involves time-limited costs that can be spread over some months. The actual cost of this model for one agency, involving 2–3 person-days per week for 12 months is approximately $25,000, depending on the level of effort required of the consultants. Employee teams can be expensive for the organization, however, and compensatory time may need to be provided for the change team.

The OD approach could be available to a broad audience of addiction programs. Areas where special expertise is required include group facilitation and conflict-resolution. This work could be done by a program ''insider'' but only if he/she has distance from the organization's problems and politics, since a major role of the ''change agent'' is to ally with all parts of the organization. Finally, expertise is needed in the area of EBPs and knowledge of the benefits and limitations of EBPs. Limitations of this evaluation of the OD approach include testing in only two organizations; and the intervention team was also the evaluation team.

OD is another option available for addiction programs interested in adopting EBPs and can enhance the effectiveness of other approaches. The model can harness the energy and creativity of the workers who will use the changes, and includes the supervisors and administrators whose support will be needed to reinforce the changes.

ACKNOWLEDGEMENTS

Research support was provided by the National Institute on Drug Abuse, Grant R25 DA13582. The authors wish to thank David Rosenbloom, Ph.D., Professor, Boston University School of Public Health, for his salient comments in reviewing this article.

REFERENCES

1. Lamb S, Greenlick MR, McCarty D. (Eds.) Bridging the Gap between Practice and Research: Forging Partnerships with Community-based Drug and Alcohol Treatment. Washington, D.C.: National Academy Press, 1998, 264.
2. Simpson DD. Effectiveness of drug-abuse treatment: A review of research from field settings. In Treating Drug Abusers Effectively. Egertson JA, Fox DM, Leshner AI, eds. Malden, MA: Blackwell, 1997; 41–73.

3. Forman RF, Bovasso G, Woody G. Staff beliefs about addiction treatment. J Subst Abuse Treat 2001; 21:1–9.
4. Simpson, DD. A conceptual framework for transferring research to practice. J Subst Abuse Treat 2002; 22:171–182.
5. Backer TE, David SL, Soucy G. Introduction: the challenge of technology transfer. In Reviewing the Behavioral Science Knowledge Base on Technology Transfer. NIDA Research Monograph #155; Backer TE, David SL, Soucy G. eds. Rockville MD: National Institute on Drug Abuse, 1995; 1–20.
6. Mosteller F. Innovation and evaluation. Science 1981; 211(4485):881–886.
7. Rogers EM. Diffusion of Innovations. 4th ed. New York, NY: The Free Press, 1995, 519.
8. Brown BS, Flynn PM. The federal role in drug abuse technology transfer: a history and perspective. J Subst Abuse Treat 2002; 22:245–257.
9. McLellan AT. Technology transfer and the treatment of addiction: what can research offer practice? J Subst Abuse Treat 2002; 22:169–170.
10. Liddle HA, Rowe CL, Quille TJ, Dakof GA, Mills DS, Sakran E, Biaggi, H. Transporting a research-based adolescent drug treatment into practice. J Subst Abuse Treat 2002; 22:231–243.
11. Carise D, Cornely, W, Gurel O. A successful researcher-practitioner collaboration in substance abuse treatment. J Subst Abuse Treat 2002; 23: 157–162.
12. Baldwin TT, Ford JK. Transfer of training: A review and directions for future research. Personnel Psychology 1988; 41:63–105.
13. Clarke N. Job/work environment factors influencing training transfer with a human service agency: some indicative support for Baldwin and Ford's transfer climate construct. International Journal of Training and Development 2002; 6(3):146–162.
14. French WL, Bell CH. Organization development: Behavioral Science Interventions for Organization Improvement. 6th ed. Upper Saddle River, NJ: Prentice-Hall, 1999, 343.
15. French WL, Bell CH. Organization Development: Behavioral Science Interventions for Organization Improvement. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1978, 271.
16. Gosin MN, Dustman PA, Drapeau AE, Harthun MI. Participatory action research: Creating an effective prevention curriculum for adolescents in the Southwestern U.S. Health Education Research 2003; 18(3):363–379.
17. French WL, Bell C. Organization Development: Behavioral Science Interventions for Organization Improvement. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1984, 347.
18. Miller WR, Zweben A, DiClemente CC, Rychtarik RG. *Motivational Enhancement Therapy Manual*, Project MATCH Monograph Series, Volume 2, NIH Publication No. 94-3723, National Institute on Alcohol Abuse and Alcoholism: Rockville, MD, 1992.
19. Miller WR, Rollnick S. Motivational Interviewing: Preparing People for Change. 2nd ed. New York, The Guilford Press 2002, 428.

ELSEVIER

Cancer Detection and Prevention 30 (2006) 34–37

Cancer Detection and Prevention

www.elsevier.com/locate/cdp

Cancer prevention trials and primary care physicians: Factors associated with recommending trial enrollment

Tracy A. Battaglia MD, MPH[a,c,*], Arlene Ash PhD[a], Marianne N. Prout MD, MPH[b,c], Karen M. Freund MD, MPH[a,c]

[a] *Women's Health Unit, Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, 720 Harrison Avenue, DOB Suite 1108, Boston, MA 02118-2334, USA*

[b] *Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA*

[c] *Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, MA, USA*

Accepted 23 September 2005

Abstract

Background: To explore the willingness of primary care providers (PCPs) to encourage enrollment of patients into cancer prevention trials. *Methods:* A self-administered survey was mailed to a random sample of PCPs in three geographic regions. Physicians were asked questions about their knowledge and attitudes towards cancer prevention trials. We presented a clinical vignette of a woman at high risk for breast cancer and asked if they would encourage her enrollment into a breast cancer chemoprevention trial (yes/no). Each survey included one of 16 possible clinical vignettes where patient characteristics (age, race socioeconomic status, physical mobility and co-morbidity) varied dichotomously. Bivariate analyses and logistic models were used to examine the independent effects of patient and physician characteristics on physician decisions. *Results:* Two hundred and sixty-six surveys (50% response) were analyzed. The mean age of respondents was 48; 54% were White, 35% Asian and 5% Black. By design physicians were evenly distributed by gender, specialty and geographic location. Overall, 53% would encourage enrollment into a breast cancer chemoprevention trial. Significant predictors of a recommendation to enroll were: geographic location in California or Georgia, younger vignette patient and anticipating an increase in patient trust after recommending enrollment. *Conclusion:* PCPs are less likely to encourage elderly patients to enroll into cancer chemoprevention trials. Decisions differ based on geographic location and perceived trust in the patient–provider relationship. To achieve successful enrollment, trial investigators must continue to educate PCPs and ensure a strong PCP–patient relationship is maintained.

Keywords: Breast neoplasms; Chemoprevention; Clinical trials; Decision making; Experimental design; High-risk patients; Patient-provider relationship; Prevention trials; Primary care providers; Primary prevention; Regional variability; Response rate; Tamoxifen; Trial enrolment

1. Introduction

Recruiting subjects to cancer chemoprevention trials is a new challenge resulting from expanding prevention modalities. With cancer incidence the endpoint, chemoprevention trials require large numbers of at-risk participants for extended periods. Thus, recruitment strategies target ''healthy'' asymptomatic individuals who may not perceive themselves as at-risk. Such eligible participants are not typically under oncology specialty care. As a result, primary care physicians (PCPs), who are increasingly called upon to incorporate risk assessment into their practice [1], are important partners for trial investigators.

Lack of physician recommendation is a leading reason for inadequate accrual in both cancer treatment [2–4] and chemoprevention trials [5,6]. This study sought to identify factors associated with PCP decisions to encourage patient enrollment into a breast cancer chemoprevention trial.

2. Methods

Previous work [7] informed the development of a 51-item questionnaire to assess physician knowledge and attitudes

* Corresponding author. Tel.: +1 617 638 8036; fax: +1 617 638 8026.
E-mail address: Tracy.Battaglia@bmc.org (T.A. Battaglia).

0361-090X/$30.00
doi:10.1016/j.cdp.2005.09.005

Table 1
Survey measurements: physician knowledge and attitudes

	Number of questions	Measurement	Scale
Awareness of breast cancer chemoprevention trials	5	Sum of positive response to five yes/no questions	Scale 0–5; 0, not aware; 5, very aware
Knowledge of breast cancer chemoprevention trial outcomes	4	Sum of correct response to four yes/no questions	Scale 0–4; 0, none correct; 4, all correct
Experience with tamoxifen	2	Sum of positive response to two yes/no questions	Scale 0–2; 0, no personal experience; 2, experienced
Trust score	1	Response to a single five-point likert question: "*If you encouraged this patient to enroll in this trial, how would it change her trust in you as her doctor*?"	Scale 1–5; 1, trust you much less; 5, trust you much more
Control score	1	Response to a single three-point likert question: "*If this patient were to enroll in a breast cancer prevention trial, do you feel that care of this patient from your office will . . .*"	Scale 1–3; 1, decrease; 3, increase

(Table 1) regarding chemoprevention trials. To reflect geographic variation in cancer care [8,9], the self-administered survey was mailed to PCPs in three regions: Detroit, MI; Atlanta, GA and San Francisco, CA. Physicians were randomly selected from local Board of Registration Databases and stratified by gender, specialty (family practice, internal medicine) and geographic region. Non-responding physicians received up to three surveys and eight phone calls. The third survey included a prepaid US$ 10 incentive.

We used a *fractional* factorial experimental design [10] to examine medical decision-making in the absence of observational confounding. We developed several versions of a clinical vignette describing a physician encounter with an at-risk woman who recently underwent a benign breast biopsy. In each version, the patient presents with a medically identical "case" however, the patient may differ by five dichotomous characteristics: age (80 versus 65), race (Black versus White), socioeconomic status (middle versus low-income), co-morbidity (none versus hypertension and diabetes) and mobility (ambulatory versus aided by walker). Using the five dichotomous factors in a complete factorial design would yield 2^5 or 32 possible vignette combinations of all five characteristics. In order to allow for fewer versions of the vignette, a balanced *fraction* (in this case 1/2 = 16) of all possible vignette combinations was selected. The subset chosen was well-balanced (e.g. half were 80, half White, etc.) and therefore allows for statistically powerful estimates of all main effects and two-way interactions. Each participant was presented only one of the 16 versions of the clinical vignette and asked whether they would encourage her enrollment into a trial randomizing to tamoxifen versus placebo to reflect the ongoing National Surgical Adjuvant Breast and Bowel Project Breast Cancer Prevention Trial (P-1) [11]. Halfway through data collection, the P-1 trial was unblinded after finding a 49% reduction in invasive breast cancer among those taking tamoxifen. Thereafter, the vignette asked participants whether they would encourage enrollment into a trial randomizing to tamoxifen versus raloxifene to reflect the P-2 trial [12].

Preliminary analyses found no difference in physician decision-making pre and post P-1 trial findings, so data were combined for future analyses. Chi-square and Wilcoxen rank sum tests were used for bivariate comparisons. A multiple logistic regression model included significant variables from bivariate analyses; missing values were assigned the mean value of the responding participants.

3. Results

Of 1211 surveys mailed, 565 were eligible (6 deceased, 46 retired, 144 not practicing primary care, 450 incorrect address) and 281 (50%) responded. Fifteen were dropped for not answering the main outcome question, leaving 266 surveys for analysis. Reflecting study design, respondents were evenly distributed by gender, specialty and geographic location. Mean age of participants was 48; 54% identified themselves as White, 35% Asian, 5% Black; 39% had one or more patients enrolled in some clinical trial and the median number of patients seen weekly was 85.

Overall, 53% of participants encouraged the woman in the clinical vignette to enroll into a breast cancer chemoprevention trial. Physicians in Michigan were less likely to encourage enrollment than physicians in California or Georgia (40% versus 62% and 59%, $p = 0.007$). Though neither awareness of the existing breast cancer chemoprevention trials nor knowledge of the P-1 trial outcome were associated with physician decisions, 70% of physicians who encouraged enrollment were experienced with using tamoxifen compared with only 58% of those who did not encourage enrollment ($p = 0.04$) (Table 2). The majority of physicians did not feel that encouraging trial enrollment would change patient trust (70% trust score = 3) or the control they had in patient care (71% control score = 2). However, 31% of physicians who encouraged trial enrollment felt that patient trust would increase (trust score = 4 and 5) after encouraging the trial compared with only 6% among those who did not encourage trial enrollment ($p < 0.0001$). Similarly, 15% of those who encouraged trial enrollment felt care from their office would increase

Table 2
Variation in decision to encourage enrollment by knowledge and attitude scores

Knowledge/attitude score	Percentage of physicians		*p*-Value	OR (95% CI)[a]
	Encourage enrollment $N = 141$	Do not encourage enrollment $N = 125$		
Awareness				
0, not aware	4	7	0.21	–
1	6	7		
2	43	30		
3	31	26		
4	14	22		
5, very aware	2	8		
Knowledge				
0, none correct	2	2	0.10	–
1	29	48		
2	21	9		
3	20	14		
4, all correct	28	27		
Experience				
0, no experience	10	15	0.04	0.4 (0.16, 1.01)
1	20	27		0.6 (0.28, 1.12)
2, experienced	70	58		xxx
Trust				
1, much less	0	6	<0.0001	
2, less	2	14		0.07 (0.01, 0.33)
3, no change	67	74		xxx
4, more	24	6		8.3 (3.3, 20.9)
5, much more	7	0		
Control				
1, decrease	15	20	0.06	1.14 (0.51, 2.56)
2, stay same	70	73		xxx
3, increase	15	7		2.6 (0.89, 7.57)

[a] Multivariate model controlled only for those variables statistically significant in bivariate analyses: vignette patient age, physician geographic location, experience with tamoxifen, trust score and control score.

(control score = 3) after encouraging the trial compared with only 7% who did not encourage trial enrollment ($p = 0.06$).

Physicians recommended trial enrollment 44% of the time when the vignette patient was 80 years old compared to 61% when she was 65 ($p = 0.006$). No differences in physician decision-making were found based on patient race, socioeconomic status, co-morbidity or mobility.

In multiple logistic regression analysis controlling for only those variable significant in bivariate analyses (vignette patient age, physician geographic location, experience with tamoxifen, trust and control score) experience with tamoxifen and control in patient care were no longer significant predictors of encouraging trial enrollment while geographic location in California or Georgia, younger vignette patient age and higher trust score remained significant (Table 2).

4. Discussion

Primary care providers are uniquely positioned to identify, educate and encourage subject participation for cancer prevention trials. This is the first reported study to look at PCP decision-making for a chemoprevention trial. Given a hypothetical case of a woman approached to enroll in a breast cancer chemoprevention trial, greater than half surveyed were willing to encourage the patient to enroll. This pattern did not change after the P-1 trial [11] released its findings in favor of tamoxifen which underscores the importance of understanding what factors do influence PCP willingness to encourage enrollment. This is especially true as more recent studies find that even in the presence of favorable evidence, PCPs are unwilling to recommend chemoprevention to high risk patients [13].

Our finding that advancing age is associated with not recommending chemoprevention trial enrollment supports the notion that provider recommendation plays a major role in the representation of older subjects in clinical trials [14,15]. Eligibility criteria for the P-1 trial [11] served as the basis for our study. Though there was no upper age limit for P-1 trial enrollment, women were considered ineligible if their life expectancy was <10 years. The difference in enrollment of 80-year-old versus 65-year-old women may reflect provider's appropriate perception of this risk–benefit ratio. About half of the 80-year-old women in the vignettes

would have been eligible for enrollment based on a favorable life expectancy since the average woman who reaches the age of 80 has a life expectancy of over 7 years [16]. The failure to find an association between physician decision and patient race, co-morbidity or physical mobility suggests that PCPs may use age alone in their decision-making. This finding goes against recommended practices to consider life expectancy when making clinical decisions among the elderly.

Regional variability in physician decision-making may reflect similar geographic variation in the use of breast conserving surgery [17,18]. This study also supports our previous work that the integrity of the patient–provider relationship may influence physician recommendation [7]. Though most physicians surveyed did not feel that patient trust or control in patient care would change whether or not they recommend enrollment, those who did feel that trust or control in patient care would increase after recommending enrollment were more likely to recommend the trial. Similarly, those who anticipate a loss of patient trust or control in their medical care were less likely to recommend the trial. Though the association of control with decision-making was not significant in multivariate modeling, further analyses found trust and control to be closely correlated. This finding may reflect physician concern over preserving the threatened physician–patient relationships in the current managed care environment [19,20].

One limitation to our study findings is the 50% response rate. Though not optimal, this is consistent with existing literature utilizing physician administered survey methodology [21]. Thirty-seven percent of our study sample were ineligible because of incorrect address using the Board of Registration Files but is consistent with our prior work [22]. This high rate may reflect trainees who are transient. Therefore our findings may not be generalizable to trainees or those recently completing training.

To maximize accrual to cancer prevention trials, our study suggests that trial investigators seeking PCP support must not only educate these physicians, especially in certain geographic locations, but they must help facilitate patient trust and develop systems to allow the provider to maintain active in their patient's care during and after trial enrollment.

Acknowledgements

This research was funded in part, by the following grants: CA 73297-01, K12-HD43444 Building Interdisciplinary Research Careers in Women's Health (BIRCWH) Faculty Scholar Award, CCCDA-03-228-01 ACS Career Development Award and the Avon Foundation. The authors are grateful to Kathryn Roloff for her insights and assistance in the preparation of the drafts of this manuscript.

References

[1] Chemoprevention of breast cancer: recommendations and rationale. U.S. Preventive Services Task Force. Ann Intern Med 2002;137:56–8.
[2] Benson AB, Prokop Pregler J, Bean JA, Rademaker AW, Eshler B, Anderson K. Oncologists' reluctance to accrue patients onto clinical trials: an Illinois Cancer Center study. J Clin Oncol 1991;9:2067–75.
[3] Winn RJ. Obstacles to the accrual of patients in clinical trials in the community setting. Semin Oncol 1994;21:112–7.
[4] Mansour EG. Barriers to clinical trials. Part III: knowledge and attitudes of health care providers. Cancer 1994;74(Suppl. 9):2672–5.
[5] Lerman C, Rimer BK, Daly M, Lustbader E, Sands C, Balshem A, et al. Recruiting high risk women into a breast cancer health promotion trial. Cancer Epidemiol Biomark Prev 1994;3(3):271–6.
[6] Kinney AY, Richards C, Vernon SW, Vogel VG. The effect of physician recommendation on enrollment in the breast cancer chemoprevention trial. Prev Med 1998;27(5 Pt 1):713–9.
[7] Frayne SM, Mancuso M, Prout MN, Freund KM. Attitudes of primary care physicians toward cancer-prevention trials: a focus group analysis. J Natl Med Assoc 2001;93(11):450–7.
[8] Nattinger AB, Gottlieb MS, Veum J, Yahnke D, Goodwin JS. Geographic variation in the use of breast-conserving treatment for breast cancer. New Eng J Med 1992;326(17):1102–7.
[9] Farrow DC, Hunt WC, Samet JM. Geographic variation in the treatment of localized breast cancer. New Eng J Med 1992;326(17):1097–101.
[10] Cochran WG, Cox GM. Experimental designs New York, NY: John Wiley & Sons; 1957.
[11] Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 1998;90(18):1371–88.
[12] Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Wolmark N. The study of tamoxifen and raloxifene: preliminary enrollment data from a randomized breast cancer risk reduction trial. Clin Breast Cancer 2002;3(2):153–9.
[13] Taylor R, Taguchi K. Tamoxifen for breast cancer chemoprevention: low uptake by high-risk women after evaluation of a breast lump. Ann Fam Med 2005;3:242–7.
[14] Yee KW, Pater JL, Pho L, Zee B, Siu LL. Enrollment of older patients in cancer treatment trials in Canada: why is age a barrier? J Clin Oncol 2003;21(8):1618–23.
[15] Hutchins LF, Unger JM, Crowley JJ, Coltman Jr CA, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. New Eng J Med 1999;341(27):2061–7.
[16] National Surgical Adjuvant Breast and Bowel Project: P-1 Protocol. Section 6.0 subject eligibility and ineligibility, September 1994.
[17] Nattinger AB, Gottlieb MS, Veum J, Yahnke D, Goodwin JS. Geographic variation in the use of breast-conserving treatment for breast cancer. New Eng J Med 1992;326(17):1102–7.
[18] Farrow DC, Hunt WC, Samet JM. Geographic variation in the treatment of localized breast cancer. New Eng J Med 1992;326(17):1097–101.
[19] Emanuel EJ, Dubler NN. Preserving the physician–patient relationship in the era of managed care. JAMA 1995;273(4):323–9.
[20] Glass RM. The impact of managed care on patients' trust in medical care and their physicians. JAMA 1996;275(21):1693–7.
[21] Asch DS, Jedrziewski MK, Christakis NA. Response rate to mail surveys published in medical journals. J Clin Epidemiol 1997;50:1129–36.
[22] Marwell SL, Fruend KM, Barry PP. Patient factors associated with breast cancer screening among older women. J Am Geriatr Soc 1996;44:1210–4.

How Do Providers Assess Antihypertensive Medication Adherence in Medical Encounters?

Barbara G. Bokhour, PhD,[1,2] Dan R. Berlowitz, MD, MPH,[1,2] Judith A. Long, MD,[3,4,5] Nancy R. Kressin, PhD[1,2]

[1]Center for Health Quality, Outcomes and Economic Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, USA; [2]Boston University School of Public Health, Boston, MA, USA; [3]Philadelphia VA Center for Health Equity Research and Promotion, PA, Philadelphia, USA; [4]University of Pennsylvania School of Medicine, PA, Philadelphia, USA; [5]Leonard David Institute of Health Economics, PA, Philadelphia, USA

BACKGROUND: Poor adherence to antihypertensives has been shown to be a significant factor in poor blood pressure (BP) control. Providers' communication with patients about their medication-taking behavior may be central to improving adherence.

OBJECTIVE: The goal of this study was to characterize the ways in which providers ask patients about medication taking.

DESIGN: Clinical encounters between primary care providers and hypertensive patients were audiotaped at 3 Department of Veterans' Affairs medical centers.

PARTICIPANTS: Primary care providers ($n=9$) and African-American and Caucasian patients ($n=38$) who were diagnosed with hypertension (HTN).

APPROACH: Transcribed audiotapes of clinical encounters were coded by 2 investigators using qualitative analysis based on sociolinguistic techniques to identify ways of asking about medication taking. Electronic medical records were reviewed after the visit to determine the BP measurement for the day of the taped encounter.

RESULTS: Four different aspects of asking about medication were identified: structure, temporality, style and content. Open-ended questions generated the most discussion, while closed-ended declarative statements led to the least discussion. Collaborative style and use of lay language were also seen to facilitate discussions. In 39% of encounters, providers did not ask about medication taking. Among patients with uncontrolled HTN, providers did not ask about medications 33% of the time.

CONCLUSION: Providers often do not ask about medication-taking behavior, and may not use the most effective communication strategies when they do. Focusing on the ways in which providers ask about patients' adherence to medications may improve BP control.

KEY WORDS: hypertension; medication adherence; provider-patient communication.
DOI: 10.1111/j.1525-1497.2006.00397.x
J GEN INTERN MED 2006; 21:577–583.

Hypertension (HTN) affects more than 29% of the adult population[1,2] and increases the risk for adverse outcomes. Effective treatment of HTN has been shown to reduce this risk,[3,4] yet studies have consistently shown that most patients with established HTN have poorly controlled blood pressure (BP)[5,6] and that 30% to 70% of patients do not take their BP medications as prescribed[1,7–9].

Poor adherence may be partially due to problems of access and cost of medication, however, patients' beliefs about HTN and medication may also play a significant role. In a national survey, hypertensive patients reported that they discontinued antihypertensive therapy because they believed that they were cured (46%) and thought that they had been advised to stop by their provider (25%).[10] Patients' nonadherence to medications has been attributed to both intentional (i.e., a conscious decision not to take medications) and unintentional (i.e., a failure to take medications due to poor understanding or forgetfulness) reasons.[11] And yet, providers may be unaware of patients' medication-taking behavior and patients' understanding of how to use medications. Without this information, it is difficult for providers to distinguish between drug efficacy problems and medication adherence issues. Effective communication is key to providers' assessment of patients' adherence to medications.

A patient-centered approach in which the provider engages the patient in a process of shared decision making has been identified as an important factor in improving patient adherence.[12–14] Studies of patient-physician communication about medication-taking have found little evidence of joint patient and physician involvement in decision making and information sharing during consultations about medications,[14] and a dearth of in-depth questioning of patients about their medication taking behaviors.[15]

One effective communication strategy which has repeatedly helped improve clinical outcomes such as treatment adherence is "patient-centered counseling."[16] This multifaceted strategy fosters clinicians' abilities to identify barriers to treatment adherence relevant to each individual patient.[16] In this paper we focus on one specific facet, provider's assessment of a patient's medication adherence, which we posit as a crucial element for effective decision making about HTN management. We define an effective communication strategy as one that elicits detailed information from the patient about how he/she is taking his/her antihypertensive medications. Although asking questions is a core feature of providers' clinical assessments, little research has focused on *how* questions are asked. Fifteen years ago, Steele et al.,[17] examining conversations about HTN medications, found that using direct and information-intensive approaches to assessing adherence were more effective than indirect approaches in detecting adherence problems, yet it is not clear how the growing literature and emphasis on

The authors have no conflicts of interest to report.

Address correspondence and requests for reprints to Dr. Bokhour: Center for Health Quality, Outcomes and Economic Research, ENRM Veterans Hospital (152), 200 Springs Road, Bedford, MA 01730 (e-mail: bokhour@bu.edu).

Manuscript received March 10, 2005
Initial editorial decision May 27, 2005
Final acceptance December 14, 2005

patient-physician communication may have influenced contemporary interactions.

We conducted a study to examine the ways in which providers ask patients about medication taking in clinical encounters using a qualitative sociolinguistic approach to analyze audiotapes of naturally occurring clinical encounters.[18,19] We conducted an in-depth analysis of the forms of language used by providers in order to characterize the different communication strategies that they used to ask about patients' medication-taking behavior, how patients responded, and the relationship between these strategies and patients' BP control.

METHODS

The data were collected as part of the Physician Intervention to Improve Control of Hypertension (PITCH) project, funded by Department of Veterans Affairs, HSR&D. The goal of the PITCH project was to assess the impact of a provider intervention to improve communication with patients about HTN. In this paper we report the results from analyses of baseline audiotaped patient-provider encounters before implementation of the intervention.

Participants

Participants were recruited in primary care clinics at 3 large urban Veterans' Affairs Medical Centers. Patients who were African American or Caucasian and had a documented diagnosis of HTN in at least 2 encounters in a single calendar year were eligible for the study. A convenience sample of eligible patients who presented in the clinic for a non-urgent primary care appointment during the recruitment period were approached by a research assistant and asked to have their appointment audiotaped. We recruited equal numbers of African-American and Caucasian patients. Providers were approached by study investigators at each site and were eligible to participate if they were primary care physicians, physician assistants, or nurse practitioners participating in the parent project, and were the treating provider for enrolled study participants. Three providers from each site were audiotaped in clinical encounters with a total of 39 patients (9 to 15 patients per site); logistical considerations required we stop recruitment at this point. The Institutional Review Boards at all 3 institutions approved the study and all patients and providers provided written informed consent.

Data Collection

We audiotaped primary care visits of hypertensive patients with their providers. Providers and patients were told that they were participating in a study to examine communication between patients and providers. Research assistants set up a tape recorder in the exam room, started the recording, and left the room. Electronic medical records were reviewed after the visit to determine the BP measurement for the day of the taped encounter.

Analysis. All encounters were transcribed verbatim. Through a process of open inductive coding, we identified 23 different communication activities related to HTN–that is, any communication sequence in which HTN or BP was referenced. We then identified segments in which the provider asked the patient how s/he was taking prescribed medications and called this activity, "taking medication." One investigator, an expert in sociolinguistic analyses (B.B.), examined these segments and sub-coded the provider's utterances based on sociolinguistic discourse markers and structure,[19] including (1) the structure, i.e., the use of interrogatives versus declaratives and the use of open-ended versus closed-ended statements; (2) the verb form indicating temporality (assessment of general behavior vs. specific time-limited behavior); (3) the content of language used; and (4) the style of interaction. The content was categorized based on types of terminology used in discussing medication. The style of interaction was identified by the ways in which the provider responded to patients' expressed problems or concerns. Based on these analyses, we developed a taxonomy of communication strategies that providers used to ask patients about their medication-taking behavior. A second investigator (N.K.) reviewed coded segments and, through iterative consensus, agreed upon the taxonomy. We also examined the content and extent of patient responses to different ways of being asked about medication-taking behaviors. Patient's BP control was determined to be either controlled (< 140/90) or uncontrolled (> 140/90), and we examined the different strategies used in these 2 groups.

RESULTS

One encounter was eliminated from analysis due to poor audio quality, leaving a total of 38 dialogs for analysis. Patients (Table 1) were all male with an average age of 65.9 years. More than two-thirds had graduated from high school and 39% had some level of higher education. Patients were largely poor, with over 42% earning $20,000 or less per year and 37% earning between $20,001 and $40,000. Fifty percent self-identified as African American and 50% self-identified as Caucasian.

There were 15 interactions (39%) in which no segments of text were coded "taking medication." At no time during these encounters did the providers ask about their patients' medication-taking behavior. In 2 cases there was no discussion about HTN at all. In the others, communication about HTN included explanations of HTN and its sequelae, provision of new prescriptions for medications, and discussions of diet and exercise to control HTN. In the 23 remaining encounters (61%), there was at least 1 instance of asking about medications. Below we describe the communication strategies providers used in these encounters to ask about medication-taking behavior.

Table 1. Patient Characteristics (n=38)

Male (%)	100%
Age	
Mean (SD)	65.9 (11.5)
Race (%)	
African American	50% (n=19)
Caucasian	50% (n=19)
Education	
Mean (SD)	12.1 y (2.6)
< High School	26% (n=10)
High School diploma	37% (n=14)
Some higher education	37% (n=14)
Annual Income	
$20,000 or less	42% (n=16)
$20,001 to $40,000	39% (n=15)
$40,001 to $80,000	8% (n=3)
Missing data	11% (n=4)

Communication Strategies Used for Assessing Medication-Taking Behavior

We examined 4 dimensions that characterize how providers ask patients about their medications based on: (1) structure, (2) temporality, (3) content of the question(s), and (4) style. We discuss excerpts from 3 different encounters that illustrate these dimensions (Table 2). We then briefly discuss how medication is talked about in the remainder of the encounter.

Structure

We identified 6 different linguistic structures that providers used to ask patients about medications. We first identified whether the "asking" was open or closed ended. We then identified whether the asking was in the form of a question (interrogative) or in the form of a statement (declarative). Table 3 outlines the taxonomy of ways of asking by identifying (1) the type of question and (2) the possible types of response facilitated. Generally speaking, open-ended questions require patients to give information, whereas closed-ended questions require single word answers only, often yes, no or a simple number.

Closed-Ended Questions

Many providers directly asked patients about their medication taking using an interrogative closed-ended form of questioning, such "are you taking," "do you take," and "did you take." Patients often replied with a single word response, yes or no, providing little additional information about their use and understanding of antihypertensives.

In 12 interactions, providers used only declarative and closed-ended questions such as "so you are taking," or "so you took." This strategy is "leading" in that patients may perceive the statement as an assumption of fact, therefore finding it difficult to negate or contradict. This can be seen in example 1 (Table 2). The provider initiates this sequence with a question about having trouble taking medication (line 1), an important aspect of patient-centered communication. However, when the patient replies "no," the provider does not assess whether the patient is actually taking his medication as prescribed. Note that "so you are taking," is made as a statement rather than posed as a question. Although the provider lists the medications, he does not ask the patient to confirm that he had been in fact taking these medications. The patient responds with "yes," simply indicating agreement with the provider. He provides no detailed information about whether he knows which pills are which, or how much or how often he is taking 2 of the medications.

The provider likely leaves this conversation believing that the patient is adherent to his medication due to the "agreement." It is not clear, however, that the patient is accurately or consistently taking his medication based on the information provided in this encounter.

Open-Ended Strategies

In contrast, an interrogative, open-ended question, such as "which medications are you taking" or "how often" or "when," is treated as a request for the patient to provide information. This type of question requires more than a yes/no response from the patient. In all 5 instances where the "which medications" strategy was used, patients replied with multiple word answers, which in turn generated a discussion of how they were taking their medication. In example 2, the patient has controlled BP of 138/70 and the provider assesses how much of each medication the patient was taking, when he was taking it and if he knows when to take the different medications that were prescribed. The patient does not identify names of the medications; however, he is clear about which pills he has been taking. In lines 6 to 9, the provider discovers that the patient has not been taking the second dose of 1 medication. This more complex response allows the provider to evaluate whether the patient is knowledgeable about the medication he is taking and helps her assess if the patient is taking the medication as prescribed.

Using more than 1 communication strategy to ask about medication-taking behavior was also effective in eliciting information from the patient. This is demonstrated in example 3 in which the patient's HTN is clearly out of control (BP 188/114). Having examined the pharmacy refill data accessible through the computerized patient record, the provider sees that the patient has not been refilling the prescriptions, leading to extensive probing of the patient's medication-taking behavior.

This provider begins by asking if the patient is taking his medication (line 1, strategy 1). He continues by asking how many different kinds of medication the patient is taking (line 5, strategy 6), and follows up with questions about how often the medication is being taken (lines 20 and 42; strategy 5) and when (lines 9, 13, and 15; strategy 2). At this point the provider is aware that the patient is not taking his medication as prescribed. Later, he asks directly about medications and follows up with a more declarative statement to confirm the behavior (line 40, strategy 4).

Through the use of multiple strategies, the provider is able to assess the patients' medication-taking behavior, and to identify potential reasons for the lack of BP control, and plan intervention accordingly. After this interchange the provider asks the patient about problems he is having taking medication, explains the potential impact of HTN on the patient's health, makes changes to the medication, providing written instructions to the patient, and arranging for follow-up. Near the end of the appointment, the provider reiterates the medication plan, and is explicit with the patient that no change will be made because "we're not sure exactly that you've been taking your medicines every single day, day in and day out."

Temporality

The temporal nature of the question asked was identified based on the form of the verb used—an infinitive form (-ing) indicating a general behavior, versus a past tense form (did) indicating a single event (see Table 3). In all 3 examples, the providers begin with questions about patients' general, time unlimited behaviors with respect to medication taking, i.e., "are you taking." Only when the provider in example 3 finds that he cannot reconcile the information he is getting from the patient with the information he has on the medical record does he shift to asking about time-specific medication taking, i.e., "did you take it this morning." He thereby gains information about the patient's actual behavior.

Table 2. Examples of Asking About Medications and Associated Findings

Example 1	
1 PR: So, are you having any problems with any of your heart – your blood pressure medications?	*Structure*: Closed and Declarative *Temporality*: Unlimited
2 PT: No.	*Content:* Medical terminology, but reframes using lay terms
3 PR: No? Okay. So, if it ain't broke, we're not going to fix it. Okay. You're still taking the Amlodipine, Isordil. What else? And the Metoprolol. Those are your three blood pressure medicines.	*Style*: Neutral *Patient response:* Single word, yes/no *Blood pressure:* 139/85
4 PT: Yes.	
5 PR: They're okay. And the Metoprolol? Let's see. All those have been – all those have at least a couple of refills on them. And the Metoprolol you're taking half a tab a day?	
6 PT: Yes	
7 PR: Okay	
Example 2	
1 PR: So right now, in the morning, which pills are you taking? Do you know?	*Structure*: Open & Interrogative
2 PT: Well, I take my high blood pressure pills.	*Temporality*: Unlimited
3 PR: Which ones, 'cause you're on a bunch of them now?	*Content:* Lay terminology
4 PT: I take that little square one, as you said.	*Style:* Collaborative
5 PR: Right. Now do you take one dose or two of those little square ones?	*Patient response:* Descriptive of pills and how medications are taken
6 PT: You – it says only one a day. One.	*Blood pressure:* 138/70
7 PR: Right. But you're supposed to take a 20 milligram one and a 40 milligram one. Do you take both of those or do you just take one pill?	
8 PT: I – I take – I just take one square one.	
9 PR: One square one. And how about the round one? Are you taking the round one still?	
10 PT: The round one, yeah. And then I take one aspirin.	
11 PR: And how about at bedtime?	
12 PT: I take the one that's marked for bedtime, I take that.	
Example 3	
1 PR: Are you taking all of your blood pressure medicines?	*Structure*: Mixed closed and open-ended
2 PT: Yeah, I've been taking them.	*Temporality*: mixed unlimited and limited
3 PR: You got them with you?	*Content:* Lay terminology, but reframes using medical terminology
4 PT: No.	*Style*: Not collaborative; confrontational
5 PR: How many different kinds of medicines are you taking?	*Patient response:* Few word responses, attempts to discuss problems with medication-taking
6 PT: About 9.	*Blood pressure:* 118/114
7 PR: Now, almost all of them have not been refilled since August. You had a refill available. Did you know that?	
8 PT: I don't (inaudible)	
9 PR: When was the last time you took your blood pressure medicine?	
10 PT: This morning.	
11 PR: This morning? You didn't have any period when you weren't taking it?	
12 PT: (inaudible) sick probably, I didn't, sleeping at night you know.	
13 PR: So did you take it yesterday?	
14 PT: In the evening.	
15 PR: How about the day before?	
16 PT: I don't know because I was sick, I was really sick (inaudible). I mean I didn't know (inaudible)	
17 PR: Because it should have only, if you got it filled in August and it was a 90 day supply, August, September, October, November. It should have been out two months ago, if you were taking it every day.	
18 PT: No.	
19 PR: Do you miss it pretty often?	
20 PT: No. I have a feeling, I got a feeling (inaudible)	
A little later on, after the provider takes the blood pressure and notes that it is high, he continues:	
40 PR: And you took Lisinopril, Hydralazine, Felodipine, did you take all three of those?	
41 PT: (inaudible)	
42 PR: How many times a day are you taking your Hydralazine?	
43 PT: One.	
44 PR: Okay, that's supposed to be three times a day.	
45 PT: Okay.	
46 PR: Now, that one you have to take three times a day. You know, we've tried doing it with other medicines that you didn't have to take so often in the past, but they haven't worked so that one you have to take three times a day. Okay. That may be part of the problem right there.	

PR, Provider; PT, Patient.

Content

Providers differed in the extent to which they used medical jargon versus lay terminology in discussing medications. In example 1, the provider discusses medications by appearance rather than name ("the little square pills"), thereby using lay, rather than biomedically specific language. He does mention the number of milligrams but then reverts to talking about the pills in a lay manner. In contrast, the providers in examples 2 and 3 only use the generic names of medications, names that may be unfamiliar to the patient and difficult to recall. These providers do discuss dose and frequency according to the number of pills and times per day, which may be easier for patients than discussions of milligrams.

Table 3. Ways of Asking About Medication Taking

Strategy	Question Asked	Type of Question	Temporality	Possibility for Response
Closed-ended questions				
1	Are you taking X? Do you take X?	Interrogative Yes/no question	Temporally unlimited	Requires yes/no response. Little opportunity for patient to discuss medication-taking behavior
2	Did you take X?	Interrogative Yes/no question	Temporally limited	Requires yes/no response. Little opportunity for patient to discuss medication-taking behavior
3	So, you are taking X, Y and Z medications. 1. rising intonation ("?") 2. falling intonation (".")	Declarative 1. positive assumption/neutral expectation 2. positive assumption/positive expectation Closed-ended	Temporally unlimited	1. Patient perceives expectation of adherence therefore is difficult to negate (i.e., well, no actually I'm not). 2. Patient perceives provider making statement of 'fact' and therefore even harder to negate
4	So you took X [this morning]	Declarative	Temporally limited	Requires yes/no response. Positive assumption presented, therefore difficult to negate
Open-ended questions				
5	Which medications are you taking for your blood pressure?	Interrogative	Temporally unlimited	Requires patient to provide information, thereby allowing provider to assess patients' knowledge of medication
6	How often are you taking your X medication?	Interrogative Restricted to a number	Temporally unlimited	Requires patient to tell how often, but does not assess patients' knowledge of medication types

Temporality: unlimited, assessing usual behavior;.
Limited: assessing behavior of a specific kind and time.

Style of Questioning

Provider styles varied according to how collaborative the interaction was. Collaborativeness was defined as the ways in which the provider followed up on patients' utterances and concerns and subsequently focused on the patients' communication. In example 2, the provider asks if the patient knows which medications he is taking and follows up by asking for specifics, and provides a rationale for asking in line 3. Further the provider pays close attention to the responses of the patient and follows-up on the patients' cues to talk about medication by color and shape. As the encounter continues, the provider asks the patient to bring his pills to a follow-up appointment with a nurse so that she can further monitor his medication taking and review his medications at that time. The provider elicits the patient's involvement in assessing his medication-taking behavior and provides a more collaborative interaction.

In example 1, the interaction is less collaborative. The provider begins by asking about problems the patient might be having taking medications. This strategy could build collaboration, but the topic is quickly closed by the provider's use of the declarative form of questioning.

Example 3 may be construed as least collaborative. In line 7, the provider challenges the patient's statement that he has been taking his medications by saying that he sees the medication has not been refilled. Importantly, the patient twice tries to give information about being sick, and potentially how that may have interfered with his medication taking. However, the provider does not address this concern and continues questioning the patient about his medications. By not attending to the patient's concern here, the provider forgoes an opportunity to see if the patient's apparent nonadherence is intentional or non-intentional. If the provider had followed-up by asking about problems taking medications, he would have gained the needed information for good prescribing, and may have further been able to assist the patient to take his medication in the future.

Following Up

Even when providers did ask in some format about medication taking, they often did not follow-up by seeking information about barriers to taking medication as prescribed. In a few instances, providers asked about side effects patients were having or problems with paying for medications. None of the providers asked questions about patients' beliefs about medications or about their understanding of HTN. Rather the tendency was for providers to be directive, instructing patients, in some instances multiple times, about the importance of taking antihypertensives, providing rationale for changing medications and giving instructions regarding how to take them.

Asking About Medication Taking and BP Control

Table 4 shows how often patients who had controlled versus uncontrolled BP were asked about medication taking. In 9/15 (60%) encounters in which the provider did not ask the patient about his medication taking, the patient's BP was uncontrolled (i.e., <140/90) that day. Although providers were somewhat more likely to ask about medications when BP was uncontrolled (67%), 33% of those with uncontrolled BP were never asked about medication taking.

Table 5 shows how often providers used different asking strategies with patients who had controlled versus uncontrolled BP. When BP was uncontrolled, providers used declarative, closed-ended statements to ask about medications 55% of the time and interrogatives only 45% of the time. Within the interrogatives, providers asked open-ended questions regardless of the level of BP control. Closed-ended interrogatives (i.e., "did you take" or "are you taking") were used as a primary tool only when BP was uncontrolled.

DISCUSSION

Patients often do not adhere to prescribed antihypertensive medications. We have described how one aspect of provider

Table 4. Number of Patients Asked/Not Asked About Medications and Whether or Not BP Was Controlled ($n=38$)

	Total	Controlled	Uncontrolled
Total		11 (29%)	27 (71%)
Asked	23 (61%)	5 (45%)	18 (67%)
Not asked	15 (39%)	6 (55%)	9 (33%)

communication—asking about patients' medication-taking behavior—may influence the kinds of information patients provide. Almost 40% of those with a diagnosis of HTN were not asked at all about how they took their medication and 1/3 of these patients had uncontrolled HTN. Even when patients' HTN was uncontrolled, providers often did not directly ask about medications, discussing other aspects of HTN management instead. As a result, providers were less likely to be able to determine whether uncontrolled HTN was due to ineffective medication or poor adherence to potentially effective medications.

When providers did ask about medication taking, they sometimes used approaches that were not optimal for obtaining detailed information about adherence. We identified 4 dimensions of provider inquiry based on the structure, temporality, content and style of asking. The linguistic structure of questions can either facilitate or inhibit the amount and kind of information patients offer the provider. More closed-ended and declarative questions may make it difficult for patients to supply extensive information about their medications. Although we recognize that patients may not always tell the truth about their medication-taking behavior, we believe that some proportion of agreement with the provider's description of medication dosage and frequency may be due to the form the question takes and lack of a collaborative communication style. Providers may be able to better assess adherence by asking interrogative, open-ended questions, using strings or sequences of questions, and collaborating by following up on patients' concerns. Asking patients to report on time-limited behaviors may further supply information to providers about what patients are actually doing with regard to medication taking. These strategies may lead to a more in-depth discussion of patients' beliefs about medications, problems they are encountering taking medications, and ultimately to better adherence.

Patients' responses to questions may be affected by other aspects of provider communication. When providers ask about medications by their pharmaceutical names, patients may get confused. In contrast, study providers who asked about medications by describing the color and size of each pill elicited detail from the patients about how they took their medications. This alternative practice may improve the patient's ability to accurately report and the provider's ability to accurately assess adherence. This may be especially important for patients with low health literacy.[20–22] In addition, provider communication styles, such as the challenging style in example 3, may be detrimental to creating a therapeutic alliance with patients.[23,24]

Table 5. Ways of Being Asked About Medications and Whether or Not BP Was Controlled.

Type of Question	Controlled	Uncontrolled
Open-ended, interrogative (i.e., Which, how often)	4 (80%)	3 (17%)
Closed-ended, interrogative (i.e., Are you, did you)	0 (0%)	5 (28%)
Closed-ended, declarative (i.e., So you are)	1 (20%)	10 (55%)
Total	5 (100%)	18 (100%)

Number Indicates when Question Type was used as Sole or Primary Type ($n=23$).

The strategies described in this paper are a point of departure for understanding one aspect of HTN care, the prescription and taking of antihypertensives. As noted by Steele et al.,[17] direct and information intensive approaches may be most effective in detecting adherence, however, we found also that consideration of the linguistic structure of the questioning as well as its style and content may impact upon how patients respond to providers' questions. Patient factors, such as level of education and patient beliefs about medication, may also contribute to problems communicating about medication taking.[25] Analyses of patient-provider communication in the future will shed more light on the ways in which providers' communication strategies facilitate or hinder patients' adherence to medications as prescribed.

Many of the practices we observed reflect a provider-centered and medically centered model of disease management. Effective communication has been described as one in which a relationship-centered provider jointly partners with patients to make decisions and explore their perspectives.[26,27] Discussing patient illness representations has been shown to improve hypertensive patients' viewpoints on adherence.[12] Providers who focus solely on informing the patient of the medication regimen to be followed rely on a more profession-centered style of communication. The focus in many of the encounters we examined was on information *transfer*—giving information about medications to the patient—rather than information *exchange*, in which the provider and patient have a 2-way exchange and collaboratively discuss patient perspectives on medications.[28] Providers' focus on informing rather than on assessing patients' medication-taking behavior, or the extent to which the patient buys into this "contract," may lead to an inaccurate decision that medications prescribed are ineffective, when in fact the patient is not taking the medication as prescribed.

This study has several important limitations. We are reporting on the behavior of only 9 providers, all of who were practicing in Veterans' Affairs clinics. As such we do not know if the interactions we observed generalize to all physicians. In addition, all of the patients were men and communication patterns may differ with women. It is possible that the behaviors we observed were the result of a Hawthorne effect resulting from the fact that providers and patients were aware of the presence of the tape recorder and our interest in communication. We nonetheless observed a wide range of interaction styles and communication behaviors, suggesting that neither providers nor patients were strongly inhibited. Similarly, if there was a selection bias in favor of providers with better communication skills, our finding that these physicians often did not elicit information from patients is actually a conservative test of this phenomenon.

This study also developed a method for describing how providers discuss adherence. The taxonomy may be useful in assessing provider-patient communication about a variety of health behaviors, including but not limited to medication

adherence. It could also prove useful in medical education and in further research about patients' self-management of chronic illness.

Taking medication as prescribed is most often discussed as compliance or adherence. Conrad,[29] however, implores us to reconsider this issue as one in which patients integrate taking medication into their daily lives. As the IOM[30] also notes, we must move toward a model in which patients' needs and concerns become the focus of the encounter rather than simply a transfer of information from provider to patient. Considering the social context of health and illness in which patients take medication makes sense and ultimately creates the conditions for communication and dialogue that will lead to collaboration and change.

This study was supported by the Department of Veterans Affairs, Health Services Research and Development grant # TRH 01-038. Dr. Kressin is supported by Department of Veterans Affairs, Health Services Research and Development, RCS # 02-066-1. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. We would also like to thank Sue Eisen, Donald Miller, and Rani Elwy for comments on an earlier draft of this paper.

REFERENCES

1. **Hajjar I, Kotchen TA.** Trends in prevalence, awareness, treatment and control of hypertension in the United States, 1988–2000. JAMA. 2003;290:199–206.
2. **Joint National Committee,Chobanian A, Bakris G, et al.** Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension. 2003;42:1206–52.
3. **Amery A, Birkenhager W, Brixko P, et al.** Mortality and morbidity results from the European working party on high blood pressure in the elderly trial. Lancet. 1985;1:1349–54.
4. **SHEP Cooperative Research Group.** Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP). JAMA. 1991;265:3255–64.
5. **Jamerson K, DeQuattro V.** The impact of ethnicity on response to antihypertensive therapy. Am J Med. 1996;101:22S–32S.
6. **Winickoff R.** The persistent problem of poor blood pressure control. Arch Intern Med. 1987;147:1393–6.
7. **NHLBI Working Group.** Management of patient compliance in the treatment of hypertension: report of the NHLBI working group. Hypertension. 1982;4:415–23.
8. **Sackett DL, Snow JC.** The magnitude of compliance and noncompliance. In: Haynes RB, Taylor DW, Sackett DL, eds. Compliance in Health Care. Baltimore: Johns Hopkins University Press; 1979:11–22.
9. **Clark LT.** Improving compliance and increasing control of hypertension: needs of special hypertensive populations. Am Heart J. 1991;121 (suppl):664.
10. **Gallup G, Cotugno H.** Preferences and practices of Americans and their physicians in antihypertensive therapy. Am J Med. 1986;81: 20–4.
11. **Horne R, Clatworthy J, Polmear A, Weinman J.** Do hypertensive patients' beliefs about their illness and treatment influence medication adherence and quality of life? J Hum Hypertension. 2001;15(suppl 1): S65–S68.
12. **Theunissen N, De RIdder D, Bensing J, Rutten G.** Manipulation of patient-provider interaction: discussing illness representations or action plans concerning adherence. Patient Educ Couns. 2003;51: 247–58.
13. **Ogedegbe G, Harrison M, Robbins L, Mancuso C, Allegrant J.** Barriers and facilitators of medication adherence in hypertensive African Americans: a qualitative study. Ethnicity Disease. 2004;14:3–12.
14. **Stevenson F, Barry C, Britten N, Barber N, Bradley C.** Doctor-patient communication about drugs: the evidence for shared decision making. Soc Sci Med. 2000;50:829–40.
15. **Kjellgren K, Svensson S, Ahlner J, Saljo R.** Antihypertensive medication in clinical encounters. Int J Cardiol. 1998;64:161–9.
16. **Ockene J, Kristeller J, Goldberg R.** Increasing the efficacy of physician-delivered smoking interventions: a randomized clinical trial. J Gen Intern Med. 1991;6:1–8.
17. **Steele D, Jackson T, Gutmann M.** Have you been taking your pills? The Adherence-monitoring sequence in the medical interview. J Fam Pract. 1990;30:294–9.
18. **Morris G, Chenail R.** The Talk of the Clinic: Explorations in the Analysis of Medical and Therapeutic Discourse. Hillsdale, NJ: Lawrence Erlbaum Associates; 1995.
19. **Schiffrin D.** Approaches to Discourse. Cambridge, MA: Blackwell Publishers; 1994.
20. **Williams M, Baker D, Parker R, Nurss J.** 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.
21. **Baker D, Parker R, Williams M, et al.** The health care experience of patients with low literacy. Arch Fam Med. 1996;5:329–34.
22. **Kalichman S, Ramachandran B, Catz S.** Adherence to combination antiretroviral therapies in HIV patients of low health literacy. J Gen Intern Med. 1999;14:267–73.
23. **Mead M, Bower P.** Patient-centredness: a conceptual framework and review of the empirical literature. Soc Sci Med. 2000;51:1087–110.
24. **Kjellgren K, Ahlenr J, Saljo R.** Taking antihypertensive medication - controlling or co-operating with patients. Int J Cardiol. 1995;47: 257–68.
25. **Ong L, De Haes J, Hoos A, Lammes F.** Doctor-patient communication: a review of the literature. Soc Sci Med. 1995;40:903–18.
26. **Braddock C, Edwards K, Hasenberg N, Laidley T, Levinson W.** Informed decision making in outpatient practice: time to get back to basics. JAMA. 1999;282:2313–20.
27. **Frankel R, T S.** Getting the most out of the clinical encounter; The four habits model. Permanente J. 1999;3:1–8.
28. **Lee R, Garvin T.** Moving from information transfer to information exchange in health and health care. Soc Sci Med. 2003;56:449–64.
29. **Conrad P.** The meaning of medications: another look at compliance. Soc Sci Med. 1985;20:29–37.
30. **Institute of Medicine.** Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy of Sciences; 2001.

4 Burr ML, Ashfield-Watt PAL, Dunstan FDJ, Fehily AM, Breay P, Ashton T, et al. Lack of benefit of dietary advice to men with angina: results of a controlled trial. *Eur J Clin Nutr* 2003;57:193-200.
5 Leaf A, Kang JX, Xiao Y-F, Billman GE. Clinical prevention of sudden cardiac death by n-3 polyunsaturated fatty acids and mechanism of prevention of arrhythmias by n-3 fish oils. *Circulation* 2003;107:2646-52.
6 Raitt MH, Connor WE, Morris C, Kron J, Halperin B, Chugh SS, et al. Fish oil supplementation and risk of ventricular tachycardia and ventricular fibrillation in patients with implantable defibrillators—randomized controlled trial. *JAMA* 2005;293:2884-91.
7 Lim W, Gammack J, Van Niekerk J, Dangour A. Omega 3 fatty acid for the prevention of dementia. *Cochrane Database Syst Rev* 2006;(1):CD005379.
8 Holman RT. The slow discovery of the importance of omega3 essential fatty acids in human health. *J Nutr* 1998;128:427-33S.
9 Scientific Advisory Committee on Nutrition, Committee on Toxicity. *Advice on fish consumption: benefits and risks*. London, Stationery Office, 2004. www.food.gov.uk/multimedia/pdfs/fishreport2004full.pdf (accessed 9 Feb 2006).
10 Newson RB, Shaheen SO, Henderson AJ, Emmett PM, Sherriff A, Calder PC. Umbilical cord and maternal blood red cell fatty acids and early childhood wheezing and eczema. *J Allergy Clin Immunol* 2004;114:531-7.
11 Myers RA, Worm B. Rapid worldwide depletion of predatory fish communities. *Nature* 2003;423:280-3.

doi 10.1136/bmj.38798.680185.47

Increasing the number of medical students from under-represented minorities

Innovative method needs to be replicated elsewhere

Better health care and a reduction in health disparities occur when the physician workforce is diverse in background, ethnicity, culture, and race.[1 2] Unfortunately, there are not enough students from economically and educationally deprived backgrounds in the pipeline to achieve a sufficiently diverse physician workforce. A paper by Greenhalgh and colleagues in this issue describes an innovative and thoughtful approach of reaching and engaging these students.[3] The project steering group used "partnering" schools and teachers in inner London to identify students from deprived backgrounds who had motivation and ability in the sciences. Teachers reaffirmed the problems the students needed to overcome to be successful in gaining a medical education: lack of self determination, lack of confidence, and lack of information. The group used the summer school programme to target these deficiencies and ignite in the students a sense of purpose and confidence, which was the first step in engaging them in a career in medicine. Interacting with medical students of similar ages and experience increased self confidence and the perception that a career in medicine is possible. Focusing on the deficiencies of the environment rather than on those of the student was also key to providing the students with information about careers in medicine. Collaboration of pupils, teachers, and parents makes all of them stakeholders with a vested interest in the outcome. The action research approach, which uses cycles of observation, data gathering, and reflection, is ideal for this type of project.

In the United States attaining a diverse medical workforce raises similar issues. Additionally, men from ethnic minorities are the least represented in medical education. The educational attitudes of many students from under-represented groups have evolved to a point where education has no value at all. Despite this, some programmes have been effective in recruiting and retaining students from ethnic minorities in medical school. The Early Medical School Selection Program (EMSSP) at Boston University School of Medicine has a successful track record of 25 years of recruiting second year college students from 11 institutions with predominately black and Mexican-American students as part of a consortium. During the summers after their second and third years in college the students attend Boston University Summer School and take undergraduate courses for credit towards their bachelor's degrees. They then take their senior undergraduate year at Boston University, taking courses from the medical school curriculum. If they pass these courses they are promoted to the medical school. Ninety per cent of the students get through to the third year of medical school without academic difficulty, and 85% pass step 1 of the United States Medical Licensing Examination on the first try.[4] This is remarkable because many of these students had not previously performed well on standardised tests. The success of this programme relies on the academic and personal support the students receive during their premedical and medical education. Despite the obstacles the students face they are able to do the work and achieve careers in medicine.

Research p 762

The challenges of lack of confidence, lack of information and lack of self determination that students encounter in Greenhalgh et al's experience and ours are are important. However, if we are going to successfully address health disparities, programmes like these need to be replicated and applied elsewhere. Also, the authors' plan to follow up this cohort will be essential to assess the long term impact of this programme. Our societies are becoming more multicultural and diverse. Many of the students in this study were either first generation immigrants or the children of immigrants from 19 countries in Africa, the Caribbean, Asia, and Europe. Understanding the culture, religion, and customs of others is imperative if we are going to live together and reduce health disparities.

Phyllis L Carr *associate dean for student affairs*
(plcarr@bu.edu)

Kenneth C Edelin *associate dean for students and minority affairs*

Office of Student Affairs, Boston University School of Medicine, Boston, MA 02118, USA

Competing interests: None declared.

1 Komaromy M, Grumbach K, Drake M, Vrazizan K, Lurie N. The role of black and Hispanic physicians in providing health care for underserved populations. *N Engl J Med* 1996;334:305-10.
2 Moy E, Bartman BA. Physicians, race and care of minority and medically indigent patients. *JAMA* 1995;273:1515-20.
3 Greenhalgh T, Russell J, Dunkley L, Boynton PM, Lefford F, Chopra N. "We were treated like adults"—development of a pre-medicine summer school for 16 year olds from deprived socio-economic backgrounds: action research study. *BMJ* 2006;332:762-6.
4 Edelin KC, Ugbolue A. Evaluation of an early medical school selection program for underrepresented minority students. *Acad Med* 2001;76:1056-9.

BMJ 2006;332:740

AIDS and Behavior, Vol. 10, No. 2, March 2006 (© 2006)
DOI: 10.1007/s10461-005-9046-6

Association of Violence Victimization with Inconsistent Condom Use in HIV-Infected Persons

Cynthia H. Chuang,[1,5] Jane M. Liebschutz,[2] Nicholas J. Horton,[3] and Jeffrey H. Samet[2,4]

Published online: Jan. 26, 2006

The association of violence victimization with current condom use in HIV-infected persons was examined in this cross-sectional study. The HIV—Alcohol Longitudinal Cohort (HIV-ALC) recruited HIV-infected participants with a history of alcohol problems. Interviews assessed violence histories and current sexual behaviors. Of the 349 participants (79% men), 38% reported inconsistent condom use and 80% reported a violence history. Lifetime sexual violence was reported by 40% and lifetime physical violence (without sexual violence) by 40%. Participants reporting lifetime sexual violence had greater odds of inconsistent condom use than participants without any history of violence. A history of childhood sexual violence was also associated with greater odds of inconsistent condom use than participants without a history of childhood sexual violence. A history of sexual violence may in part explain HIV-infected persons' greater risk for transmitting HIV through high-risk sexual behaviors.

KEY WORDS: HIV; violence; risk behavior; condom use.

INTRODUCTION

Sexual contact is currently the most common mode of HIV transmission in the U.S. and worldwide (Centers for Disease Control and Prevention [CDC], 2001). Although many HIV-infected persons substantially change their sexual practices after HIV diagnosis to reduce transmission risk (Metsch *et al.*, 1998), others continue to engage in unsafe sexual contact putting their partners at risk for disease (Erbelding *et al.*, 2000; Kalichman *et al.*, 2002; Kwiatkowski and Booth, 1998; Marks *et al.*, 1999). Prevention efforts have largely targeted HIV-negative populations deemed at high risk for acquiring HIV. Although these efforts are successful at reducing risky sexual behaviors (Kamb *et al.*, 1998; National Institute of Mental Health [NIMH] Multisite HIV Prevention Trial Group, 1998) and should continue, targeting HIV-infected populations about the continued risks of transmitting disease is increasingly being emphasized (CDC, 2003). Recognizing predictors of high-risk sexual practices among the HIV-infected is a step toward understanding the chain of events that may lead to putting sexual partners at risk for infection. This study investigated the association of a history of physical or sexual violence with inconsistent condom use in an HIV-infected cohort with a history of alcohol problems.

The association of a personal history of violence with high-risk sexual behavior among populations at increased risk for HIV infection has been described

[1]Division of General Internal Medicine, Department of Medicine, Penn State College of Medicine, Milton S. Hershey Medical Center, Hershey, Pennsylvania.
[2]Department of Medicine, Boston University School of Medicine, Section of General Internal Medicine, Clinical Addiction Research and Education (CARE) Unit, Boston Medical Center, Boston, Massachusetts.
[3]Department of Mathematics, Smith College, Northampton, Massachusetts.
[4]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts.
[5]Correspondence should be directed to Cynthia H. Chuang, MD, MSc, Penn State Milton S. Hershey Medical Center, Division of General Internal Medicine, HU15, 500 University Drive, P.O. Box 850, Hershey, Pennsylvania 17033-0850; e-mail: cchuang@psu.edu.

(Cohen *et al.*, 2000; DiIorio *et al.*, 2002; El-Bassel *et al.*, 2001; Hamburger *et al.*, 2004; Hillis *et al.*, 2001; Lenderking *et al.*, 1997; Mullings *et al.*, 2000; NIMH HIV Prevention Trial Group, 2001; Parillo *et al.*, 2001; Paul *et al.*, 2001). These studies show that violence is associated with sexual risk behaviors that increase the risk of acquiring STDs, including HIV. Extending this finding to HIV-infected populations, one could hypothesize that a history of violence is associated with sexual risk taking and transmission of the virus even after HIV diagnosis and knowledge of infectivity. In a cross-sectional analysis of HIV-infected men who have sex with men (MSM), O'Leary *et al.* (2003) report a significant association between a history of childhood sexual violence and unprotected anal sex in the previous 90 days with partners of HIV-negative or unknown serostatus.

Physical and sexual violence is very common; it is estimated that 25% of U.S. women have been victimized during their lifetime, with 1.5% having been abused within the past year (Tjaden and Thoennes, 2000). However, the prevalence of violence among the HIV-infected is even more alarming (Bedimo *et al.*, 1997; Cohen *et al.*, 2000). Sixty-six percent of HIV-infected women in one cohort reported a history of violence, with 21% having been victimized in the past year, and 31% reporting a history of sexual violence prior to age 18 (Cohen *et al.*, 2000). Another study of HIV-infected women found that 32% had a history of sexual violence during their lifetime (Bedimo *et al.*, 1997). The study of physical and sexual violence among HIV-infected men has been limited.

Previous research has described high prevalence of inconsistent condom use among HIV-infected men and women of all transmission risk categories with a lifetime history of alcohol problems (Ehrenstein *et al.*, 2004). The current study examined whether inconsistent condom use was associated with a history of physical or sexual violence. It is known that sexual violence is associated with sexual risk behavior among populations at risk for HIV; this study examines whether this association is also true among individuals aware of their HIV infection.

METHODS

Participant Recruitment

The HIV—Alcohol Longitudinal Cohort (HIV-ALC) study recruited HIV-infected individuals with a history of alcohol problems with the primary aim of evaluating the effect of alcohol use on HIV progression (Samet *et al.*, 2003). The current study is a cross-sectional analysis using the baseline data of the 349 participants of the HIV-ALC study. Patients were recruited principally from the Boston Medical Center HIV Diagnostic Evaluation Unit (Samet *et al.*, 1995), a weekly clinic for engaging HIV-infected persons into medical care. Participants were also recruited from other sites: the Beth Israel Deaconess Medical Center, a respite facility for homeless persons, a methadone clinic, Boston Medical Center's primary care practices, referrals by friends, and through posted flyers at homeless shelters and HIV/AIDS social service agencies in the Boston area.

Participants were eligible for enrollment if they had confirmed HIV infection, a lifetime history of alcohol problems (defined as $\geq$ 2 positive responses to the CAGE questionnaire (Ewing, 1984), and were $\geq$18 years of age. Those patients recruited from the Boston Medical Center HIV Diagnostic Evaluation Unit who did not meet CAGE criteria were eligible if one of two attending physicians made a specific diagnosis of alcohol abuse or dependence. Other inclusion criteria were fluency in English or Spanish, Mini-Mental State Examination (Folstein *et al.*, 1975) score $\geq$21, and likelihood of residence in the Boston area for the next 2 years. Recruitment began in June 1997 and ended in July 2001.

Data Collection and Survey Instrument

All participants were interviewed in-person by a research associate in a private room using a standardized instrument. The instrument included items on demographics, exposure to interpersonal violence, alcohol and drug use, depressive symptoms, HIV transmission risk category, and sexual behaviors. The questions used to assess violence histories were adapted from a previous study designed to describe interpersonal violence among persons with a history of substance abuse (Liebschutz *et al.*, 2002). Alcohol and drug severity and consumption were measured using both quantity and frequency questions assessing the prior 30 days and the Addiction Severity Index, which has documented reliability and validity (McLellan *et al.*, 1985). Depressive symptoms were measured by the 20-item Centers from Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977). This scale identifies depressive symptoms over the past week. Items

assessing sexual risk behaviors over the past 6 months were derived from the Risk Assessment Battery (Navaline *et al.*, 1994). Interviews were conducted in either English or Spanish. For the measures used in this analysis, the Spanish version was created by translating the questions to Spanish and then back translated to check for accuracy.

Definition of Violence Variables

To assess for physical violence, the participants were asked, "Have you ever been physically abused or assaulted (for example: kicked, hit, choked, shot, stabbed, burned, or held at gunpoint)?" To assess for sexual violence, the participants 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, participants were asked their age at the time of the first episode.

Two violence variables were created for these analyses: lifetime violence and childhood sexual violence. The lifetime violence variable was created to define violence experienced at any age. A participant's lifetime violence variable was then categorized into one of three groups: no violence history, a sexual violence history (with or without physical violence), or a physical violence history alone (without sexual violence). Most persons reporting sexual violence also reported physical violence, so it was not possible to truly isolate these two types of violence. The three categories were chosen for the lifetime violence variable because they were felt to best describe a person's violence experience. The childhood sexual violence variable was dichotomous, defined as the report of sexual violence occurring prior to age 13.

Definition of Inconsistent Condom Use

Participants were asked about their current sexual practices. Inconsistent condom use was defined as not using condoms at all sexual encounters in the past 6 months. Consistent condom use was defined as using condoms at all sexual encounters or not having sex in the past 6 months. During the interviews, sex was defined as any vaginal intercourse, anal intercourse or oral sex.

Definition of Covariates

Covariates that have previously been shown to be associated with sexual risk taking or had clinical face validity were chosen. Covariates used were sex, age, education (±high school graduation), marital status (married vs. not married), homelessness, race/ethnicity (Black, White, other), CES-D score coded as a continuous variable ranging from 0 to 60 with higher scores indicating more depressive symptoms (Radloff, 1977), HIV transmission risk category (heterosexual, injection drug use, MSM), alcohol use in the past 30 days, any cocaine use in the past 30 days, any heroin use in the past 30 days, and current antiretroviral use. Homelessness was defined as at least one night in a shelter or on the street in the past 6 months. Alcohol use in the past 30 days was categorized as hazardous, moderate, or abstinent. These categories were derived from the NIAAA definition for hazardous use based on >14 drinks/week for men and >7 drinks/week for women, or >3 drinks on one occasion for men and >2 drinks on one occasion for women. Moderate alcohol use was defined as any drinking less than hazardous.

Statistical Analysis

Using chi-square tests for categorical variables and *t* tests for continuous variables, characteristics of participants currently engaging in inconsistent condom use were compared with participants without inconsistent condom use or not sexually active in the past 6 months. Unadjusted and adjusted logistic regression models were used to analyze the relationship between the victimization variables (lifetime violence and childhood sexual violence) and inconsistent condom use. Adjusted logistic regression models controlled for the following covariates: sex, age, education, marital status, homelessness, race/ethnicity, CES-D score, HIV transmission risk category, alcohol use in the past 30 days, any cocaine use in the past 30 days, any heroin use in the past 30 days, and current antiretroviral use. For all analyses, two-tailed tests were performed using $p < .05$ as criterion for statistical significance.

RESULTS

The HIV-ALC cohort ($N = 349$) was recruited from the following locations: 56% from the Boston Medical Center HIV Diagnostic Evaluation Unit;

Table I. Characteristics of HIV-Infected Persons with a History of Alcohol Problems Stratified by Consistency of Condom Use ($N = 349$)

Characteristic	Inconsistent condom use[a], no. (%)	Consistent condom use or no sex[b], no. (%)	Test statistic[c]
Male	98 (74)	178 (82)	3.01
Age (years), mean (*SD*)	40.1 (7.1)	40.9 (7.5)	0.98
Graduated high school	90 (68)	120 (55)	5.68*
Married	10 (8)	17 (8)	0.008
Homeless	38 (29)	63 (29)	0.002
Race/ethnicity			3.94
Black	59 (45)	95 (44)	
White	50 (38)	66 (30)	
Other	23 (17)	56 (26)	
CES-D, mean score (*SD*)	23.5 (13.1)	21.8 (13.0)	−1.20
HIV transmission risk category			1.92
Heterosexual	30 (23)	48 (22)	
Injection drug use	72 (55)	133 (61)	
Men who have sex with men (MSM)	29 (22)	36 (17)	
Alcohol use in past 30 days			7.82*
Abstinent	64 (48)	137 (63)	
Moderate	15 (11)	22 (10)	
Hazardous	53 (40)	58 (27)	
Any cocaine use in past 30 days	48 (36)	36 (17)	17.56**
Any heroin use in past 30 days	21 (16)	16 (7)	6.31*
Current antiretroviral use	72 (55)	133 (61)	1.54

[a] $n = 132$.
[b] $n = 217$.
[c] Test statistics are expressed as t scores for continuous variables and χ^2 for categorical variables.
*$p < .05$.**$p < .001$

16% from posted flyers; 13% from Boston Medical Center's primary care practices; 5% from a respite facility for homeless persons; 4% from a methadone clinic; 4% from friend referrals; and 2% from the Beth Israel Deaconess Medical Center. Most study participants [315/349 (90%)] met the eligibility criteria of at least two out of four positive responses to the CAGE questionnaire (Ewing, 1984); the remainder qualified on the basis of clinical assessment [34/349 (10%)].

Characteristics of the cohort are shown in Table I. The majority of the cohort was male and most participants described injection drug use as their risk factor for HIV transmission. The participants are compared by the condom use variable in Table I. Participants with inconsistent condom use were significantly more likely to have graduated high school and to have used alcohol, cocaine, or heroin in the past 30 days.

Eighty percent of the participants had experienced either physical and/or sexual violence at some point in their lives (40% physical violence only and 40% sexual violence with or without physical violence). Both women and men reported high prevalence of any lifetime history of violence (88% and 79%, respectively), but women were more likely to report a history of sexual violence than men (73% vs. 32%), $\chi^2(df = 1) = 40.2$, $p < .001$. Childhood sexual violence (prior to age 13) was reported by 26% of the cohort, with women more likely to have experienced childhood sexual violence than men (39% vs. 22%), $\chi^2(df = 1) = 8.5$, $p = .001$. Among the men, MSM and heterosexual men were both equally likely to report lifetime violence (82% and 77%, respectively), however MSM were more likely to have experienced sexual violence compared with heterosexual men (57% vs. 24%), $\chi^2(df = 1) = 25.3$, $p < .01$. MSM were also more likely to have experienced childhood sexual violence than heterosexual men (34% vs. 18%), $\chi^2(df = 1) = 6.90$, $p < .01$.

Separate unadjusted models examined the relationship of both violence variables with inconsistent condom use. In the unadjusted lifetime violence model, sexual violence was significantly associated with inconsistent condom use compared with those with no history of violence (OR = 2.42, 95% CI 1.29–4.53). Lifetime physical violence was not significantly associated with inconsistent condom use in the unadjusted model. Childhood sexual violence was significantly associated with inconsistent condom use

compared with those without a history of childhood sexual violence (OR 2.14, 95% CI 1.31–3.51), in the unadjusted model.

Logistic regression analyses were performed to model the effect of violence victimization on inconsistent condom use, controlling for covariates. In the lifetime violence model, the magnitude of the association of sexual violence with inconsistent condom use remained significant after controlling for covariates adjusted (OR = 2.88, 95% CI 1.39–5.96). Lifetime physical violence was not significantly associated with inconsistent condom use adjusted (OR = 1.39, 95% CI 0.69–2.79). Childhood sexual violence remained significantly associated with inconsistent condom use adjusted (OR = 2.25, 95% CI 1.31–3.89), in the multivariable analysis. Sex, age, race/ethnicity, education, marital status, homelessness, CES-D score, HIV transmission risk category, current alcohol, heroin, and antiretroviral use were not significant predictors of inconsistent condom use in either of the multivariable violence models. However, cocaine use in the past 30 days (adjusted OR = .54, 95% CI 1.28–5.03), lifetime violence (adjusted OR = 2.69, 95% CI 1.36–5.29), and childhood sexual violence predicted inconsistent condom use in both multivariable models.

DISCUSSION

Sexual violence, occurring either during childhood or at any age, was found to be significantly associated with currently engaging in inconsistent condom use in this cohort of HIV-infected persons with alcohol problems. Several findings in our study deserve emphasis. First, almost 40% of this HIV-infected cohort reported engaging in inconsistent condom use, putting their sexual partners at risk for infection. Secondly, the prevalence of lifetime violence in this cohort of HIV-infected persons was striking at 80%; this is higher than previously reported prevalences of 66–68% among HIV-infected women (Bedimo *et al.*, 1997; Cohen *et al.*, 2000). This finding may be due to the inclusion criteria of a history of alcohol problems in our sample, which has been associated with interpersonal violence (Jasinski *et al.*, 2000; Simpson and Miller, 2002). The finding of increased sexual risk taking associated with past sexual violence further defines characteristics of HIV-infected persons who may be at higher risk for transmitting the disease to their sexual contacts.

The association of violence with high-risk sexual behavior has been previously reported in one study of HIV-infected MSM (O'Leary *et al.*, 2003). This current study adds to existing literature in several ways. First, this cohort comprises HIV-infected men and women of all transmission risk categories, allowing study of the association of violence with current sexual behavior in a more heterogeneous sample. Second, it is relevant that these results were found in a cohort with a history of alcohol problems, because alcohol has been shown to increase sexual risk behavior and may interfere with efforts to improve inconsistent sexual behavior (NIMH Multisite HIV Prevention Trial Group, 2002). Third, the effect of various types of violence was studied. The outcome of inconsistent condom use was associated with lifetime sexual violence and childhood sexual violence, but not with lifetime physical violence.

Sexual violence has been associated with lower expectations for safe condom use, less assertiveness about birth control, and less assertiveness about refusing unwanted sex (Thompson *et al.*, 1997). The long-term consequences of sexual violence that increase the likelihood for risk taking are different than the long-term consequences of physical violence (Leonard and Follette, 2002; Loeb *et al.*, 2002), which could explain why no association between physical violence and inconsistent condom use was found. Although a person who has experienced any type of violence is more likely to experience more adverse medical and psychological consequences than individuals with no violence history (Collins *et al.*, 1999; Liebschutz *et al.*, 1997; Martin *et al.*, 1999; Quinlivan and Evans, 2001; Rosenberg *et al.*, 2000; Wisner *et al.*, 1999), establishing specifically whether sexual violence occurred may further delineate risks for sexual transmission of HIV.

This study has certain limitations. Generalizability of these findings is limited because of the sampling methodology and the restriction to HIV-infected persons with alcohol problems. Also, there is the potential for recall and reporting bias with self-report of sensitive data. Other methods may yield higher reports of inconsistent behaviors, but would be unlikely to alter the direction of our findings (Newman *et al.*, 2002). The interviews, however, were performed confidentially by staff trained to facilitate patient comfort, in part an effort to minimize potential for inaccurate reporting. Another limitation is the inability to determine if the findings differed by gender. When stratified by gender, the sample sizes of the gender groups were too small to establish any significant associations in the main analyses. However, the distribution and prevalence of inconsistent condom use did not differ significantly by gender, therefore

it is unlikely that gender is confounding the associations in the main analyses. Similarly, although the HIV-ALC survey did assess whether assault was perpetrated by strangers or by someone known to the victim, the sample sizes were too small to establish significant associations stratified in this way.

Another potential limitation in this study is in the way the outcome variable was defined. During the interviews, sex was defined as any vaginal intercourse, anal intercourse, or oral sex. Thus, behaviors of varying degrees of actual HIV transmission risk, such as unprotected anal intercourse and unprotected oral sex, were both categorized as inconsistent condom use. Also, the frequency of inconsistent condom use is unknown. The inconsistent condom use variable is therefore a heterogeneous one, with the actual risk of transmitting disease varying between individuals. Additionally, no information was collected on the serostatus of the participants' sexual partners. Although there is risk of transmitting resistant or more virulent strains to an already HIV-infected sexual partner, there obviously is only risk of seroconverting a sexual partner that is not HIV-infected. Therefore, the outcome of inconsistent condom use could overestimate the actual risk of transmitting a new HIV infection in certain situations.

Some HIV-infected persons are continuing to engage in inconsistent condom use, putting sexual partners at risk for acquiring disease. Both a history of lifetime sexual violence, and more specifically childhood sexual violence identified HIV-infected persons with recent inconsistent condom use in this study. Such information can inform clinical practice and the design of HIV prevention programs aimed at reducing risky sexual behaviors among HIV-infected persons. A history of sexual violence may identify a subgroup more at risk for engaging in risky sexual behaviors. In these populations, more detailed screening of risk behaviors, communicating prevention messages, discussing strategies for risk reduction, and reinforcing safer behaviors may be warranted. Similarly, how victims of sexual violence respond to HIV prevention programs aimed at reducing sexual risk behaviors will need further study. Such efforts will aid in development of strategies to encourage safer sexual behaviors among HIV-infected persons.

ACKNOWLEDGMENTS

Data management was provided by DM-STAT, Inc., Medford, MA, in particular Nicole Tibbets, BA. The authors appreciate the contributions of the clinical staff of the HIV Diagnostic Evaluation Unit at Boston Medical Center, including Colleen LaBelle, RN, and Jennifer Doyle. We also acknowledge the invaluable work provided by the research associates and the project managers, Jacqueline Savetsky, MPH, and Seville Meli. Finally, the authors appreciate the contributions of Howard Libman, MD, at the Beth Israel Deaconess Medical Center. Support for this study came from the following grants from the National Institute on Alcohol Violence 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.

REFERENCES

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

Centers for Disease Control and Prevention (2001). *HIV/AIDS Surveillance Report*, p. 13.

Centers for Disease Control and Prevention (2003). Incorporating HIV prevention into the medical care of persons living with HIV: Recommendations of the CDC, the Health Resources and Services Administration, the National Institutes of Health and the HIV Medicine Association of the Infectious Disease Society of America. *MMWR*, *52* (No. RR012).

Cohen, M., Deamant, C., Barkan, S., Richardson, J., Young, M., Holman, S., Anastos, K., Cohen, J., and 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*, 560–565.

Collins, K., Schoen, C., Joseph, S., Duchon, L., Simantov, E., and Yellowitz, M. (1999). *Health concerns across a woman's lifespan: The Commonwealth Fund 1998 Survey of Women's Health*. New York: The Commonwealth Fund.

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

Ehrenstein, V., Horton, H. J., and Samet, J. H. (2004). Inconsistent condom use among HIV-infected patients with alcohol problems. *Drug and Alcohol Dependence*, *73*, 159–166.

El-Bassel, N., Simoni, J. M., Cooper, D. K., Gilbert, L., and Schilling, R. F. (2001). Sex trading and psychological distress among women on methadone. *Psychology of Addictive Behaviors*, *15*, 177–184.

Erbelding, E. J., Stanton, D., Quinn, T. C., and Rompalo, A. (2000). Behavioral and biologic evidence of persistent high-risk behavior in an HIV primary care population. *AIDS*, *14*, 297–301.

Ewing, J. A. (1984). Detecting alcoholism. The CAGE questionnaire. *JAMA*, *252*, 1905–1907.

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

Hamburger, M. E., Moore Koenig, L. J., Vlahov, D., Schoenbaum, E. E., Schuman, P., and Mayer, K. (2004). Persistence of inconsistent condom use: Relation to abuse history and HIV serostatus. *AIDS and Behavior*, *8*, 333–344.

Hillis, S. D., Anda, R. F., Felitti, V. J., and Marchbanks, P. A. (2001). Adverse childhood experiences and sexual risk behaviors in women: A retrospective cohort study. *Family Planning Perspectives*, *33*, 206–211.

Jasinski, J. L., Williams, L. M., and Siegel, J. (2000). Childhood physical and sexual abuse as risk factors for heavy drinking among African-American women: A prospective study. *Child Abuse and Neglect*, *24*, 1061–1071.

Kalichman, S. C., Rompa, D., Luke, W., and Austin, J. (2002). HIV transmission risk behaviors among HIV-positive persons in serodiscordant relationships. *International Journal of STD and AIDS*, *13*, 677–682.

Kamb, M. L., Fishbein, M., Douglas, J. M., Jr., Rhodes, F., Rogers, J., Bolan, G., Zenilman, J., Hoxworth, T., Malotte, C. K., Iatesta, M., Kent, C., Lentz, A., Graziano, S., Byers, R. H., and Peterman, T. A. (1998). Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: A randomized controlled trial. Project RESPECT Study Group. *JAMA*, *280*, 1161–1167.

Kwiatkowski, C., and Booth, R. (1998). HIV-seropositive drug users and unprotected sex. *AIDS and Behavior*, *2*, 151–159.

Lenderking, W. R., Wold, C., Mayer, K. H., Goldstein, R., Losina, E., and Seage, G. R., III (1997). Childhood sexual abuse among homosexual men. Prevalence and association with inconsistent sex. *Journal of General Internal Medicine*, *12*, 250–253.

Leonard, L. M., and Follette, V. M. (2002). Sexual functioning in women reporting a history of child sexual abuse: Review of the empirical literature and clinical implications. *Annual Review of Sex Research*, *13*, 346–388.

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

Liebschutz, J. M., Mulvey, K. P., and Samet, J. H. (1997). Victimization among substance-abusing women. Worse health outcomes. *Archives of Internal Medicine*, *157*, 1093–1097.

Loeb, T. B., Williams, J. K., Carmona, J. V., Rivkin, I., Wyatt, G. E., Chin, D., and Asuan-O'Brien, A. (2002). Child sexual abuse: Associations with the sexual functioning of adolescents and adults. *Annual Review of Sex Research*, *13*, 307–345.

Marks, G., Burris, S., and Peterman, T. A. (1999). Reducing sexual transmission of HIV from those who know they are infected: The need for personal and collective responsibility. *AIDS*, *13*, 297–306.

Martin, S. L., Clark, K. A., Lynch, S. R., Kupper, L. L., and Cilenti, D. (1999). Violence in the lives of pregnant teenage women: Associations with multiple substance use. *American Journal of Drug and Alcohol Abuse*, *25*, 425–440.

McLellan, A. T., Luborsky, L., Cassiola, J., Griffith, J., Evans, F., Bass, H., and 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.

Metsch, L., McCoy, C., Lai, S., and Miles, C. (1998). Continuing risk behaviors among HIV-seropositive chronic drug users in Miami, Florida. *AIDS and Behavior*, *2*, 161–169.

Mullings, J. L., Marquart, J. W., and Brewer, V. E. (2000). Assessing the relationship between child sexual abuse and marginal living conditions on HIV/AIDS-related risk behavior among women prisoners. *Child Abuse and Neglect*, *24*, 677–688.

National Institute of Mental Health Multisite HIV Prevention Trial Group (1998). The NIMH Multisite HIV Prevention Trial: Reducing HIV sexual risk behavior. *Science*, *280*, 1889–1894.

National Institute of Mental Health Multisite HIV Prevention Trial Group (2001). 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 and Health*, *33*, 163–180.

National Institute of Mental Health Multisite HIV Prevention Trial Group (2002). Predictors of sexual behavior patterns over one year among persons at high risk for HIV. *Archives of Sexual Behavior*, *31*, 165–176.

Navaline, H. A., Snider, E. C., Petro, C. J., Tobin, D., Metzger, D., Alterman, A. I., and 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*, S281–S283.

Newman, J. C., Des Jarlais, D. C., Turner, C. F., Gribble, J., Cooley, P., and Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health*, *92*, 294–297.

O'Leary, A., Purcell, D., Remien, R., and Gomez, C. (2003). Childhood sexual abuse and sexual transmission risk behaviour among HIV-positive men who have sex with men. *AIDS Care*, *15*, 17–26.

Parillo, K. M., Freeman, R. C., Collier, K., and Young, P. (2001). Association between early sexual abuse and adult HIV-risky sexual behaviors among community-recruited women. *Child Abuse and Neglect*, *25*, 335–346.

Paul, J. P., Catania, J., Pollack, L., and Stall, R. (2001). Understanding childhood sexual abuse as a predictor of sexual risk-taking among men who have sex with men: The Urban Men's Health Study. *Child Abuse and Neglect*, *25*, 557–584.

Quinlivan, J. A., and Evans, S. F. (2001). A prospective cohort study of the impact of domestic violence on young teenage pregnancy outcomes. *Journal of Pediatric and Adolescent Gynecology*, *14*, 17–23.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Rosenberg, H. J., Rosenberg, S. D., Wolford, G. L., II, Manganiello, P. D., Brunette, M. F., and Boynton, R. A. (2000). The relationship between trauma, PTSD, and medical utilization in three high risk medical populations. *International Journal of Psychiatry in Medicine*, *30*, 247–259.

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

Samet, J. H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D. E., and 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.

Simpson, T. L., and Miller, W. R. (2002). Concomitance between childhood sexual and physical abuse and substance use problems. A review. *Clinical Psychology Review*, *22*, 27–77.

Thompson, N. J., Potter, J. S., Sanderson, C. A., and Maibach, E. W. (1997). The relationship of sexual abuse and HIV risk behaviors among heterosexual adult female STD patients. *Child Abuse and Neglect*, *21*, 149–156.

Tjaden, P., and Thoennes, N. (2000). *Extent, nature, and consequences of intimate partner violence. Findings from the National Violence Against Women Survey.* Washington, DC: US Department of Justice.

Wisner, C. L., Gilmer, T. P., Saltzman, L. E., and Zink, T. M. (1999). Intimate partner violence against women: Do victims cost health plans more? *Journal of Family Practice*, *48*, 439–443

Intoxication Before Intercourse and Risky Sexual Behavior in Male Veterans With and Without Human Immunodeficiency Virus Infection

Robert L. Cook, MD, MPH, Kathleen A. McGinnis, MS,† Kevin L. Kraemer, MD, MSc,* Adam J. Gordon, MD, MPH,*‡ Joseph Conigliaro, MD, MPH,§ Stephen A. Maisto, PhD,¶ Jeffrey H. Samet, MD, MA, MPH,‖ Stephen Crystal, PhD,** David Rimland, MD,††‡‡ Kendall J. Bryant, PhD,§§ R. Scott Braithwaite, MD,¶¶‖‖ and Amy C. Justice, MD, PhD‖‖*

Background: Male veterans represent a large population at risk for acquiring or transmitting human immunodeficiency virus (HIV) infection. We sought to determine the prevalence of risky sexual behavior among veterans with and without HIV infection and to assess the relationship of intoxication before intercourse and other measures of drug and alcohol use to risky sexual behavior in this population.

Methods: We analyzed baseline data on 1009 HIV-positive (mean age 49 years) and 710 HIV-negative male veterans (mean age 55 years) who were participating in the Veterans Aging Cohort 5-Site Study (VACS 5). Participants completed a written questionnaire that included measures of alcohol and drug use and risky sexual behavior.

Results: Compared with HIV-negative veterans, HIV-positive veterans were more likely to report 5 or more sexual partners in the past year (14% vs. 4%, $P < 0.01$), less likely to report not using a condom at last intercourse (25% vs. 75%, $P < 0.01$), and similarly likely to report having 2 or more partners and inconsistent condom use (10% vs. 10%). Among sexually active HIV-positive veterans, intoxication before intercourse was significantly associated with having 5 or more sexual partners in the past year (odds ratio [OR] 1.8, 95% confidence interval [CI] 1.1–2.8), inconsistent condom use (OR 1.8, 95% CI 1.2–2.7), and the combined measure of 2 or more partners and inconsistent condom use (OR 1.8, 95% CI 1.1–3.0). Intoxication before intercourse was not significantly associated with these behaviors in HIV-negative veterans, although similar trends were noted.

Conclusion: Risky sexual behavior was common among male veterans attending outpatient clinics and is more common among HIV-positive veterans who use alcohol and drugs in sexual situations. Asking HIV-positive men a single question about intoxication before intercourse could help to identify men at increased risk of engaging in risky sexual behavior, and specific advice to avoid intoxication in sexual situations could help to reduce risky sexual behavior.

Key Words: HIV infection, HIV-positive status, prevention, risk behaviors, substance abuse

(*Med Care* 2006;44: S31–S36)

From the *Department of Medicine and †Center for Urban and Social Research, University of Pittsburgh and ‡Center for Health Equity Research and Promotion, VA Medical Center, Pittsburgh, Pennsylvania; §Department of Medicine, University of Kentucky, Lexington, Kentucky; ¶Department of Psychology, Syracuse University, Syracuse, New York; ‖Department of Medicine, Boston University School of Medicine, Boston Massachusetts; **Institute for Health, Health Care Policy, and Aging Research, Rutgers University, New Brunswick, New Jersey, ††Department of Medicine, Emory University School of Medicine and ‡‡VA Medical Center, Atlanta, Georgia; §§National Institute on Alcohol Abuse and Alcoholism, Bethesda Maryland; and ¶¶Department of Medicine, Yale University and ‖‖West Haven VA Medical Center, New Haven, Connecticut.

Primary study funding was from National Institute on Alcohol and Alcohol Abuse (3U01 AA 13566), National Institute of Aging (K23 AG00826), Robert Wood Johnson Generalist Faculty Scholar Award, an Interagency Agreement between National Institute on Aging, National Institute of Mental Health and the Veterans Health Administration, and the Veterans Health Administration Office of Research and Development and Public Health Strategic Health Care Group.

Data were presented in part at the 2003 Annual Meeting of the Research Society on Alcoholism, Ft. Lauderdale, Florida.

Reprints: Robert L. Cook, MD, MPH, 230 McKee Place, Suite 600, University of Pittsburgh, Pittsburgh, PA 15213. E-mail: cookrl@upmc.edu.

ISSN: 0025-7079/06/4400-0031

The number of individuals infected with the human immunodeficiency virus (HIV) continues to increase worldwide, with the majority of new infections caused by sexual transmission.[1,2] Clinicians are encouraged to provide HIV-prevention counseling to their patients, and the need to provide such counseling to HIV-positive persons is increasingly being noted.[3] HIV-positive individuals who engage in risky sexual activity can transmit the virus to previously HIV-negative individuals, transmit or receive new or resistant HIV strains with individuals who are already HIV-positive,[4] and acquire new sexually transmitted infections.[5–8]

The CDC/IDSA/HRSA guidelines for incorporating HIV prevention into medical care of persons living with HIV state that clinicians should screen for "barriers to abstinence or correct condom use (eg, . . . alcohol and other drug use before or during sex)."[3] However, compliance with these recommendations is poor. A recent study of physicians who care for HIV-positive individuals found that 60% routinely

counseled newly infected patients about HIV prevention, but only 14% counseled established patients.[9] Physicians were less likely to provide prevention counseling to their HIV-positive patients if the patient was male or if the patient had an ongoing drug or alcohol use problem.[9] Discussing the role of alcohol and drug use in HIV prevention can be challenging because of the large number of potential risk factors, including the consumption of alcohol, methamphetamine, cocaine, or amyl nitrates,[10–17] as well as the large numbers of items included in commonly recommended assessment instruments for hazardous alcohol or drug use.[18,19]

To our knowledge, the association of alcohol and drug use to risky sexual behavior has not been reported in a veteran population. We performed our study in veterans because the Veterans Administration (VA) is the largest single provider of HIV care in the nation, serving at least 19,000 HIV-positive veterans in 2003. Moreover, the men attending clinics at VA hospitals often are older than 50 years of age, and it is unclear whether these relatively older men are engaging in risky sexual behavior. Our primary study aim was to investigate whether a single construct (intoxication with alcohol or drugs before sexual intercourse) or other measures of drug and alcohol use were associated with risky sexual behavior in HIV-positive and HIV-negative male veterans.

METHODS

Recruitment

Study participants were enrolled in the Veterans Aging Cohort 5-Site Feasibility Study (VACS 5), a 5-site study of veterans with and without HIV infection. Details of study methods are reported elsewhere.[20] In brief, participants were recruited between September 2001 and June 2002 from infectious disease and primary care clinics at Department of Veterans Affairs Medical Centers in 5 urban areas: Atlanta, Houston, Los Angeles, the Bronx, and Manhattan. Target enrollment was set at 65% of HIV-positive patients receiving care for each Infectious Disease clinic in fiscal year 2002 and was stratified based on age and race for each site. Targets for controls were set for each site based upon the local HIV-positive enrollment targets.

The final sample of HIV-positive patients represented 49% of the HIV-positive patients seen in these Infectious Disease clinics during this time. Nonparticipants were more likely to be black, have fewer clinic visits, higher viral loads, and to have been on HAART for a shorter period of time ($P < 0.05$).[20] The institutional review boards at each site and the coordinating site approved the study protocol, and each participant signed an informed consent document.

Survey Instrument

Each participant completed a written survey that included questions concerning demographic characteristics, current depression, sexual behavior, and drug and alcohol use. We categorized race (white vs. nonwhite), annual household income (<$12,000 vs. ≥$12,000), and current marital status (married or living with long-term partner vs. other). We defined depression as the presence of a score of ≥10 on the Patient Health Questionnaire (PHQ-9), a 9-item measure in which scores of ≥10 have a sensitivity of 88% and specificity of 88% for major depression.[21] Men reported whether they had ever had sex with another man, and whether they had ever used injection drugs.

Sexual Behavior Measures

To assess the number of sexual partners, we asked the following open-ended question: "During the past 12 months, with how many persons have you had sex?" (with sex defined as oral, vaginal, or anal sex). We assessed condom use by asking the following question: "Thinking back about the last time you had sex, did you or your partner use a condom?" We did not have information on the serostatus of participant's sexual partners.

We created 3 outcome measures of risky sexual behavior. First, we categorized persons as having "inconsistent condom use" if they did not use a condom at their last sexual intercourse. Second, we created a combined measure of 2 or more partners in the past year and inconsistent condom use. This combined definition would avoid classifying men with risky sex if they had inconsistent condom use within a long-term monogamous relationship. Finally, we categorized persons as having multiple sexual partners if they reported 5 or more sexual partners in the past year.

Alcohol and Drug Use Measures

Alcohol consumption was assessed using the 10-item Alcohol Use Disorders Identification Test (AUDIT).[18,19] Persons were classified as having hazardous alcohol use if their total score on the AUDIT was ≥8 or if their answer to one of the AUDIT items indicated that they were binge drinking (consuming ≥6 drinks in 1 sitting at least once a month).[22] Drug use was assessed by the Drug Abuse Screening Test (DAST), a 10-item scale that assesses various consequences of drug use.[23] Persons were classified as having hazardous drug use if their score on the DAST was ≥3. Men reported whether they had ever used any injection drugs; other specific drugs were not assessed. Substance use before sex was assessed by the single item, "the last time you had sexual intercourse, were you intoxicated or high from alcohol or drugs?"

Statistical Analyses

All analyses were conducted using Stata Statistical Software, Release 8 (Stata Corp., College Station, TX). Comparisons between proportions were made using the χ^2 distribution. We excluded the 43 subjects (2.5%) who were women, because factors associated with their risky sexual behavior were likely to be different from those in men and the number of women was too small to make definitive conclusions. All analyses were stratified by HIV status.

We determined the proportion of men who reported various drug and alcohol consumption and the proportion who reported each of the 3 categories of risky sexual behavior. Comparisons between HIV-positive and HIV-negative men were done by χ^2 analysis for categorical variables and t test for continuous variables. For additional analyses, we excluded persons reporting zero sexual partners in the past 12 months (n = 496, 29%).

We conducted bivariate analyses to determine factors associated with the combined measure of risky sexual behavior (≥2 sexual partners and inconsistent condom use) in sexually active HIV-positive and HIV-negative men. Then, we used multivariable logistic regression to determine the association of intoxication before sex, hazardous alcohol use, and hazardous drug use to each of the 3 sexual behavior outcome measures (inconsistent condom use, ≥2 sex partners and inconsistent condom use, and ≥5 sexual partners). Each of the multivariate models was adjusted for age, race, marital status, and depression. In 1 series of logistic models, we included the single item "intoxicated before intercourse" as the only measure of drug and alcohol use. In a second set of models, we included scale measures of hazardous alcohol use and hazardous drug use as the measures of alcohol and drug abuse, but did not include "intoxication before intercourse." We used c-statistics to compare the overall fit of 2 sets of models.

RESULTS

Of the 1719 men included in this analysis, 1009 were HIV-positive and 710 were HIV-negative. Compared with HIV-negative men, the HIV-positive men were younger, more likely to be black, less likely to be married, had lower incomes, and were more likely to have a history of sex with other men (Table 1). Hazardous alcohol consumption was similar in HIV-positive and HIV-negative men, whereas hazardous drug use was significantly more common in HIV-positive men.

Among men who reported being intoxicated before the last intercourse, 19% consumed alcohol only in the past year; 17% consumed drugs only, and 61% consumed both alcohol and drugs. One-third (33%) of persons who reported being intoxicated at the time of their last intercourse did not meet the definition for either hazardous alcohol use or hazardous drug use.

TABLE 1. Characteristics of Study Participants

	HIV+ (n = 1009)	HIV− (n = 710)	P
Age, yr (mean, SE)	49.3 (0.28)	55.4 (0.38)	<0.001
Race/ethnicity (%)			<0.001
Black	55	44	
White	28	38	
Hispanic	13	13	
Other	4	5	
Married (%)	12	39	<0.001
Income <$12,000 (%)	49	38	<0.001
HIV risk factor			
Sex with men, ever (%)	55	6	<0.001
IVDU, ever (%)	29	11	<0.001
Alcohol/drug use (%)			<0.001
No alcohol drug use	24	32	
Alcohol use only	31	45	
Drug use only	16	9	
Both drug and alcohol use	29	15	
Hazardous alcohol use (%)	19	21	0.20
Hazardous drug use (%)	34	18	<0.001
Intoxication before sex*	22	9	<0.001
Sexual partners (past year)			<0.001
0	32	26	
1	32	55	
2–4	22	15	
≥5	14	4	
Inconsistent condom use*	25	75	<0.001
Multiple sex partners (>2) and inconsistent condom use*	10	10	0.75

*Among those who were sexually active in past 12 months.
Hazardous alcohol use: AUDIT score ≥8 or consumed 6 or more drinks in one sitting at least monthly. Hazardous drug use: Total score of Drug Abuse Screening Test ≥3.
IVDU indicates intravenous drug use.

Sexual Behaviors of HIV-Positive and HIV-Negative Male Veterans

The proportion of men who were currently sexually active was similar in the HIV-positive men and HIV-negative men (68% and 74%; Table 1). HIV-negative men were much more likely to be in monogamous relationships, whereas HIV-positive men were significantly more likely to report 2 or more sexual partners (36% vs. 19%; $P < 0.001$) and 5 or more partners (14% vs. 4%, $P < 0.001$) in the past year. Having 2 or more sexual partners remained relatively common in men who were 60 years or older (26% of HIV-positive and 12% of HIV-negative men) and in men who were currently married (12% of men in each group).

Of the 1161 men who were sexually active, HIV-positive men were more likely than HIV-negative men to have used a condom at last intercourse (73% vs. 27%, $P < 0.001$; Table 1). In sexually active HIV-positive and HIV-negative men, 1 in 10 men had both 2 ore more sexual partners and inconsistent condom use in the past year. HIV-positive men were more likely to report 2 or more sexual partners and inconsistent condom use if they were white, not married, currently depressed, ever had sex with men, reported hazardous drug use, or reported intoxication before intercourse (Table 2). In HIV-negative men, the only variables significantly associated with this measure of risky sexual behavior were ever having sex with men and hazardous alcohol use (Table 2).

Factors Associated With Risky Sexual Behavior: Multivariate Analysis

HIV-Positive Men

Among sexually active HIV-positive men, intoxication before intercourse was significantly associated with each of the 3 risky sexual behaviors that we assessed. Specifically, men who were intoxicated before their last intercourse were significantly more likely to report inconsistent condom use (odds ratio [OR] 1.8, 95% confidence interval [CI] 1.2–2.7); 2 or more sexual partners and inconsistent condom use (OR 1.8, 95% CI 1.1–3.0); and having 5 or more sexual partners in the past year (OR 1.8, 95% CI 1.1–2.8; Table 3). Hazardous drug use was significantly associated with both inconsistent condom use and the combined measure of 2 or more sexual partners and inconsistent condom use, whereas hazardous alcohol use was not significantly associated with any of the sexual behavior outcomes in HIV-positive men.

TABLE 2. Factors Associated With Risky Sexual Behavior (RSB: ≥2 Sexual Partners and Inconsistent Condom Use) Among HIV-Positive and HIV-Negative Male Veterans: Bivariate Analysis

	HIV-Positive Men			HIV-Negative Men		
Risky Sexual Behavior	**RSB Yes (n = 94)**	**RSB No (n = 570)**	***P***	**RSB Yes (n = 71)**	**RSB No (n = 439)**	***P***
Age, years (mean, SE)	46.9 (0.99)	48.2 (0.35)	0.17	53.2	53.8	0.56
Race						
Black	47%	60%	0.001	58%	46%	0.21
White	43%	24%		25%	35%	
Hispanic	9%	12%		11%	15%	
Other	2%	4%		6%	4%	
Low income (<$12,000/yr)	44%	45%	0.89	34%	35%	0.9
Married/long-term relationship	4%	15%	0.005	36%	43%	0.30
Depressed	34%	21%	0.005	18%	20%	0.68
Sex with men (ever)	77%	55%	<0.001	13%	4%	0.002
IVDU (ever)	29%	30%	0.10	11%	12%	0.86
Intoxicated before sex	33%	20%	0.004	13%	8%	0.11
Hazardous alcohol use	20%	17%	0.46	34%	22%	0.03
Hazardous drug use	49%	33%	0.002	19%	19%	0.87

Hazardous Alcohol Use: AUDIT score ≥8 or consumed 6 or more drinks in one sitting at least monthly. Hazardous Drug Use: Total score of Drug Abuse Screening Test ≥3.
IVDU indicates intravenous drug use.

TABLE 3. Relationship of Alcohol and Drug Use to Various Risky Sexual Behaviors in 657 HIV-Positive Male Veterans: Multivariate Analysis

	Inconsistent Condom Use (Nonuse at Last Intercourse)		≥2 Partners in Past Year and Inconsistent Condom Use		≥5 Sexual Partners in Past Year	
	OR	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**
HIV-positive men						
Intoxicated before sex at last intercourse*	1.8	(1.2–2.7)	1.8	(1.1–3.0)	1.8	(1.1–2.8)
Hazardous alcohol use (AUDIT ≥ 8)†	1.2	(0.8–2.0)	0.9	(0.5–1.6)	0.7	(0.4–1.2)
Hazardous drug use (DAST ≥ 3)‡	1.9	(1.3–2.9)	2.4	(1.4–3.9)	1.5	(0.97–2.2)
HIV-negative men						
Intoxicated before sex at last intercourse*	0.8	(0.4–1.6)	1.7	(0.8–3.8)	2.2	(0.8–6.4)
Hazardous alcohol use (AUDIT ≥ 8)†	1.1	(0.6–1.8)	2.0	(1.1–3.4)	0.6	(0.2–1.6)
Hazardous drug use (DAST ≥ 3)‡	0.6	(0.4–1.1)	0.8	(0.4–1.7)	6.6	(2.8–15.6)

*Multivariate analysis adjusted for age, race, marital status, and depression.
†Multivariate analysis adjusted for age, race, marital status, DAST, and depression.
‡Multivariate analysis adjusted for age, race, marital status, AUDIT, and depression.

HIV-Negative Men

The relationship of alcohol and drug use to risky sexual behavior in HIV-negative men was less consistent (Table 3). Hazardous alcohol use was significantly associated with the combined outcome of 2 or more sexual partners and inconsistent condom use (OR 2.0, 95% CI 1.1–3.4), whereas hazardous drug use was significantly associated with having ≥5 partners in the past year (OR 6.6, 95% CI 2.8–15.6; Table 3).

DISCUSSION

Risky sexual behavior was relatively common in this large, middle-aged sample of more than 1700 HIV-positive and HIV-negative male veterans. One in 10 veterans, regardless of HIV status, reported 2 or more sexual partners and inconsistent condom use in the past year. This behavior was not limited to young or single men. Among the HIV-positive men, 26% who were 60 years or older and 12% who were currently married reported having multiple sexual partners. These findings are consistent with research suggesting that one-third of HIV-positive persons regularly engage in risky sexual behavior[10–17,24–27] and many are acquiring new sexually transmitted infections.[5–8] The findings from this study also confirm the importance of explicitly addressing alcohol and drug use as an important component of HIV prevention counseling, both in HIV-positive and HIV-negative men.

In HIV-positive men, we found that a single question about whether men were intoxicated before intercourse was equally predictive of risky sexual behavior compared with more detailed assessments of hazardous alcohol and drug use. This finding is consistent with current recommendations to

assess for alcohol and drug use in sexual situations during HIV-prevention counseling among HIV-positive men,[3] and with other research that found similar associations.[11,14,16,28]

We used 3 different measures of risky sexual behavior. Although these measures are somewhat overlapping and associated with one another, it is possible that substance use could be more strongly associated with some aspects of risky sexual behavior than others (eg, nonuse of condoms or multiple sexual partners). In the HIV-positive men, the relationships were similar for each of the 3 sexual behaviors that we assessed. However, in the HIV-negative men, the association of hazardous drug use was stronger for multiple sexual partners than for the other relationships that we examined. This finding suggests that drugs may be related to finding new partners in different settings, but that that they may not directly affect use of condoms. These findings emphasize the need to consider different aspects of sexual behavior when assessing HIV risks.

Our data are consistent with other information indicating that there may be differences in the relationship between alcohol and drug use and risky sexual behavior by HIV status.[29] Hazardous alcohol consumption was associated with at least 1 measure of risky sexual behavior in HIV-negative men but in not in the HIV-positive men. In contrast, hazardous drug use was associated with 2 measures of risky sexual behavior in HIV-positive men but not in the HIV-negative men. It is not clear whether this finding reflects underlying differences in the preferred substance of the 2 populations of men, whether a diagnosis of HIV infection itself leads to changes in alcohol and drug use, or whether the findings simply reflect other underlying differences in the HIV-positive and HIV-negative men in this sample. For example, HIV-positive men may have different motivations to drink or use drugs, such as to escape feeling guilty about their sexual behavior or helping cope with stress related to living with HIV infection.[24,30]

Our measures of sexual behavior were somewhat limited and could not differentiate between specific types of sexual behavior (eg, oral, vaginal, and anal sex) nor determine characteristics of the participants' sexual partners (eg, the HIV serostatus of partners). Thus, it is possible that some persons with relatively safe behavior (eg, oral sex without using a condom) might be grouped together with persons who engaged in higher risk behaviors (eg, anal sex without a condom). Our combined measure of risky sexual behavior ($\geq$2 sexual partners and inconsistent condom use) has been used in other recently-reported research,[17] and the assessment of condom use at last intercourse provided the opportunity to make event-specific comparisons (eg, intoxication before last intercourse vs. condom use at last intercourse). Although we used validated measures of risky behavior and substance use, it is likely that some participants underreported these behaviors. Because our measure of condom use asked only about use during the last intercourse, it is likely to underestimate the proportion of persons who do not always use condoms and who may be engaging in risky sexual behavior.

Other study limitations included a lack of detail regarding the specific drugs that participants used, and a sample size that was somewhat smaller in HIV-negative men, which limited the power to detect statistically significant differences to the same degree as in HIV-positive men. The generalizability of the findings to nonveteran populations is not clear. However, the majority of previous research on alcohol and drug use and risky sexual behavior in HIV-positive men has been among samples of men who had sex with men, and our sample included additional risk groups. Finally, the cross-sectional nature of the data limited the ability to ascribe causality.

In summary, our results demonstrate that many male veterans, both HIV-positive and HIV-negative, are engaging in risky sexual behavior. These findings reinforce the need for clinicians who provide care to veterans to conduct regular HIV risk assessments and provide prevention counseling.[3,31–33] Such counseling may be particularly important for persons who are already HIV-infected and continue to engage in risky sexual behavior. Our findings suggest that a single question about intoxication before intercourse can be used to identify HIV-positive men with an increased propensity for risky sexual behavior. Therefore, routine inquiries about intercourse should be included as part of routine HIV-prevention assessments, and specific advice to avoid being intoxicated in sexual situations may help to prevent risky sexual behavior and its resulting spread of HIV and other STDs.

ACKNOWLEDGMENTS

The authors acknowledge the helpful contributions of the study coordinators at each of the VACS sites and the veterans who provided the data for this study. We also acknowledge additional site investigators for VACS 5, including Sheldon Brown, MD, (Bronx), Michael Simberkoff, MD (Brooklyn/Manhattan), Maria C. Rodriguez-Barradas, MD (Houston), and Matthew Goetz, MD (Los Angeles).

REFERENCES

1. Steinbrook R. The AIDS epidemic in 2004. *N Engl J Med.* 2004;351:115–117.
2. Kates J, Sorian R, Crowley JS, et al. Critical policy challenges in the third decade of the HIV/AIDS epidemic. *Am J Public Health.* 2002;92:1060–1063.
3. Centers for Disease Control and Prevention. Incorporating HIV Prevention into the Medical Care of Persons Living with HIV: Recommendations of the CDC, the Health Resources and Services Administration, the NIH, and the HIV Medicine Association of the Infectious Diseases Society of America. *MMWR.* 2003;52:1–24.
4. Blackard JT, Mayer KH. HIV superinfection in the era of increased sexual risk-taking. *Sex Transm Dis.* 2004;31:201–204.
5. Collins T, Celum CL. The clinical manifestations and treatment of sexually transmitted diseases in human immunodeficiency virus-positive men. *Clin Infect Dis.* 2001;32:611–622.
6. Scheer S, Chu PL, Klausner JD, et al. Effect of highly active antiretroviral therapy on diagnoses of sexually transmitted diseases in people with AIDS. *Lancet.* 2001;357:432–435.
7. Fox KK, del Rio C, Holmes KK, et al. Gonorrhea in the HIV era: a reversal in trends among men who have sex with men. *Am J Public Health.* 2001;91:959–964.
8. Centers for Disease Control and Prevention. Primary and secondary syphilis among men who have sex with men–New York City, 2001. *MMWR.* 2002;51:853–856.
9. Metsch LR, Pereyra M, del Rio C, et al. Delivery of HIV prevention counseling by physicians at HIV medical care settings in 4 US cities. *Am J Public Health.* 2004;94:1186–1192.

10. Morin SF, Steward WT, Charlebois ED, et al. Predicting HIV transmission risk among HIV-infected men who have sex with men. *J Acquir Immune Defic Syndr.* 2005;40:226–235.
11. Purcell DW, Moss S, Remien RH, et al. Illicit substance use, sexual risk, and HIV-positive gay and bisexual men: differences by serostatus of casual partners. *AIDS.* 2005;19:S37–S47.
12. Denning PH, Campsmith ML. Unprotected anal intercourse among HIV-positive men who have a steady male sex partner with negative or unknown HIV serostatus. *Am J Public Health.* 2005;95:152–158.
13. Palepu A, Raj A, Horton NJ, et al. Substance abuse treatment and risk behaviors among HIV-infected persons with alcohol problems. *J Subst Abuse Treat.* 2005;28:3–9.
14. Benotsch EG, Kalichman SC, Kelly JA. Sexual compulsivity and substance use in HIV-seropositive men who have sex with men: prevalence and predictors of high-risk behaviors. *Addict Behav.* 1999;24:857–868.
15. Vanable PA, McKirnan DJ, Buchbinder SP, et al. Alcohol use and high-risk sexual behavior among men who have sex with men: the effects of consumption level and partner type. *Health Psychol.* 2004;23: 525–532.
16. Kalichman SC, Weinhardt L, DiFonzo K, et al. Sensation seeking and alcohol use as markers of sexual transmission risk behavior in HIV-positive men. *Ann Behavior Med.* 2002;24:229–235.
17. Stein M, Herman DS, Trisvan E, et al. Alcohol use and sexual risk behavior among human immunodeficiency virus-positive persons. *Alcohol Clin Exp Res.* 2005;29:837–843.
18. Babor TF, Ramon de la Fuente J, Saunders J, et al. *AUDIT. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care.* Geneva: World Health Organization; 1992;2:1–40.
19. Maisto SA, Conigliaro J, McNeil M, et al. An empirical investigation of the factor structure of the AUDIT. *Psychol Assess.* 2000;12:346–353.
20. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): overview and description. *Med Care.* 2006;44:(Suppl 2): S13–S24.
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606–613.
22. Gordon AJ, Maisto SA, McNeil M, et al. Three questions can detect hazardous drinkers. *J Fam Pract.* 2001;50:313–320.
23. Maisto SA, Carey MP, Carey KB, et al. Use of the AUDIT and the DAST-10 to identify alcohol and drug use disorders among adults with a severe and persistent mental illness. *Psychol Assess.* 2000;12:186–192.
24. Kalichman SC. HIV transmission risk behaviors of men and women living with HIV-AIDS: prevalence, predictors, and emerging clinical interventions. *Clin Psychol Sci Pract.* 2000;7:32–47.
25. Erbelding EJ, Stanton D, Quinn TC. Behavioral and biologic evidence of persistent high-risk behavior in an HIV primary care population. *AIDS.* 2000;14:297–301.
26. Crepaz N, Marks G. Towards an understanding of sexual risk behavior in people living with HIV: a review of social, psychological, and medical findings. *AIDS.* 2002;16:135–149.
27. Kalichman SC, Rompa D, Luke W, et al. HIV transmission risk behaviours among HIV-positive persons in serodiscordant relationships. *Int J STD AIDS.* 2002;13:677–682.
28. Purcell DW, Parsons JT, Halkitis PN, et al. Substance use and sexual transmission risk behavior of HIV-positive men who have sex with men. *J Subst Abuse.* 2001;13:185–200.
29. Stall R, Paul JP, Greenwood G, et al. Alcohol use, drug use and alcohol-related problems among men who have sex with men: the Urban Men's Health Study. *Addiction.* 2001;96:1589–1601.
30. Robins AG, Dew MA, Davidson S, et al. Psychosocial factors associated with risky sexual behavior among HIV-seropositive gay men. *AIDS Educ Prev.* 1994;6:483–492.
31. Kelly JA, Kalichman SC. Behavioral research in HIV/AIDS primary and secondary prevention: recent advances and future directions. *J Consult Clin Psychol.* 2002;70:626–639.
32. Schreibman T, Friedland G. Human immunodeficiency virus infection prevention: strategies for clinicians. *Clin Infect Dis.* 2003;36:1171–1176.
33. Kamb ML, Fishbein M, Douglas JM Jr, et al. Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: a randomized controlled trial. Project RESPECT Study Group. *JAMA.* 1998;280:1161–1167.

Prevalence of Torture Survivors Among Foreign-Born Patients Presenting to an Urban Ambulatory Care Practice

Sondra S. Crosby, MD,[1,2] Marie Norredam, MD,[3] Michael K. Paasche-Orlow, MD, MPH,[1] Linda Piwowarczyk, MD, MPH,[2,4] Tim Heeren, PhD,[5] Michael A. Grodin, MD[2,4,5]

[1]Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Boston Center for Refugee Health and Human Rights, Boston, MA, USA; [3]Department of Health Services Research, Institute of Public Health, University of Copenhagen, Copenhagen K, Denmark; [4]Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA; [5]Boston University School of Public Health, Boston, MA, USA.

BACKGROUND: The prevalence of torture among foreign-born patients presenting to urban medical clinics is not well documented.

OBJECTIVE: To determine the prevalence of torture among foreign-born patients presenting to an urban primary care practice.

DESIGN: A survey of foreign-born patients.

PATIENTS: Foreign-born patients, age ≥ 18, presenting to the Primary Care Clinic at Boston Medical Center.

MEASUREMENTS: Self-reported history of torture as defined by the UN, and history of prior disclosure of torture.

RESULTS: Of the 308 eligible patients, 88 (29%) declined participation, and 78 (25%) were not included owing to lack of a translator. Participants had a mean age of 47 years (range 19 to 76), were mostly female (82/142, 58%), had been in the United States for an average of 14 years (range 1 month to 53 years), and came from 35 countries. Fully, 11% (16/142, 95 percent confidence interval 7% to 18%) of participants reported a history of torture that was consistent with the UN definition of torture. Thirty-nine percent (9/23) of patients reported that their health care provider asked them about torture. While most patients (15/23, 67%) reported discussing their experience of torture with someone in the United States, 8 of 23 (33%) reported that this survey was their first disclosure to anyone in the United States.

CONCLUSION: Among foreign-born patients presenting to an urban primary care center, approximately 1 in 9 met the definition established by the UN Convention Against Torture. As survivors of torture may have significant psychological and physical sequelae, these data underscore the necessity for primary care physicians to screen for a torture history among foreign-born patients.

KEY WORDS: torture; prevalence.
DOI: 10.1111/j.1525-1497.2006.00488.x
J GEN INTERN MED 2006; 21:764–768.

Recent events at Abu Ghraib and Guatánamo Bay have raised the public's awareness of torture. The United States of America became a signatory of The United Nations Convention Against Torture and Other Cruel, Inhuman, and Degrading Treatment or Punishment (C.A.T.) in 1988.[1] The American Medical Association's Code of Ethics states that physicians must not "countenance, condone, or participate in the practice of torture or other forms of cruel, inhuman, or degrading procedures, whatever the offence of which the victim of such procedures is suspected, accused or convicted."[2,3] The C.A.T. defines survivors of torture as those who have endured "severe pain or suffering, whether physical or mental, . . . when such pain or suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or any other person acting in an official capacity."[4] The Offices of the United Nations High Commissioner for Human Rights has documented the broad prevalence of torture around the world.[5]

The purpose of torture is to break the mind, body, and spirit of victims and to send fear into communities. Torture often includes beatings, psychological torment, rape, burning, suspension, electrical shocks, and detention under inhumane conditions.[6] Torture is one of the most traumatizing of human experiences and can result in significant long-term medical and psychological sequelae.[7–12]

Several community-based surveys of the prevalence of torture survivors in specific immigrant communities have been published. Marshall et al.[13] surveyed a cohort of 586 Cambodian adults in Long Beach, California, who lived in Cambodia during the Khmer Rouge reign, and found that fully 54% reported a history of torture. Jaranson et al.[14] surveyed 622 Somali and 512 Oromo refugees in Minneapolis-St. Paul and found the prevalence of torture to be 36% and 55%, respectively. We are aware of 2 clinic-based surveys examining the prevalence of torture. Eisenman et al.[15] reported a prevalence of torture of 6.6% among a sample of 121 foreign-born patients presenting to an ambulatory clinic in New York City. Significantly, none of these patients were recognized to be survivors of torture by their treating physicians. In addition, Eisenman et al.[16] surveyed 638 Latino adult patients in 3 community-based primary care clinics. Fifty-four percent reported political violence, and 8% reported torture. Those exposed to political violence had higher rates of physical and psychological problems compared with those not exposed to political violence. In Eisenman's[16] study, only 3% of the patients who had experienced political violence reported telling any clinician, and none reported that their current physician asked about a history of political violence.

These studies all have relevance to primary care physicians (PCPs) serving immigrant populations. However, despite these important data, it is unclear how well these studies are replicable across a broad spectrum of immigrant primary care populations. Insufficient data have been presented on demographic factors that can be used by PCPs to identify

None of the authors have any conflicts of interest to declare.

Address correspondence and requests for reprints to Dr. Crosby: Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, ACC-5, 850 Harrison Avenue, Boston, MA 02128 (e-mail: scrosby@bu.edu).

Manuscript received July 29, 2005
Initial editorial decision October 28, 2005
Final acceptance March 14, 2006

high-risk groups and guide screening practices in the primary care setting.

The primary objectives of this study are to determine the prevalence of torture among foreign-born patients presenting to an inner city primary care clinic in Boston and to evaluate whether or not patients have discussed their history of torture with their physicians. A secondary objective was to evaluate demographic variables (age, sex, duration in the United States, and region of origin) as associated factors with having a history of torture.

METHODS

This is a survey of foreign-born patients presenting from July 2003 to August 2004 to the Boston Medical Center Primary Clinic, a large urban medical center. On days covered by project staff, study interviewers screened all patients presenting for care through registration data or interview. Foreign-born patients were asked to participate in a survey while waiting to see their primary care providers. Inclusion criteria were as follows: age ≥ 18 years, foreign-born (outside of the United States and territories), and presenting to the Primary Care Clinic at Boston Medical Center. Patients were excluded if they were unwilling to participate in the survey. In addition, subjects were not able to participate if a professional interpreter was not available. Patients referred to the Primary Care Clinic by the Boston Center for Refugee Health and Human Rights, who were known to be survivors of torture, were also excluded from this survey.

Data Collection

After obtaining informed consent, a research associate interviewed subjects using a structured instrument to ascertain baseline information including the following: age, sex, country of origin, year of arrival in the United States, and to ascertain whether the subject had a history of torture. Questions on personal and family exposure to torture were modified from Eisenman et al.[15,17] While our wording differs, Montgomery and Foldspang[18] found a sensitivity of 82% and a specificity of 92% on comparing self-reported exposure with torture on a structured interview with in-depth psychological assessment. We asked, "Were you ever harmed or threatened by the following: government, police, military or rebel soldiers?" Patients who endorsed this question were then asked, "Some people in your situation have experienced torture. Has that ever happened to you or your family? Please explain." Patients were considered to have either personal or family exposure to torture if they answered affirmatively to both questions and if the explanation given was consistent with the UN C.A.T. definition of torture, as described above. Subjects who identified trauma at the hands of robbers or bandits, or whose history of trauma was not clearly government sponsored or inflicted by a group that the government was not willing or was unable to control, were not defined as conforming to the UN definition of torture.

Patients reporting a history of torture were referred with their consent to the Boston Center for Refugee Health and Human Rights, a comprehensive program that cares for survivors of trauma and torture at Boston Medical Center. No information was retained for subjects who refused to participate. The study was approved by the Institutional Review Board at Boston Medical Center.

Statistical Analysis

The percent of patients reporting a personal or family history of torture is described through exact 95% confidence intervals (CI) for a binomial percentage. To evaluate regional variations, subjects' countries of origin were categorized as follows: (1) Central America, South America, and the Caribbean, (2) Asia, or (3) Africa. The percent of patients with a history of torture is compared across demographic groups through the χ^2 test of independence (for nominal characteristics such as sex, area of origin) and the χ^2 test for linear trend in percentages (for ordinal categorizations [quintiles] of age and years in the United States). Results significant at a 2-tailed $P<.05$ level are considered statistically significant. Computations were performed using STATA, version 8 (Stata, College Station, TX).[19]

RESULTS

Of the 453 identified as potentially foreign-born based on registration data, 145 (32%) were excluded due to being born in the United States or territories. Of the 308 eligible patients, 78 (25%) were not asked to participate owing to the lack of a professional translator. Of the 230 eligible patients asked to participate, 142 (62% of the 230 asked, 46% of 308 eligible) consented to participate in the survey. This sample had a mean age of 47 years (range 19 to 76), was mostly female (58%, 82/142), had been in the United States for an average of 14 years (range 1 month to 53 years), and came from 35 different countries (Table 1). Fully, 16% (23/142, 95% CI 11% to 23%) reported a history of being personally tortured or having a family member tortured. Among these patients, 9 of 23 (39%) reported both personal and family member experience of torture, 9 (39%) of 23 reported only a history of being personally tortured, and 5 (22%) of 23 reported only a history of torture experienced by a family member. Among the 18 patients who reported personal experience of torture, the UN definition of torture was met in 16 (89%) of these cases to reveal a prevalence of 11% (16/142, 95% CI 7% to 18%). Among the 5 patients who reported only a history of torture experienced by a family member, the UN definition of torture was met in 4 (80%) of these cases.

Most patients (15/23, 67%) reported discussing their experience of torture with people in the United States. However, only 39% (9/23) had ever discussed their experience of torture with a health care provider and 8 (33%) of 23 reported that this survey was their first disclosure to anyone in the United States.

Subjects in the United States for a shorter period of time had a significantly higher rate of reporting a history of personal or family torture than subjects who had been in the United States longer ($P<.01$) (Fig. 2). For example, subjects in the United States for ≤ 3.5 years (lowest quintile) had a 38% (95% CI 19% to 57%) rate of personal or family torture and subjects in the United States for >30 years (highest quintile) had a 4% (95% CI 0% to 12%) rate of personal or family torture.

Significant regional variation is apparent. Whereas 6% (95% CI 1% to 11%) of subjects from Central America, South America, and the Caribbean reported a history of personal or family torture, 18% of subjects (95% CI 0% to 45%) from Asia, and fully 41% of subjects (95% CI 24% to 57%) from Africa reported a history of personal or family torture ($P<.001$).

Table 1. Demographics of Study Population with a History of Torture Exposure

Patient Group	% (n's) With A History of Torture Exposure	P Value*
Overall sample	16 (23/142)	—
Sex		.86
Males	17 (10/60)	
Females	16 (13/82)	
Age (quintiles)		.70
19 to 32	21 (6/29)	
33 to 45	6 (2/31)	
46 to 50	22 (6/27)	
51 to 60	26 (8/31)	
61to 76	4 (1/24)	
Years in the United States (quintiles)		<.01
<3.5	38 (11/29)	
3.5 to 9	14 (4/29)	
9.1 to 15.5	11 (3/27)	
15.6 to 22	13 (4/30)	
23 to 53	4 (1/26)	
Area of origin		<.01
Central America, South America, and the Caribbean	6 (6/94)	
Asia	18 (2/11)	
Africa	41 (15/37)	

P-value comparing the percent with a history of torture exposure across patient groups via the χ^2 test of independence for nominal variables (sex, area of origin) and the χ^2 test for trend for ordinal variables (age, years in the United States).

There was no statistical trend relating the rate of reporting a history of torture with subjects' age (P=.70) or sex (P=.86).

DISCUSSION

We found that 11% (16/142, 95% CI 7% to 18%) of participants reported a history of being personally tortured in a manner that met the UN definition of torture. Two associated factors for higher rates of torture were region of origin (Africa) and having been in the United States for a shorter period of time. However, the observation that subjects in the United States for a shorter period of time were more likely to be tortured is likely due to the fact that the duration for which subjects had been in the United States was associated with the region of origin (P<.01). For example, while half of all subjects who had been in the United States for less than 3.3 years (first quintile) were of African origin, only 5% of those who had been in the United States for greater than 31 years (fifth quintile) were of African origin.

Our finding that 39% of subjects reporting torture had disclosed this history to a health care provider is extraordinarily high when compared with the existing literature. Eisenman et al.[16] reported only 3% ever telling a clinician about political violence, and 0% reported that their current clinician asked about political violence. There are several hypotheses for this unexpected finding. The greater Boston area is a resettlement site for many immigrant communities, and there are multiple organizations that serve the needs of these populations, including specific attention to the needs of survivors of torture and human rights abuses. Some of these organizations are Physicians for Human Rights, The International Institute of Boston, and The Boston Center for Refugee Health and Human Rights. The presence of, and awareness about these organizations in communities may increase awareness of torture among both patients and providers. We have also provided training about caring for survivors of torture to health care providers in multiple clinical sites in the greater Boston area, directly raising awareness of this topic among primary care providers.

Our findings on the prevalence of torture in African primary care patients (41%) is similar to that reported by Jaranson et al.[14] in a community-based population of East African refugees (25% to 69%). The prevalence of torture in subjects from Central and South America and the Caribbean was 6%, compared with 8% reported by Eisenman[16] in Latino primary care patients. Our prevalence of torture was higher than Eisenman[15] reported in a primary care sample of 121 patients in NYC (6.6% vs 11%). It is possible that differences in subjects' countries of origin may account for this difference.

Several important limitations should be considered when interpreting these data. Few studies have examined the validity of self-reported history of torture. The reference that we cite (Montgomery and Foldspang[18]) reports good validity of a personal report of torture to a clinical determination of torture as defined by the Tokyo Declaration. The validity of our questions in determining torture as established by the UN Convention has not been directly established. We did not evaluate socioeconomic status and it is possible that the prevalence of torture may vary with socioeconomic status. We excluded subjects who were known to the Boston Center for Refugee Health and Human Rights, which is based at Boston Medical Center. This was appropriate, as two-thirds of patients seen at Boston Center for Refugee Health and Human Rights have been referred for care at the Boston Center for Refugee Health and Human Rights by outside sources (attorneys and resettlement agencies) and including such patients would inappropriately enrich our sample. Conversely, as one-third of the patients at the Boston Center for Refugee Health and Human Rights are referred from within Boston Medical Center excluding such patients will deplete the sample and yield an underestimation of the true prevalence of torture in the clinic population. The magnitude of this effect, however, is quite small due to the relative sizes of the Primary Care Clinics at Boston Medical Center (>20,000 unique patients/year, >33% foreign-born) and The Boston Center for Refugee Health and Human Rights (359 patients last year). Inclusion of subjects removed from the sampling pool due to internal referral would have increased our prevalence estimate from 11% (95% CI 7% to 18%) to 13%. It is also possible that the prevalence we report is an overestimate due to the presence of a specialized center for survivors of torture within the institution. Although patients of the Boston Center for Refugee Health and Human Rights were not included in this survey, relatives and acquaintances of such patients, who themselves likely would have a high rate of exposure to violence, may have been drawn to the institution for this reason. It is possible that overestimation of prevalence rates could be due to high utilization of health care services, as seen with domestic violence populations.[20] Weighting the subject selection process by health care utilization was not possible because the surveys were administered anonymously.

We collected no information on nonparticipants. While it is possible that some survivors of torture would choose to

avoid this study because of fear and stigma, we are unable to confirm this conjecture. While the interview instrument used in this study has been validated previously, we were not able to confirm that the instrument operates effectively across the many cultures represented by the participants in this study. We did not ask whether participants were refugees or asylum seekers. This might have provided useful data for primary care providers, as immigration status may be an important easily identifiable associated factor. It is unclear whether the point prevalence we report can be generalized to the foreign-born patients in other primary care practice settings. It is important to realize that the actual point prevalence of torture will vary among clinical practices depending on the proportion of foreign-born patients from different countries and various parts of the world. In addition, prevalence will change over time with country-specific situations, such as wars, oppressive leaders, and politics.

The high prevalence of torture in foreign-born primary care patients highlights the importance of clinical interview and exam skills for primary care providers to identify patients who have experienced torture or potential vicarious trauma. Lack of recognition and treatment may result in significant psychological and physical sequelae.

The clinical presentation of survivors of torture has been shown to be highly varied.[21,22] For example, patients may present to their primary care providers with chronic headache or organic brain syndromes due to head trauma, nerve palsies due to suspension, genital pain due to genital torture, foot pain due to falanga, chest pain, abdominal pain, hearing loss, or dental trauma.[10] Often, there are no telltale marks, and physicians are not generally trained to detect the specific sequelae of torture.[23] In addition, mental illness, including posttraumatic stress disorder, depression, anxiety, adjustment disorder, and psychosomatic illness, are all prevalent in torture survivors, but may not be easily diagnosed in the absence of an appropriate history.[24,25] This lack of recognition may result in unnecessary investigations, or labeling patients as "hypochondriacs." At worse, the lack of a history will result in failure to get treatment and prolongation of suffering.

Our results showed that this survey was their first disclosure to anyone in the United States of being personally tortured or having a family member tortured for one-third of the subjects. Survivors of torture may try to avoid medical care due to fear of further persecution, deportation, and humiliation. They may not identify themselves to physicians, even when seeking services. Such patients may harbor a basic mistrust of physicians and may be reluctant to tell their caregivers about their history. For communities without dedicated immigrant and refugee services, providers may need more diligence to elicit a torture history in foreign-born patients.

In our population, variables associated with a higher risk of torture were recent arrival to the United States, and immigration from the African and Asian continents. We believe that clinicians should routinely ask patients from the African and Asian continents who are recent arrivals to the United States about a history of torture. Further studies of large numbers of foreign-born patients across a broad spectrum of primary care practices are needed to stratify risk factors for torture in clinical settings, and to provide further guidance to clinicians for torture history screening in primary care settings.

Screening programs, educational initiatives, and interventions for treatment should be further studied. Physicians seeing immigrant patients in their practices should be familiar with the general backgrounds of their patients' countries of origin, common medical and psychological sequelae of torture, and should be knowledgeable about specialized referral centers for survivors of torture. The Boston Center for Refugee Health and Human Rights has a web course available for providers on caring for survivors of torture (www.bcrhhr.org). Information about specialized treatment centers for survivors of torture can be found at The National Consortium of Torture Treatment Programs Web site.[26]

We thank Alison Sullivan, Jamayla Culpepper, Simran Kaur, and Sharl Mendler for assistance in administration of the survey. We thank Jennifer Carroll, MD, for contributions in study design.

REFERENCES

1. Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, Office of the High Commission for Human Rights, United Nations Web site. Available at: http://www.ohchr.org/english/countries/ratification/9.htm. Accessed March 1, 2005.
2. **American Medical Association.** Code of Medical Ethics Current Opinions with Annotations, 2004–2005. American Medical Association; 2005.
3. The role of the physician and the medical profession in the prevention of international torture and in the treatment of its survivors. American College of Physicians. Ann Intern Med. 1995;122:607–13.
4. United Nations Convention against Torture and Other Cruel, Inhumane, or Degrading Treatment or Punishment, New York, 1984.
5. United Nations High Commissioner for Refugees, Statistics. Available at: http://www.unhcr.org/.Accessed March 18, 2005.
6. **Rasmussen OV.** Medical aspects of torture. Dan Med Bull. 1990;37, 1–88(Suppl 1):1–88.
7. **Basoglu M, Paker M, Ozmen E, Tasdemir O, Sahin D.** Factors related to long-term traumatic stress responses in survivors of torture in Turkey. JAMA. 1994;272:357–63.
8. **Goldfeld AE, Mollica RF, Pesavento BH, Faraone SV.** The physical and psychological sequelae of torture. Symptomatology and diagnosis. JAMA. 1988;259:2725–9.
9. **Moisander PA, Edston E.** Torture and its sequel—a comparison between victims from six countries. Forensic Sci Int. 2003;137:133–40.
10. **Moreno A, Grodin MA.** Torture and its neurological sequelae. Spinal Cord. 2002;40:213–23.
11. **Nice DS, Garland CF, Hilton SM, Baggett JC, Mitchell RE.** Long-term health outcomes and medical effects of torture among US Navy prisoners of war in Vietnam. JAMA. 1996;276:375–81.
12. **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.
13. **Marshall GN, Schell TL, Elliott MN, Berthold SM, Chun CA.** Mental health of Cambodian refugees 2 decades after resettlement in the United States. JAMA. 2005;294:571–9.
14. **Jaranson JM, Butcher J, Halcon L, et al.** Somali and Oromo refugees: correlates of torture and trauma history. Am J Public Health. 2004;94: 591–8.
15. **Eisenman DP, Keller AS, Kim G.** Survivors of torture in a general medical setting: how often have patients been tortured, and how often is it missed? West J Med. 2000;172:301–4.
16. **Eisenman DP, Gelberg L, Liu H, Shapiro MF.** 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. 2003;290:627–34.
17. **Eisenman DP, Tracy K, Keller AS.** Development of a screening questionnaire for detecting victims of torture in medical settings. J Gen Intern Med. 1988;13:126.
18. **Montgomery E, Foldspang A.** Criterion-related validity of screening for exposure to torture. Dan Med Bull. 1994;41:588–91.

19. **StataCorp.** Stata Statistical Software: Release 7.0. College Station, TX: Stata Corporation; 2001.
20. **Ulrich YC, Cain KC, Sugg NK, Rivara FP, Rubanowice DM, Thompson RS.** Medical care utilization patterns in women with diagnosed domestic violence. Am J Prev Med. 2003;24:9–15.
21. **Hougen HP, Kelstrup J, Petersen HD, Rasmussen OV.** Sequelae to torture. A controlled study of torture victims living in exile. Forensic Sci Int. 1988;36:153–60.
22. **Rasmussen OV, Lunde I.** Evaluation of investigation of 200 torture victims. Dan Med Bull. 1980;27:241–3.
23. **Weinstein HM, Dansky L, Iacopino V.** Torture and war trauma survivors in primary care practice. West J Med. 1996;165:112–8.
24. **Goldfeld AE, Mollica RF, Pesavento BH, Faraone SV.** The physical and psychological sequelae of torture. Symptomatology and diagnosis. JAMA. 1988;259:2725–9.
25. **Holtan N, Antolak K, Johnson DR, Ide L, Jaranson J, Ta K.** Unrecognized torture affects the health of refugees. Minn Med. 2002;85: 35–9.
26. **National Consortium of Torture Treatment Programs.** Available at: http://ncttp.westside.com/default.view. Accessed March 29, 2005.

Accountability in Research, 12:33–45,

ISSN: 0898-9621 print
DOI: 10.1080/08989620590918916

EXCLUDING PARTICULAR INFORMATION FROM CONSENT FORMS

NEAL DICKERT, B.A.

Department of Health Policy and Management, Bloomberg School of Public Health and Phoebe R. Berman Bioethics Institute, The John Hopkins University

NANCY KASS, Sc.D.

Phoebe R. Berman Professor of Bioehtics and Public Health, Department of Health and Policy and Management, Bloomberg School of Public Health and Phoebe R. Berman Bioethics Institute, The John Hopkins University

MICHAEL PAASCHE-ORLOW, M.D., M.P.H.

Assistant Professor, Department of Medicine,
Boston University School of Medicine

HOLLY TAYLOR, M.P.H., Ph.D.

Assistant Professor, Department of Health Policy and Management,
Bloomberg School of Public Health and Phoebe R. Berman
Bioethics Institute, The Johns Hopkins Universtiy

Although the informed consent process is crucial to protecting human research subjects, there are cases when particular information within the consent form may present risks to those subjects. In this paper, we examine a case in which including the sponsor's name on the consent form may allow the form to serve as a surrogate for subjects' HIV status.

There is no literature addressing the ethical acceptability of excluding particular information from consent forms, and there exists little regulatory guidance on this issue. We argue that excluding information from the consent form is, in fact, obligatory when that information is disclosed orally during the

We would like to thank the investigators who brought this case to our attention, who wish to remain anonymous. We also would like to thank the members of the Program in Research Ethics (PRE) of the Johns Hopkins University Phoebe R. Berman Bioethics Institute, as well as David Wendler for thoughtful contributions to this manuscript.

Neal Dickert was primarily responsible for drafting the paper. Nancy Kass, Micahel Paasche-Orlow, and Holly Taylor worked with the primary author in developing and editing the paper.

Address correspondence to Neal Dickert, B.A., Department of Health Policy and Management, Phoebe R. Berman Bioethics Institute, John Hopkins University Bloomberg School of Public Health, Hampton House 348, 624 N. Broadway, Baltimore, MD 21205. E-mail: ndicker@jhmi.edu.

consent process but its presence on the form poses risks to the subjects the consent process is designed to protect. Further, we argue that the regulations ought to be amended to reflect this obligation.

***Keywords**: informed consent, consent form, international research, ethics, human subjects research, confidentiality*

The doctrine of informed consent is well-established as a cornerstone of biomedical research ethics, based on an assumption that informing potential subjects and emphasizing voluntary decision making respects and protects their autonomy. Although the consent form is designed to facilitate the process of informed consent and thus advance the interests of potential research subjects, difficulties may arise when the consent form itself poses a risk to subjects. It is particularly problematic that existing national and international regulations do not adequately address this potential conflict.

Consider the following actual case. Investigators from a U.S. institution are conducting a trial in Africa examining the potential benefits of a preventive intervention on transmission of sexually transmitted diseases. The study will assess the efficacy of this intervention in both HIV-infected and HIV-uninfected subjects, and study procedures will be identical for both populations. Although the investigators consider the study of both populations to be essentially one study, there are different funding sources for studying the infected and uninfected subjects. In keeping with the Council for International Organizations of Medical Sciences (CIOMS) guideline 5, it is the policy of one sponsor that the investigators must disclose the identity of the sponsor to study participants (CIOMS, 2002). Because that sponsor wants its name to appear only on the forms of the subjects whose enrollment it is funding, the sponsor wants separate consent forms for the two populations, with the forms differing only in the sponsor that is named.

The investigators plan to obtain written informed consent from each subject and to provide full oral disclosure of all important aspects of the study to participants, including a discussion of their HIV status and the identity of the sponsor funding the arm of the study in which participants are enrolled. Not including subjects' HIV status on consent forms is relatively standard practice. However, the investigators in this case also wish to omit the name of the sponsor of the study from the form and to use the same consent form for the infected and uninfected subjects. They

are concerned that using two different forms including the name of the sponsor will allow the forms to serve as a surrogate for subjects' HIV status. Given the very close living conditions in this particular population, investigators worry that individuals may not be able to keep consent forms private. Having worked in this community for years, investigators believe that the presence of information on the consent form that could identify particular subjects as having HIV could lead to significant discrimination against those individuals. Based on their experience with this community, they also are concerned that uninfected subjects might use their forms (containing the name of the sponsor associated with the study of *uninfected* subjects) as "certificates of negativity" in order to engage in potentially high-risk sexual activity for themselves and partners. In other words, they are concerned that these forms, if misused or misunderstood, will facilitate activities that may increase risks of HIV transmission.

The issue at hand is whether investigators ought to be permitted, or even required, to exclude a particular piece of information from the consent form when including that information poses a risk to the subject and, potentially, to the community at large. No information will be withheld from participants during the consent discussion. At issue is only what appears on the consent document that is signed by participants and given to them for future reference.

We are aware of no literature or regulation that addresses this question directly. We will briefly review existing regulations and guidelines, as well as the body of literature on alterations of the consent process, in an effort to gain any insight or guidance regarding this case and others like it, in which particular pieces of information pose risks when present on consent forms. Following this review, we argue that this type of situation indeed represents a gap in existing regulations, and that excluding potentially harmful information from the consent form is sometimes not only ethically acceptable but, in fact, ethically obligatory.

Regulatory Guidance

The Common Rule (U.S. 45 CFR 46) does not directly address the inclusion of the sponsor or selective omission of particular pieces of information from consent forms. The regulations explicitly allow for an Institutional Review Board (IRB) to waive the requirement for a signed form altogether if "the only record

linking the subjects and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality" (45 CFR 46.117.c). In this case, the subject should be given the option of signing a consent form but may be allowed to decline. The IRB also may waive the requirement for a signed form if "the research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context" (45 CFR 46.117.c). As the study at hand involves a minor surgical intervention and monitoring for sexually transmitted disease, it poses greater than minimal risk and thus is not eligible for waiver of a signed consent form. Furthermore, although the risk presented by a signed form containing the name of the sponsor may be one primarily of confidentiality, the risk does not derive from the form itself. Rather, the risk to subjects derives from the presence of the sponsor's name, which could serve as a surrogate for the subject's HIV status. Investigators express no desire to waive the requirement for a signed form altogether.

IRBs may approve "a consent procedure which does not include, or which alters, some or all of the elements of informed consent" (45 CFR 46.116.c) under the following conditions:

1. The research involves no more than minimal risk to subjects;
2. The alteration must not affect the rights and welfare of the subjects;
3. The research could not practicably be carried out without the waiver or alteration; and
4. Whenever appropriate, the subjects will be provided with additional pertinent information after participation (45 CFR 46.116.c).

Again, the proposed study does not meet these criteria due to the fact that it poses greater than minimal risk to the subjects. While we believe the *alteration* proposed (removing the name of the sponsor from the consent form) poses no greater than minimal risk, the regulations appear to require that the *study* itself not pose more than minimal risk in order for any alterations to be allowed.

International guidelines are silent on the issue of particular exclusions from the consent form. As mentioned earlier, guideline 5 of the CIOMS guidelines includes the sponsor of the study

in the list of information that ought to be included on the consent form. However, the commentary on guideline 4 discusses circumstances in which the requirement for a signed consent form may be waived. Similar to the Common Rule, it states that waivers of signed consent forms may be "approved when existence of a signed consent form would be an unjustified threat to the subject's confidentiality" (CIOMS, 2002).

Relevant Literature

We are aware of no literature that directly addresses the issue of excluding particular information from consent forms. Perhaps most closely related are papers discussing the acceptability of withholding relevant information from participants during the consent process (deception) and of waiving the requirement for signed or written consent.

If deception is necessary in order to answer an important research question, it is generally agreed that investigators should provide research participants with the maximal amount of information possible that may be material to their decision (Murray, 1980; Wendler, 1996). Although withholding the sponsor's identity from subjects altogether may be seen as deceptive, investigators in this case fully intend to disclose sponsors' identities orally. The study is not deceptive, and this body of literature is relevant only in that it may suggest a need to ensure that investigators do inform subjects of the sponsors' names to the extent that this information is material to the subjects' decision about participation.

More germane to this topic is work discussing situations in which signed or written consent forms may be omitted (Appelbaum, Lidz, and Meisel, 1987; Levine, 1986; Wendler and Rackoff, 2001). This literature is consistent with the provisions within the CIOMS guidelines (2001) and the federal regulations (45 CFR 46) that allow written consent to be waived. In some settings, written consent is impractical, and oral consent (e.g., for telephone surveys) or no consent (e.g., chart review studies) are considered acceptable. In other settings, written consent is avoided because participants are illiterate, and using a written form is at best unhelpful and at worst disrespectful. Wendler and Rackoff (2001), for example, discussed a wide range of situations in which they believe that the requirement for a signature, in particular, is

inappropriate. Such cases may involve fears of persecution, differing cultural views of signatures, physical inability to perform the act of signing, false impressions of subjects on the binding nature of research participation, and risks to confidentiality. They argued that in each of these cases, the signature itself is not an essential part of the consent process, and numerous alternatives exist to the standard signed consent form that allow for all essential elements of the informed consent process to occur. While these authors did not discuss excluding particular information on the form, their arguments for waiving the need for a signature at times when it undermines the interests of subjects would seem to support this practice.

Other scholars also have argued that the need for a signed, written consent form can be outweighed by important subject interests, particularly when the form threatens confidentiality (Appelbaum et al., 1987; Levine, 1986). In short, there is precedent for and arguments to support avoiding the use of consent forms when information could be stigmatizing and risky to participants, and for avoiding the requirement for a *signed* consent form when the signature runs contrary to subjects' interests.

The recent report of the National Bioethics Advisory Commission (NBAC) on international research ethics also emphasizes that the *prima facie* need for documented informed consent must be considered in light of the specific circumstances of the research study being performed.

> Federal policy should require investigators to document that they have obtained voluntary informed consent, but should be flexible with respect to the form of such documentation. Especially when individuals can easily refuse or discontinue participation or when signed forms might threaten confidentiality, IRBs should permit investigators to use other means of identifying that informed consent has been obtained. (NBAC, 2001)

The report goes on to assert that there exists no case law in which written informed consent has been required and recommends that measures to increase protections for privacy and confidentiality are needed.

Although it can be an important part of the consent process, the consent form is not the whole process. Research subjects have important interests that may conflict with and at times override the need for a signed, written consent form.

The Value of the Consent Form

Given the above discussion, it is important to consider the purpose of the consent form and the ethical value it possesses. Some sense of the ethical value of the form will allow us to make an assessment about whether countervailing moral considerations might ever warrant omitting particular information from the form. The most obvious, and perhaps primary, purpose of the form is legal. By documenting what participants were told about the study and that they voluntarily agreed to participate, a signed document potentially offers some protection to the investigator, institution, and participants (45 CFR 46.117; Levine, 1986; Macklin, 1999; Meisel and Kuczewski, 1996). Second, the consent form may serve as an "adjunct" to the consent process (Appelbaum et al., 1987). It may be given to participants prior to talking with the investigator in order to give them information ahead of time (Neptune, 1996), or investigators may refer to the document during a consent discussion in order to ensure that all relevant topics are covered. For participants who decide to enroll in research, the form may be a useful document to which they may refer over the course of a study in order to remember study details and answer questions (ACHRE 1996). Finally, the consent form gives IRBs a window into the consent process for the purpose of evaluating its adequacy (ACHRE, 1996).

Although the signed, written consent form is the paradigmatic way to satisfy legal demands for documentation, other forms of documentation exist, and they have not been shown to be any less valuable legally. In fact, there is no case law or data to establish the primacy of signed, written documentation over other methods, such as video or audio documentation of the participant's agreement. Where signed consent is avoided because participants are illiterate or object to the signing of forms, or where a written consent form poses a risk to confidentiality, it is not uncommon, in our experience, for investigators themselves to sign a document attesting that they have disclosed relevant information and answered questions, and that the participant joined voluntarily. Indeed, the Food and Drug Administration (FDA) explicitly advises the use of video documentation in certain settings (U.S. Food and Drug Administration, 1998).

Despite near universal use of written consent forms in biomedical research and a body of literature studying the process of informed consent (Sugarman et al., 1999), the effectiveness of consent forms in facilitating the consent process is largely unknown. A recent study demonstrated that consent forms for Phase I oncology trials appeared clear in describing the risks and benefits of these studies (Horng et al., 2002). The authors thus challenge the view that consent forms are responsible for the significant misunderstandings participants are known to have about the nature of Phase I studies (Daugherty et al., 2000; Rodenhuis et al., 1984). Other studies have demonstrated marginal benefits at best when looking at the utility of written information at all (Layton and Korsen, 1994) and written information provided at different time points (Neptune, 1996).

On the other hand, early work on providing written information for cancer treatment suggested that patients had a more complete understanding of their diagnosis and treatment options when given written information as part of the consent for treatment (Morrow, Gootnick, and Scmale, 1978). Similarly, there are anecdotal reports that participants like having forms to which they can refer later (ACHRE, 1996), and some believe they convey the voluntary nature of research (Kass, Maman, and Atkinson, in review). In addition, several studies have soundly demonstrated that consent forms do not achieve specified goals of readability and comprehensibility (Paasche-Orlow, Taylor, and Brancati, 2003; Silva and Sorrell, 1988), suggesting that their potential utility may be far from maximized. Although some studies have demonstrated increases in subject understanding as a result of consent form improvements, these increases generally have been small (Flory and Emanuel, 2004; Joffe et al., 2001). Some of the same studies and others have highlighted deeper and more substantial problems with the consent process in general (Joffe et al., 2001; Titus and Keane, 1996).

Although the data on the utility and value of written consent forms are, to some degree, conflicting, we can at least conclude that the value of consent forms does not appear to be overwhelming. That is, consent forms are not so important that concern for the form overrides other important interests that subjects may have.

Might it be particularly valuable to subjects to have the name of the sponsor on the consent form—more valuable than for

other types of information? It is difficult to see why this information would have any special reason to be included. From a legal perspective, it seems that documenting oral disclosure of the sponsor would be sufficient. Similarly, subjects are unlikely to need to refer back to the name of the sponsor or reread information about the sponsor. If they do, they can easily contact investigators if they forget the sponsor's identity. In contrast, there may be clear benefits to subjects to have written information on potential adverse effects of an intervention or contact information if such an event should occur. In short, the primary benefit of disclosing the research sponsor to human subjects is to achieve transparency and to allow the subjects to take into account who is sponsoring the research study. There seems to be no special role for a consent form in achieving this benefit.

When Can Information Be Excluded?

In cases where participants could potentially face significant risk by including information such as the name of the sponsor on the consent form, the moral imperative to reduce harms to research subjects seems to trump any requirement for complete documentation on the form. In this case, the risks posed to HIV-infected individuals by having consent forms serve as a surrogate for their HIV status is a significant issue. Similar risks are likely to be present in studies of other diseases with potential stigma, such as mental illnesses and other sexually transmitted diseases, particularly in settings where consent forms are not private documents. Which pieces of information pose risk also will vary in different cases. In this case, it is the name of the sponsor; in others, different pieces of information could place subjects at risk.

The investigators' concern that HIV-uninfected individuals' consent forms could serve as "certificates of negativity" and facilitate or encourage risky behavior is a more complicated concern about communal risks, as opposed to risks to individual subjects. The degree to which IRBs ought to protect community members from risky behavior on the part of research subjects is a topic that warrants further discussion but is outside the scope of this paper.

Given the significant concerns at stake, a simple solution exists that achieves the goals of all parties involved and allows for all of the goals of the consent form to be fulfilled. First, investigators should assure the IRB that they will continue to provide complete oral disclosure of all relevant aspects of the study to each subject. Second, the investigators should simply state on the consent form that the name of the sponsor and the subjects' HIV status will be discussed with each subject in person, but that this information does not appear on the form in an effort to protect the subjects' confidentiality.

It is crucial that important interests of participants not be superceded by a consent form orthodoxy that rigidly requires exhaustive consent forms in all cases. We believe that, in this case and others like it, it is not only permissible for the investigator to omit the sponsor from the form; it is obligatory to omit this information.

In fact, it would represent a failure of the IRB system not to require investigators to omit this potentially damaging information that is already fully disclosed. Founded on the principle of local review, the IRB (both in the host country and in the United States) has an obligation to consider the special circumstances of the subject population and the community in which the subjects live, including local views and practices relevant to confidentiality (Puglisi, 2000). It is important to note that it is not easy for IRBs in sponsoring countries to be fully aware of cultural or social differences in host countries, and there will certainly be gaps. However, sponsoring country IRBs should attempt to understand host communities as well as possible and attend carefully to concerns of which they are aware. Otherwise, IRBs abrogate their commitment to conduct rigorous and sensitive local review. Furthermore, IRBs are charged with the duty of identifying and balancing competing interests that are at stake in particular protocols. To weigh subjects' interests in confidentiality with the importance of including the sponsor's name on the consent form is entirely within the purview of IRBs.

Sponsors of research, particularly in the international setting, also have an obligation to be sensitive to the realities of the communities in which they sponsor research, as well as to allow IRBs and investigators the latitude necessary to protect subjects' interests. Modifying the consent process in the way that we propose in no way involves lowering or altering the ethical standards

to which sponsors desire to adhere. Rather, these modifications are required in order to realize those ethical goals.

Conclusion

Investigators and IRBs, in this case and in other, similar cases, have an obligation to place important interests of the participants over any need to include the name of the sponsor on a written consent form. As a result of this case study, we suggest that regulations and guidelines ought to be amended to allow explicitly the exclusion of particular information from the consent form when including that information runs counter to overriding interests of the participants and their community, and when omitting that information poses no significant risks to subjects. In crafting such modifications to the regulations, and in implementing any changes, it will be important to pay close attention to the types of information that have special reason to be documented explicitly. Potential adverse effects, for example, have special reason to be given to subjects in written form; there are surely others, as well.

This argument is not about withholding information; it is about what ought to appear on the consent form. The form ought to mention that the name of the sponsor is omitted intentionally, and investigators should discuss the reasons for this omission with subjects. However, we must be vigilant in preventing consent form orthodoxy from compromising the interests of the individuals the informed consent process is expressly designed to protect.

References

Advisory Committee on Human Radiation Experiments, (ACHRE) (1996). Research Proposal Review Project. In *Final Report of the Advisory Committee on Human Radiation Experiments* (New York: Oxford University Press).

Advisory Committee on Human Radiation Experiments (ACHRE). (1996). Subject Interview Study. In *Final Report of the Advisory Committee on Human Radiation Experiments* (New York: Oxford University Press).

Appelbaum, P. S., Lidz, C. W., and Meisel, A. (1987). *Informed Consent: Legal Theory and Clinical Practice* (New York: Oxford University Press).

Council for International Organizations of Medical Sciences (CIOMS). (2002). *International Ethical Guidelines for Biomedical Research Involving Human Subjects* (Geneva: Author).

Daugherty, C. K., Banik, D. M., Janish, L., and Ratain, M. J. (2000). Quantitative Analysis of Ethical Issues in Phase I Trials: A Survey Interview Study of 144 Advanced Cancer Patients. *IRB*, 22(3): 6–14.

Flory, J. S., and Emanuel, E. J. (2004).Interventions to Improve Research Participants' Understanding in Informed Consent for Research. *Journal of the American Medical Association*, 292: 1593–1601.

Horng, S. et al. (2002). Descriptions of Benefits and Risks in Consent Forms for Phase 1 Oncology Trials. *New England Journal of Medicine*, 347(26): 2134–2140.

Joffe, S. et al. (2001) Quality of Informed Consent in Cancer Clinical Trials: A Cross-Sectional Survey. *Lancet*, 358(9295): 1772–1777.

Kass, N. E, Maman, S., and Atkinson, J. (*in review*) Is "Informed Choice" a Reality for Research Participation in Developing Countries? The Experience in Three International Research Studies.

Layton, S., and Korsen, J. (1994). Informed Consent in Oral and Maxillofacial Surgery: A Study of the Value of Written Warnings. *British Journal of Oral and Maxillofacial Surgery*, 32: 34–46.

Levine, R. J. (1986). *Ethics and Regulation of Clinical Research* (2nd ed.) (New Haven, CT: Yale University Press).

Macklin, R. (1999). Understanding Informed Consent. *Acta Oncologica*, 38: 83–87.

Meisel, A., and Kuczewski, M. (1996). Legal Myths About Informed Consent. *Archives of Internal Medicine*, 156: 2521–2526.

Morrow, G., Gootnick, J., and Scmale, A. (1978). A Simple Technique for Increasing Cancer Patients Knowledge of Informed Consent to Treatment. *Cancer*, 42: 793–799.

Murray, T. H. (1980). Commentary: Was This Deception Really Necessary? *IRB*, 2(10): 7–8.

National Bioethics Advisory Commission (NBAC). (2001). *Ethical and Policy Issues in Research Involving Human Participants: Report and Recommendations of the National Bioethics Advisory Commision* (Bethesda: U.S. Government).

Neptune, S. M. (1996). Take-Home Informed Consent for Intravenous Contrast Media. *Investigative Radiology*, 31(2): 109–113.

Paasche-Orlow, M. K., Taylor, H. A., and Brancati, F. L. (2003). Readability Standards for Informed-Consent Forms as Compared with Actual Readability. *New England Journal of Medicine*, 348(8): 721–726.

Puglisi, T. (2000). IRB Knowledge of Local Research Context. Available at <http://www.hhs.gov/ohrp/humansubjects/guidance/local.htm>. Accessed on 8/10/04.

Rodenhuis, S. et al. (1984). Patient Motivation and Informed Consent in a Phase I Study of an Anticancer Agent. *European Journal of Cancer Clinical Oncology*, 20: 457–462.

Silva, M. C., and Sorrell, J.M. (1988). Enhancing Comprehension of Information for Informed Consent. *IRB*, 10(1): 1–5.

Sugarman, J. et al. (1999). Empirical Research on Informed Consent: An Annotated Bibliography. *Hastings Center Report*, 29(1): S1–42.

Titus, S. L., and Keane, M. A. (1996). Do You Understand? An Ethical Assessment of Researchers' Description of the Consenting Process. *Journal of Clinical Ethics*, 7(1): 60–68.

U.S. Food and Drug Administration. (1998). Information Sheets: A Guide to Informed Consent. Available at <http://www.fda.gov/oc/ohrt/irbs/informedconsent.html>. Accessed on 8/23/04.

Wendler, D. (1996). Deception in Medical and Behavioral Research: Is It Ever Acceptable? *The Milbank Quarterly*, 74(1): 87–114.

Wendler, D., and Rackoff, J. E. (2001). Informed Consent and Respecting Autonomy: What's a Signature Got to Do With It? *IRB*, 23(3): 1–4.

Age and the Risk of Warfarin-Associated Hemorrhage: The Anticoagulation and Risk Factors In Atrial Fibrillation Study

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

OBJECTIVES: To assess whether older age is independently associated with hemorrhage risk in patients with atrial fibrillation, whether or not they are taking warfarin therapy.

DESIGN: Cohort study.

SETTING: Integrated healthcare delivery system.

PARTICIPANTS: Thirteen thousand five hundred fifty-nine adults with nonvalvular atrial fibrillation.

MEASUREMENTS: Patient data were collected from automated clinical and administrative databases using previously validated search algorithms. Medical charts were reviewed from patients hospitalized were for major hemorrhage (intracranial, fatal, requiring ≥2 units of transfused blood, or involving a critical anatomic site). Age was categorized into four categories (<60, 60–69, 70–79, and ≥80), and multivariable Poisson regression was used to assess whether major hemorrhage rates increased with age, stratified by warfarin use and adjusted for other clinical risk factors for hemorrhage.

RESULTS: A total of 170 major hemorrhages were identified during 15,300 person-years of warfarin therapy and 162 major hemorrhages during 15,530 person-years off warfarin therapy. Hemorrhage rates rose with older age, with an average increase in hemorrhage rate of 1.2 (95% confidence interval (CI) 1.0–1.4) per older age category in patients taking warfarin and 1.5 (95% CI = 1.3–1.8) in those not taking warfarin. Intracranial hemorrhage rates were significantly higher in those aged 80 and older (adjusted rate ratio = 1.8, 95% CI = 1.1–3.1 for those taking warfarin, adjusted rate ratio = 4.7, 95% CI = 2.4–9.2 for those not taking warfarin) than in those younger than 80.

CONCLUSION: Older age increases the risk of major hemorrhage, particularly intracranial hemorrhage, in patients with atrial fibrillation, whether or not they are taking warfarin. Hemorrhage rates were generally comparable with those reported in previous randomized trials, indicating that carefully monitored warfarin therapy can be used with reasonable safety in older patients. J Am Geriatr Soc 54:1231–1236, 2006.

Key words: aging; anticoagulation; hemorrhage; atrial fibrillation

Anticoagulation therapy with warfarin effectively reduces the risk of ischemic stroke associated with atrial fibrillation but also increases the risk for major hemorrhage.[1–4] Although previous randomized trials of warfarin for atrial fibrillation reported low rates of hemorrhage, these studies included few patients aged 80 and older and selected their patients more rigorously than in actual clinical practice.[1,5] It is uncertain whether the low bleeding rates observed in trial settings apply to patients with atrial fibrillation in usual clinical care. Prior studies have also provided conflicting evidence as to whether older age is an independent risk factor for warfarin-associated hemorrhage.[6–8] As increasing numbers of elderly patients take warfarin for atrial fibrillation,[9] more-precise data are needed on hemorrhage rates in the oldest patients.

Most observational studies of warfarin did not specifically address the risk of hemorrhage in patients with atrial fibrillation. Studies generally included patients taking warfarin for mechanical heart valves or venous thromboembolic diseases, conditions that may have different clinical characteristics and target anticoagulation intensities than atrial fibrillation. Previous studies also lacked sufficient outcome events to examine the association between age and different types of hemorrhage, an important concern because risk factors and consequences of hemorrhage differ between intracranial and extracranial hemorrhages.[10] To address these concerns, data from the large AnTicoagulation and Risk Factors In Atrial Fibrillation (ATRIA) cohort were analyzed to assess whether older age is independently

From the *University of California at San Francisco, San Francisco, California; †Division of Research, Kaiser Permanente of Northern California, Oakland, California; ‡General Internal Medicine and Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Massachusetts; and §Clinical Epidemiology Unit, Massachusetts General Hospital, Boston, Massachusetts.

Results from this study were presented at the 2004 American Heart Association Scientific Session.

Address correspondence to Margaret C. Fang, MD, MPH, University of California, San Francisco, 533 Parnassus Ave., Box 0131, San Francisco, CA 94143. E-mail: mfang@medicine.ucsf.edu

DOI: 10.1111/j.1532-5415.2006.00828.x

0002-8614/06/$15.00

associated with risk of intracranial and extracranial hemorrhage in patients with nonvalvular atrial fibrillation.

METHODS

ATRIA is a cohort study of 13,559 adults aged 18 and older with diagnosed atrial fibrillation and who received care within Kaiser Permanente of Northern California, a large integrated healthcare-delivery system. Details of the cohort assembly have been described previously.[11] Cohort members were identified by searching electronic inpatient, outpatient, and electrocardiographic databases for physician-assigned *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9) diagnoses of atrial fibrillation (427.31) between July 1, 1996, and December 31, 1997. To focus on nontransient, nonvalvular atrial fibrillation, patients with ICD-9 diagnoses of mitral stenosis, valvular repair or replacement, transient postoperative atrial fibrillation, or concurrent hyperthyroidism were excluded.[11]

Hemorrhage Outcomes Assessment

Hospitalization and billing databases were searched electronically through August 31, 1999, for primary and secondary discharge diagnoses of intracranial hemorrhage and primary discharge diagnoses of extracranial hemorrhage. Secondary discharge diagnoses of extracranial hemorrhage were not included, because initial validation studies revealed that few of these events represented true incident major hemorrhages.[12] Hemorrhages that occurred as a complication of a hospitalization for another problem were not included. Because Kaiser Permanente is an integrated healthcare system, hemorrhagic events that incurred medical bills were identified even if they occurred at non-Kaiser Permanente medical facilities.

A clinical outcomes committee reviewed the hospitalization medical records of all potential outcome events using a formal study protocol. Hemorrhagic events were categorized as occurring on or off warfarin based on chart review. The international normalized ratio (INR) at presentation before reversal of excessive anticoagulation was obtained from review of the medical record and the inpatient health plan laboratory database. Intracranial hemorrhages associated with major head trauma (e.g., neurosurgical procedure, high-velocity motor vehicle accident, and skull fracture) were excluded, because interest was primarily in the risk of spontaneous hemorrhages. Major extracranial hemorrhages were defined as fatal, requiring transfusion of two or more units of packed blood cells, or hemorrhage into a critical anatomic site (other than intracranial).

Patient Characteristics

Data on patient age and sex were obtained from electronic databases. Diagnoses of specific medical conditions were obtained through automated searches of electronic clinical inpatient and ambulatory databases for relevant ICD-9 codes during the 5 years before study enrollment. This search strategy was validated by reviewing a subset of medical records, as has been described previously.[13] Clinical risk factors in the Outpatient Bleeding Index (prior ischemic stroke, prior gastrointestinal hemorrhage, anemia (hemoglobin <13 g/dL in men and <12 g/dL in women), serum creatinine >1.5 mg/dL, coronary artery disease, and diabetes mellitus)[14] and other potential risk factors for hemorrhage (prior hematuria, prior other hemorrhage, hepatic cirrhosis, and mechanical fall complicating a previous hospitalization) were specifically searched for. Although nonsteroidal antiinflammatory drug (NSAID) and aspirin use increase hemorrhage risk,[6,15] their use was not captured comprehensively in the pharmacy database because of their availability without prescription. Attempts were made to ascertain whether aspirin was a potential confounder by reviewing the outpatient medical charts of 232 subjects not taking warfarin, as part of a previously described case-control study.[12] Information on aspirin use by outpatient warfarin users was lacking; instead, aspirin use is provided from chart review of 467 inpatients taking warfarin who were hospitalized for a thromboembolic or hemorrhagic event. The rationale was that, if aspirin use was not associated with older age, it would not confound the association between age and hemorrhage.

Warfarin Exposure

Warfarin exposure was ascertained from automated clinical, pharmacy, and laboratory databases using an algorithm previously validated through chart review.[12] Patients who were newly started on warfarin were defined as those without a previous filled prescription for warfarin in the 12 months before cohort enrollment and with no more than one outpatient INR measurement in the year before enrollment. The start date of warfarin was considered the date the prescription was dispensed. Length of warfarin exposure was calculated based on the number of days supplied per prescription. If there were gaps between prescriptions, patients were considered to be continually taking warfarin if the gaps were less than 60 days or if there were INR measurements at least every 42 days.

For patients taking warfarin, information on anticoagulation intensity and INR variability were obtained, because these factors have been associated with hemorrhage risk.[3] An adapted linear interpolation method was used to determine the person-time each patient spent at different INR intervals.[16] If a person was taking warfarin according to pharmacy records but the interval between INR measurements was greater than 8 weeks, the INR values for this extended period were not interpolated and instead were categorized as "not available"; 18% of the INR person-time of warfarin users met this criterion. INR variability was calculated by using the modified formula reported previously.[3]

Statistical Analyses

Patients who were taking warfarin were analyzed separately from those who were not. Because individual patients could have multiple periods on or off warfarin, a generalized estimating equation approach was used to account for correlation of observations obtained from the same patient.[17] Each patient was followed prospectively until the occurrence of a hemorrhagic event, the end of the study period, disenrollment from the health plan, or death. Patients suffering an intracranial hemorrhage remained at risk for an extracranial hemorrhage and vice versa. Event rates

for each age category were calculated by dividing the number of hemorrhagic events by the total person-years of follow-up in the specified warfarin category.

Multivariable Poisson regression models were used to assess whether age was associated with higher hemorrhage risk, grouping age into four ordinal categories (<60, 60–69, 70–79, and ≥80) and testing for an increasing trend in hemorrhage risk per incremental age category while adjusting for other potential risk factors for hemorrhage. Only those covariates with $P < .1$ according to univariate analysis were included in the final multivariable model for each outcome to avoid overfitting of the models. For patients receiving warfarin therapy, potential covariates also included INR intensity (dichotomized as ≥4.0 and <4.0) and INR variability (dichotomized as >90th and ≤90th percentile of variability).[3,18] Finally, separate multivariable models were developed to assess the effect of age on the risk of intracranial and major extracranial hemorrhages.

All analyses were performed using SAS software, version 9.1 (SAS Institute, Inc., Cary, NC). The institutional review boards of the collaborating institutions approved the study.

RESULTS

The cohort was followed for a median of 2.4 years (interquartile range 1.8–2.8 years). The mean age ± standard deviation of the cohort was 71 ± 15, and 28% of the total person-years were from patients aged 80 and older. Older patients were more likely to have risk factors for ischemic stroke, such as hypertension and coronary disease, and risk factors for hemorrhage, such as prior gastrointestinal hemorrhage, anemia, and renal insufficiency (Table 1). INRs of 4.0 and higher and high INR variability did not vary significantly by age ($P = .90$ and .08, respectively, Table 1).

Hemorrhagic Events

A total of 170 major hemorrhages were identified during 15,300 person-years of follow-up in patients taking warfarin, and 162 hemorrhages were identified during 15,530 person-years in patients not taking warfarin. There were 72 intracranial and 98 major extracranial hemorrhages that occurred in patients taking warfarin (annualized rate 0.47%, 95% confidence interval (CI) = 0.37–0.59 and 0.64%, 95% CI = 0.53–0.78, respectively) and 46 intracranial and 116 major extracranial hemorrhages that occurred in patients not taking warfarin (annualized rate 0.29%, 95% CI = 0.22–0.39 and 0.75%, 95% CI = 0.62–0.90, respectively). Eighty-nine percent of the 214 major extracranial hemorrhages involved the gastrointestinal tract.

Twenty-nine percent of the cohort, with a mean age of 70 ± 11, was considered to be newly started on warfarin. Only three intracranial and four major extracranial hemorrhagic events were observed during the first month after starting warfarin. The annualized rate of intracranial hemorrhage in the first month of taking warfarin was 0.92%, compared with 0.46% after the first month (crude relative rate = 2.0, 95% CI = 0.6–6.7). The rate of major extracranial hemorrhage was 1.2% in the first month of taking warfarin, compared with 0.61% afterward (crude relative rate = 2.0, 95% CI = 0.7–5.8). There were too few events to assess whether hemorrhage rates differed by age in the early phase of warfarin therapy.

Age and Risk for Hemorrhage

Upon univariate analysis, the rate of all major hemorrhages rose with older age, in patients taking and not taking warfarin. In patients who were taking warfarin, the average

Table 1. Clinical Characteristics of 13,559 Patients with Atrial Fibrillation, Stratified by Age and Warfarin Status

	On Warfarin					Off Warfarin				
	Age (number of person-years)					Age (number of person-years)				
Patient Characteristics	Mean Age ± SD	<60 (1,453)	60–69 (3,269)	70–79 (6,767)	≥80 (3,818)	Mean Age ± SD	<60 (2,493)	60–69 (2,946)	70–79 (5,253)	≥80 (4,934)
		%					%			
Women	74 ± 11	22.9	34.9	44.9	53.7	74 ± 14	26.8	39.2	45.7	54.8
Prior ischemic stroke	74 ± 11	7.7	9.2	12.5	15.6	77 ± 9	1.4	3.8	6.5	8.2
Diagnosed hypertension	72 ± 11	39.6	52.0	55.7	54.7	73 ± 12	28.2	51.1	53.6	52.3
Known coronary disease	73 ± 10	15.9	28.7	33.5	33.8	75 ± 10	9.3	25.7	30.9	31.8
Diabetes mellitus	71 ± 10	15.7	23.1	19.4	13.9	72 ± 11	9.4	17.1	18.6	10.8
Prior fall during hospitalization	76 ± 10	1.1	2.1	2.5	5.0	79 ± 12	1.0	2.4	4.0	8.7
Prior gastrointestinal bleed	73 ± 10	2.1	1.9	3.4	3.1	76 ± 12	2.0	3.4	5.8	8.3
Prior hematuria	73 ± 10	0.9	2.2	1.8	1.8	75 ± 11	0.7	1.5	1.9	2.1
Prior other bleed	73 ± 11	0.6	0.7	1.0	1.3	72 ± 11	0.5	1.3	0.9	1.0
Hepatic cirrhosis	69 ± 9	0.6	0.8	0.5	0.3	71 ± 12	0.9	1.2	1.4	0.7
Anemia*	74 ± 10	6.8	8.6	12.6	18.4	76 ± 12	6.8	12.7	17.6	25.3
Serum creatinine >1.5 mg/dL	74 ± 10	3.8	7.6	10.1	13.6	76 ± 11	4.6	7.4	8.7	12.5
INR ≥4.0	72 ± 2	1.7	1.7	1.7	1.8	—	—	—	—	—
INR variability >90th percentile	71 ± 3	3.8	3.8	3.4	3.7	—	—	—	—	—

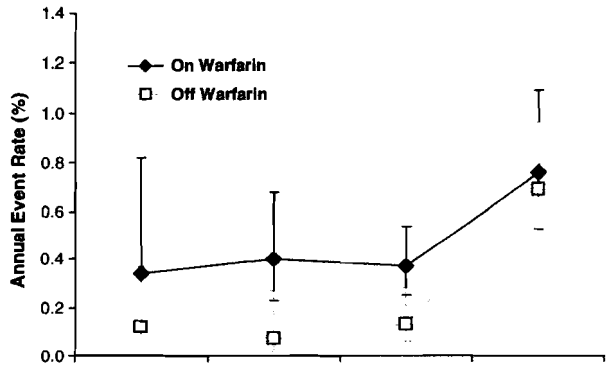
* Anemia defined as hemoglobin <13 g/dL in men and <12 g/dL in women.
SD = standard deviation; INR = international normalized ratio.

relative increase in the rate of hemorrhage from one age category to the next older age category was 1.3 (95% CI = 1.1–1.6). In patients who were not taking warfarin, the increase was 1.7 (95% CI = 1.5–2.1) per age category. Adjusting for potential risk factors for hemorrhage somewhat attenuated these results; the relative increase in hemorrhage rate in warfarin users was 1.2 (95% CI = 1.0–1.4) per age category once prior ischemic stroke, history of gastrointestinal hemorrhage, anemia, renal insufficiency, INR at presentation, and INR variability were adjusted for. For patients not taking warfarin, the relative increase in hemorrhage rate was 1.5 (95% CI = 1.3–1.8) per age category (adjusting for prior stroke, gastrointestinal hemorrhage, anemia, renal insufficiency, and diabetes mellitus).

Upon univariate analyses, the unadjusted rate of intracranial hemorrhage remained relatively flat from age 60 to 80 and then increased sharply at 80. This increase was observed in patients taking warfarin (adjusted rate ratio (RR) = 1.8, 95% CI = 1.1–3.1, comparing patients aged ≥80 with those <80) and in patients not taking warfarin (adjusted RR = 4.7, 95% CI = 2.4–9.2, comparing patients aged ≥80 with those <80) (Figure 1).

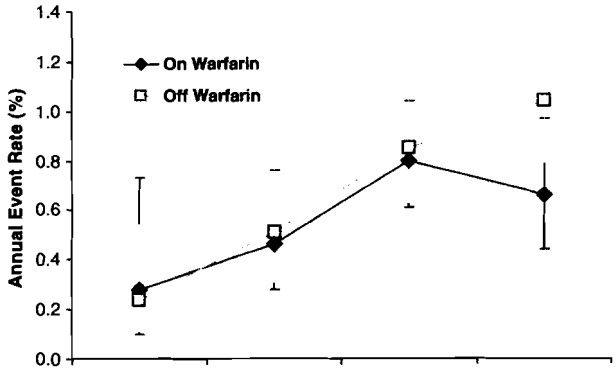
The relationship between extracranial hemorrhage and age was less consistent. In patients taking warfarin, rates of extracranial hemorrhage were significantly higher at age 70 to 79 than in those younger than 60, with an adjusted RR of 2.0 (95% CI = 1.1–8.2). At age 80 and older, the risk leveled off, with an adjusted RR of 2.1 (95% CI = 0.7–6.4) (Figure 2). In contrast, the increase in extracranial hemorrhage rates in patients who were not taking warfarin rose in a linear fashion (Figure 2), with each older age category associated with an increase in hemorrhage rate of 1.3 (95% CI = 1.1–1.7) after multivariable adjustment.

Aspirin was unlikely to be a significant confounder of the relationship between age and hemorrhage. In a sample of 232 outpatients not taking warfarin described in the Methods section, the proportion of patients taking aspirin did not vary significantly across age categories (39.4% documented aspirin use in patients aged <60, 57.1% in patients aged 60–69, 44.0% in patients aged 70–79, and 53.1% in patients aged ≥80, chi-square $P = .33$). Of 467 inpatients taking warfarin and hospitalized with an outcome event, aspirin use was uncommon, with only 8.4% of patients taking both medications, and no differences in aspirin use by age were observed (chi-square $P = .78$).



Age	<60	60–69	70–79	≥80
On warfarin				
Events, n	5	13	25	29
Rate (95% CI)	0.34 (0.14–0.82)	0.40 (0.23–0.68)	0.37 (0.25–0.54)	0.76 (0.53–1.10)
Off warfarin				
Events, n	3	2	7	34
Rate (95% CI)	0.12 (0.04–0.37)	0.07 (0.02–0.27)	0.13 (0.6–0.28)	0.69 (0.49–0.96)

Figure 1. Unadjusted age-specific rates of intracranial hemorrhage of 13,559 patients with nonvalvular atrial fibrillation taking and not taking warfarin. CI = confidence interval.



Age	<60	60–69	70–79	≥80
On warfarin				
Events, n	4	15	54	25
Rate (95% CI)	0.28 (0.10–0.73)	0.46 (0.28–0.76)	0.80 (0.61–1.0)	0.66 (0.44–0.97)
Off warfarin				
Events, n	6	15	44	51
Rate (95% CI)	0.24 (0.11–0.54)	0.51 (0.31–0.85)	0.85 (0.63–1.1)	1.04 (0.79–1.4)

Figure 2. Unadjusted age-specific rates of major extracranial hemorrhage of 13,559 patients with nonvalvular atrial fibrillation taking and not taking warfarin. CI = confidence interval.

DISCUSSION

The risk of major hemorrhage rose with older age whether subjects were taking warfarin or not in this large cohort of adults with nonvalvular atrial fibrillation. A notable finding was that the rates of intracranial hemorrhage—the most devastating form of hemorrhage—rose strikingly at the age of 80 and older.

Reported rates of warfarin-associated hemorrhage in patients with atrial fibrillation vary considerably, probably representing at least some differences in the baseline bleeding risk of the source populations. Rates ranged from 0.76% per year[19] to approximately 10% per year in a study of patients aged 75 and older.[20] Prior studies have also provided conflicting data on whether older age is a risk factor for warfarin-associated hemorrhage.[3,6–8,21,22] Advantages of the current study include a considerably higher number of outcome events, more person-years of observation accumulated in the oldest patients, and separate analyses between intracranial and extracranial hemorrhages.

The relationship between age and hemorrhage risk was found to be somewhat attenuated after adjustment for other medical conditions, suggesting that clinical factors related to aging mediate some of the higher risk. That an independent relationship still remained after multivariable adjustment indicates that other age-related features contribute to hemorrhage risk and were not identified in this study. It is probable that age-related risk factors such as cerebral amyloid angiopathy and leukoaraiosis explain some of the association between age and risk of intracranial hemorrhage.[23,24] It is also likely that the greater prevalence of

gastrointestinal pathology in older adults contributes to the higher risk of extracranial hemorrhages.[25,26] Although there is concern that older patients are more difficult to manage on warfarin, there were no significant differences in INR control across age categories, making it unlikely that INR instability mediated the observed age effects.

The hemorrhage rates observed in patients who were taking warfarin were similar to the hemorrhage rates in patients who were not taking warfarin. Because prior studies have demonstrated that warfarin clearly increases hemorrhage risk, this finding suggests that clinicians may selectively prescribe warfarin for patients who have a lower intrinsic risk for hemorrhage. This probable selection effect was more apparent in extracranial hemorrhages, which clinicians may more easily anticipate than intracranial hemorrhages. Intracranial hemorrhages are difficult to predict, and several of the known risk factors—namely, prior stroke and hypertension—are also indications for prescribing warfarin for atrial fibrillation.

Hemorrhage rates in this observational study of actual clinical care were similar to the low rates reported in randomized trial settings, in which there was a large net benefit observed with warfarin.[1,27] The majority of patients in this cohort were prevalent users of warfarin, which may have contributed to the low rates of hemorrhage. Patients newly started on warfarin may be more likely to suffer hemorrhagic complications. Indeed, higher rates of hemorrhage were observed in patients during their first month on warfarin therapy, but the small number of events makes such an estimate imprecise. It would be valuable for future assessments of warfarin safety to focus specifically on the initial phase of drug therapy. It is also possible that low hemorrhage rates were observed, because dedicated anticoagulation clinics, which have been shown to improve anticoagulation control, cared for most patients.[28] Anticoagulation in the appropriate INR range is integral to maximizing stroke prevention while minimizing hemorrhage risk.[10,29] This study demonstrates that warfarin can be used with reasonable safety in clinical settings, even in elderly patients, if monitored carefully.

This study had several limitations. As an observational study of actual clinical practice, it was subject to nonstandardized data collection, resulting in periods of missing warfarin exposure and unavailable INR data. Because warfarin treatment was not randomly assigned, confounding by contraindication may occur in that physicians are less likely to anticoagulate patients at higher risk for hemorrhage. Such residual confounding is less likely to affect the estimates of intracranial hemorrhage, because there are few validated clinical risk factors for intracranial hemorrhage. Identifying only hospitalizations with a primary diagnosis of extracranial hemorrhage may have missed some extracranial events, although previous validation studies of the cohort indicate that secondary discharge diagnoses of extracranial hemorrhages rarely represented valid incident events and that including such events in the search algorithm would have added negligibly to the total number of major extracranial bleeds. Finally, comprehensive data on nonprescription aspirin and NSAID use were not available, because nonprescription medications were not routinely recorded in the health plan's pharmacy database. If older patients were more likely to take NSAIDs, then use of NSAIDs may confound some of the age effect. There was no significant association between aspirin use and age in samples of outpatients and inpatients, making it less likely that aspirin use, at least, confounded the relationship between age and hemorrhage.

This study provides clear evidence that major hemorrhage rates rise with older age in patients with atrial fibrillation, whether taking warfarin or not. In particular, the rates of intracranial hemorrhage rose sharply at the age of 80 and older. Because patients with atrial fibrillation are generally older and have multiple comorbid conditions, care must be taken to minimize modifiable risk factors for hemorrhage when using warfarin, particularly avoiding excessively high anticoagulant intensities. These results also support vigilance when initially starting warfarin therapy, although the absolute rates of major hemorrhage in this study were reassuringly low even in the oldest patients and comparable with rates reported in randomized trials. These findings indicate that well-managed warfarin therapy can be used safely in clinical practice to achieve substantial benefit in reducing the risk of atrial fibrillation–associated ischemic stroke.

ACKNOWLEDGMENTS

Financial Disclosure: This work was supported by Public Health Services research Grant AG15478 from the National Institute on Aging, the Eliot B. and Edith C. Shoolman Fund of Massachusetts General Hospital, and a Hartford Geriatrics Health Outcomes Research Scholars Award from the American Geriatrics Society Foundation for Health in Aging. Elaine Hylek received research grant support from Bristol-Myers Squibb (makers of brand name Coumadin).

Author Contributions: Margaret Fang, Alan Go, Elaine Hylek, and Daniel Singer contributed to study concept and design, data analysis and interpretation, and preparation of manuscript. Yuchiao Chang contributed to data analysis and interpretation, and preparation of manuscript. Lori Henault and Nancy Jensvold contributed to study concept and design and acquisition of subjects and data.

Sponsor's Role: The sponsor had no role in the design, methods, subject recruitment, data collection, analysis, or preparation of this manuscript.

REFERENCES

1. 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.
2. 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.
3. Fihn SD, Callahan CM, Martin DC et al. The risk for and severity of bleeding complications in elderly patients treated with warfarin. Ann Intern Med 1996;124:970–979.
4. Palareti G, Leali N, Coccheri S et al. Bleeding complications of oral anticoagulant treatment: An inception-cohort, prospective collaborative study (ISCOAT). Lancet 1996;348:423–428.
5. Stroke Prevention in Atrial Fibrillation Investigators. Stroke Prevention in Atrial Fibrillation Study: Final results. Circulation 1991;84:527–539.
6. Sam C, Massaro JM, D'Agostino RB Sr. et al. Warfarin and aspirin use and the predictors of major bleeding complications in atrial fibrillation (the Framingham Heart Study). Am J Cardiol 2004;94:947–951.
7. McMahan D, Smith D, Carey M et al. Risk of major hemorrhage for outpatients treated with warfarin. J Gen Intern Med 1998;13:311–316.
8. Torn M, Bollen WLEM, van der Meer FJM et al. Risks of oral anticoagulant therapy with increasing age. Arch Intern Med 2005;165:1527–1532.

9. Fang MC, Stafford RS, Ruskin JN et al. National trends in antiarrhythmic and antithrombotic medication use in atrial fibrillation. Arch Intern Med 2004; 164:55–60.
10. Fang MC, Chang Y, Hylek EM et al. Advanced age, anticoagulation intensity, and risk for intracranial hemorrhage among patients taking warfarin for atrial fibrillation. Ann Intern Med 2004;141:745–752.
11. 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.
12. Go AS, Hylek EM, Chang Y et al. Anticoagulation therapy for stroke prevention in atrial fibrillation: How well do randomized trials translate into clinical practice? JAMA 2003;290:2685–2692.
13. 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.
14. Beyth R, Quinn L, Landefeld C. Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin. Am J Med 1998;105:91–99.
15. Shorr R, Ray W, Daugherty J et al. Concurrent use of nonsteroidal anti-inflammatory drugs and oral anticoagulants places elderly persons at high risk of hemorrhagic peptic ulcer disease. Arch Intern Med 1993;153:1665–1670.
16. Rosendaal FR, Cannegieter SC, van der Meer FJ et al. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993;69:236–239.
17. Diggle P, Liang K, Zieger S. Analysis of Longitudinal Data. Oxford: Clarendon Press, 1994.
18. 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.
19. Yousef ZR, Tandy SC, Tudor V et al. Warfarin for non-rheumatic atrial fibrillation: Five year experience in a district general hospital. Heart 2004;90: 1259–1262.
20. Johnson CE, Lim WK, Workman BS. People aged over 75 in atrial fibrillation on warfarin: The rate of major hemorrhage and stroke in more than 500 patient-years of follow-up. J Am Geriatr Soc 2005;53:655–659.
21. Landefeld C, Rosenblatt M, Goldman L. Bleeding in outpatients treated with warfarin: Relation to the prothrombin time and important remediable lesions. Am J Med 1989;87:153–159.
22. van der Meer F, Rosendaal F, Vandenbroucke J et al. Bleeding complications in oral anticoagulant therapy: An analysis of risk factors. Arch Intern Med 1993;153:1557–1562.
23. Rosand J, Hylek EM, O'Donnell HC et al. Warfarin-associated hemorrhage and cerebral amyloid angiopathy. Neurology 2000;55:947–957.
24. Gorter J. Major bleeding during anticoagulation after cerebral ischemia: Patterns and risk factors. Neurology 1999;53:1319–1327.
25. Hernandez-Diaz S, Rodriguez LA. Incidence of serious upper gastrointestinal bleeding/perforation in the general population: Review of epidemiologic studies. J Clin Epidemiol 2002;55:157–163.
26. Kaplan R, Heckbert S, Koepsell T et al. Risk factors for hospitalized gastrointestinal bleeding among older persons. Cardiovascular Health Study Investigators. J Am Geriatr Soc 2001;49:126–133.
27. Olsson S. Executive Steering Committee on behalf of the SPORTIF III Investigators. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): Randomised controlled trial. Lancet 2003;362:1691–1698.
28. Chiquette E, Amato M, Bussey H. Comparison of an anticoagulation clinic with usual medical care. Arch Intern Med 1998;158:1641–1647.
29. Hylek E, Skates S, Sheehan M et al. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. N Engl J Med 1996;335:540–546.

The Impact of a Suspicious Prostate Biopsy on Patients' Psychological, Socio-behavioral, and Medical Care Outcomes

Floyd J. Fowler, Jr, PhD,[2] Michael J. Barry, MD,[1] Beth Walker-Corkery, MPH,[1] Jean-Francois Caubet, MD,[3] David W. Bates, MD, MSc,[3] Jeong Min Lee, MD,[4] Alison Hauser, MA,[2] Mary McNaughton-Collins, MD, MPH[1]

[1]The General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; [2]The Center for Survey Research, University of Massachusetts, Boston, MA, USA; [3]The General Medicine Division (DWB) and Division of Urology, Brigham and Women's Hospital, Boston, MA, USA; [4]The General Medicine Division, Boston Medical Center, Boston, MA, USA.

OBJECTIVE: To evaluate the psychological, socio-behavioral, and medical implications of apparently false-positive prostate cancer screening results.

METHODS: One hundred and twenty-one men with a benign prostate biopsy performed in response to a suspicious screening test (biopsy group) and 164 men with a normal prostate-specific antigen (PSA) test result (normal PSA group) responded to a questionnaire 6 weeks, 6 and 12 months after their biopsy or PSA test.

RESULTS: The mean (± SD) age of respondents was 61 ± 9 years (range, 41 to 88 years). One year later, 26% (32/121) of men in the biopsy group reported having worried "a lot" or "some of the time" that they may develop prostate cancer, compared with 6% (10/164) in the normal PSA group ($P<.001$). Forty-six percent of the biopsy group reported thinking their wife or significant other was concerned about prostate cancer, versus 14% in the normal PSA group ($P<.001$). Medical record review showed that biopsied men were more likely than those in the normal PSA group to have had at least 1 follow-up PSA test over the year (73% vs 42%, $P<.001$), more likely to have had another biopsy (15% vs 1%, $P<.001$), and more likely to have visited a urologist (71% vs 13%, $P<.001$).

CONCLUSION: One year later, men who underwent prostate biopsy more often reported worrying about prostate cancer. In addition, there were related psychological, socio-behavioral, and medical care implications. These hidden tolls associated with screening should be considered in the discussion about the benefits and risks of prostate cancer screening.

KEY WORDS: prostate biopsy; prostate cancer; screening.
DOI: 10.1111/j.1525-1497.2006.00464.x
J GEN INTERN MED 2006; 21:715–721.

Despite controversy over the effectiveness of screening for prostate cancer,[1] the practice is widespread in the United States.[2] The usefulness of the prostate-specific antigen (PSA) test as a screening tool for prostate cancer has recently been called into question by an early proponent of the test.[3] Stamey and colleagues examined 1,317 consecutive radical prostatectomies between 1983 and 2003, and showed that serum PSA was related to prostate cancer 20 years ago, but in recent years was only related to benign prostatic hyperplasia. Given the uncertainty about the potential benefits of prostate cancer screening, it is imperative to understand the potential for risks (including psychological distress) associated with screening. Previous work[4] demonstrated that many men with an apparently false-positive prostate cancer screening test, meaning a suspicious screening test followed by a benign biopsy (cautiously called "false positive" because some of these men may have had a false-negative biopsy result), suffered negative psychological effects about 6 weeks later. In this study, we evaluated the longer-term psychological and other effects of an apparently false-positive prostate cancer-screening test.

METHODS

Study Sample

A prospective cohort of men recently screened for prostate cancer was assembled between August 2001 and September 2002 from the primary care practices of Massachusetts General Hospital, Brigham and Women's Hospital, and Boston Medical Center to study the psychological, socio-behavioral, and medical care of patients who were not found to have cancer after a prostate biopsy. The Institutional Review Board of each institution approved the study.

Patients were identified through weekly review of pathology reports and PSA test results. The inclusion criteria were as follows: men, aged 40 and older, living in the United States, with a primary care physician at 1 of the participating institutions, and a benign prostate biopsy performed because of a suspicious screening test (biopsy group). A normal PSA group consisted of men who had a normal PSA test (defined as <2.5 ng/mL) during the same period. Exclusion criteria were as follows: a diagnosis of prostate cancer, inability to understand English, and permission from the physician not granted; previous prostate biopsy was an additional exclusion criterion in the normal PSA group.

Study Design

After telephone contact and consent were obtained, prospective participants were mailed a brief (< 10 minutes), self-administered, pretested questionnaire about 6 weeks after their benign biopsy (biopsy group) or normal screening PSA test (normal PSA group). Patients who returned the 6-week ques-

Dr. Barry is a consultant for the Foundation for Informed Medical Decision-Making (a 501 (c) 3 not for profit foundation).

Address correspondence and requests for reprints to Dr. McNaughton-Collins: General Medicine Division, Massachusetts General Hospital, 50 Staniford Street, 9th Floor, Boston, MA 02114 (e-mail: mmcnaughton-collins@partners.org).

Manuscript received July 20, 2005
Initial editorial decision August 21, 2005
Final acceptance February 23, 2006

tionnaire, which was the baseline data collection, were surveyed again at 6 months, and those who responded at 6 months were sent questionnaires at 12 months. Because of an Institutional Review Board stipulation that nonresponse after reminders be considered a refusal, patients who did not return any questionnaire were not sent further surveys. A telephone reminder and mailing to nonrespondents was employed at each time point. Please see previous publication for additional details.[4]

Response. After the initial phone contact and consent, 480 patients were deemed eligible and sent a 6-week questionnaire; 83% (239/287) in the biopsy group returned a survey and 87% (168/193) in the normal PSA group responded. One man in the biopsy group and 7 in the normal PSA group were deemed ineligible based on responses to the first questionnaire, and were excluded from analysis and further follow-up. At 6 months, 83% (139/167) of the biopsy group and 84% (194/232) of the normal PSA group returned a questionnaire. One man in each group reported being diagnosed with prostate cancer on the 6-month questionnaire, and was excluded from further analysis and follow-up. In the biopsy and normal PSA groups, 88% (121/138) and 85% (164/193), respectively, returned a 12-month questionnaire. Overall, 63% (121/193) of the original biopsy group and 57% (164/287) of the original normal PSA group returned a 12-month survey.

Responders and Nonresponders. Based on characteristics reported in Table 1 at 6 weeks, there were some differences between responders and nonresponders at 6 and 12 months. In the biopsy group, we found no difference between 6-month responders and nonresponders in these variables, and 1 difference between 12-month responders and nonresponders. Twelve-month responders in the biopsy group reported having had more PSA tests at 6 weeks than nonresponders; 8% versus 16% had 1 PSA test, 25% versus 40% had 2 to 4, 41% versus 30% had 5 to 10 and 27% versus 14% had 11 or more, $P=.04$. In the normal PSA group, at 6 months the responders were more likely to be white (90% vs 72%, $P=.007$), college educated (90% vs 74%, $P=.02$), and have had more PSA tests (14% vs 31% had 1 PSA test, 37% vs 49% had 2 to 4, 36% vs 17% had 5 to 10 and 13% vs 3% had 11 or more, $P=.007$). Twelve-month responders in the normal PSA group were older (mean age 61 vs 58, $P=.03$), more likely to be white (91% vs 78%, $P=.02$), married (80% vs 61%, $P=.004$), and college educated (91% vs 79% $P=.02$). The 12-month responders also had more PSA tests than the nonresponders (11% vs 31% had 1 PSA test,

Table 1. Baseline Characteristics of the Participants*

Characteristic	N (%)		P Value
	Biopsy Group N=167	Normal PSA Group N=232	
Mean age	61.1	59.8	P=.15
Age (y)			
< 50	16 (10)	29 (13)	P=.40
50 to 59	58 (35)	89 (38)	
60 to 69	63 (38)	77 (33)	
70 to 79	29 (17)	32 (14)	
80+	1 (1)	5 (2)	
White, not Hispanic	149 (90)	201 (87)	P=.34
Married	131 (79)	171 (74	P=.34
Education			
High school or less	31 (19)	29 (13)	P=.24
Some college or degree	69 (42)	103 (44)	
Advanced degree	66 (40)	100 (43)	
Family history of prostate cancer	23 (14)	31 (13)	P=.88
Benign prostatic hyperplasia	57 (34)	30 (13)	P<.0001
Prostatitis	38 (23)	21 (9)	P<.0001
Approximately how many prostate specific antigen tests have you had?			
1	16 (10)	35 (17)	P=.002
2 to 4	46 (29)	82 (39)	
5 to 10	61 (38)	68 (33)	
11+	38 (24)	24 (11)	
Visits to urologist over past 12 months			
0	3 (2)	179 (79)	P<.0001
1	48 (29)	28 (12)	
2+	114 (69)	21 (9)	
Overall rating of current physical health			
Excellent	49 (30)	56 (24)	P=.14
Very good	60 (36)	101 (44)	
Good	47 (28)	52 (22)	
Fair-poor	10 (6)	23 (10)	
Overall rating of current mental health			
Excellent	72 (43)	89 (38)	P=.57
Very good	56 (34)	88 (38)	
Good	33 (20)	42 (18)	
Fair-poor	6 (4)	13 (6)	

*Number for individual items vary slightly because of nonresponse.
PSA, prostate-specific antigen.

40% vs 37% had 2 to 4, 35% vs 25% had 5 to 10, and 13% vs 7% had 11 or more, $P=.007$).

To assess whether response was associated with any of the outcome variables we examined responders versus nonresponders at 6 and 12 months separately on 3 key psychological outcome variables (prostate cancer risk perception, worry about prostate cancer, and thinking about prostate cancer). There were no significant differences between 6-month responders and nonresponders on these 3 variables at 6 weeks. Additionally, we found no significant differences between 12-month responders and nonresponders on the 3 psychological variables at 6 weeks, or at 6 months.

Measurements

Questionnaire Development. A literature review of assessments of the psychological effects of suspicious prostate cancer screening results identified few relevant studies[5–8]; therefore, assessments quantifying the effects of suspicious mammograms were reviewed.[9–12] We conducted 3 focus groups of men who had a benign biopsy result in response to a suspicious screening test and 1 focus group of men who had a normal screening PSA test. A preliminary questionnaire was developed, and refined using in person pretesting ($n=5$). See previous publication for additional details.[4]

Demographics, Family History, Medical History, and Health Status. Information on age, race, marital status, education, medical history, health status, and family history of prostate cancer was collected.

Psychological Impact. Men were asked how much they had thought about prostate cancer, had worried about developing prostate cancer, what they thought was their chance of getting prostate cancer someday, and how reassured they felt as a result of their most recent PSA test.

Socio-behavioral Impact. Men were asked how much they talked with their wife or significant other about prostate cancer, and how much they thought their wife or significant other was concerned about them developing prostate cancer. Men were also asked about knowledge seeking related to prostate cancer—reading books, magazine or newspaper articles, or searching the Internet for information, and how well informed they felt about prostate cancer.

Medical Care. We performed a medical record review to document how many PSA tests and prostate biopsies the men in our study had over the 1 year of follow-up. Men were also asked how many times they had visited or called their primary care provider and urologist over the year.

Statistical Analyses

Differences in proportions between the 2 groups at all time points were compared with the Fisher's exact test for 2×2 tables, and the χ^2-test for larger tables. The Pearson exact χ^2-method was used wherever small cell counts were a concern. Tests for trend were performed using the Cochran-Mantel-Haenszel test to compare trends over time within groups (biopsy vs normal PSA), and the Breslow-day test for homogeneity of odds ratios to examine whether the difference in proportions between the groups changed over the 3 time points. Ordinal logistic regression models were utilized to assess group effect at 6 and 12 months adjusting for potential confounding factors (history of BPH or prostatitis, number of previous PSA tests at 6 weeks, and number of previous visits to an urologist reported at 6 weeks).

RESULTS

Study Sample Characteristics

There were no significant differences between the groups in age, race/ethnicity, marital status, education, health status, or family history of prostate cancer (Table 1). Compared with the normal PSA group, more of the biopsy group had histories of benign prostatic hyperplasia and prostatitis, and they reported more previous PSA tests, and visits to urologists.

Psychological Impact

Compared with the normal PSA group, the biopsy group more often reported thinking and worrying about prostate cancer at every time point and more often reported thinking their chance of getting prostate cancer was greater than average. Figure 1 shows that 2 of the 3 measures drop significantly between 6 weeks following the biopsy and the 6-month follow-up, but hold steady or rise slightly between 6 and 12 months. The perception of elevated risk of cancer rose steadily throughout the year after the biopsy. At every point, these perceptions were significantly higher among the biopsy group than among the normal PSA group.

Socio-behavioral Impact

Data for the 12-month time point are presented in Table 2. Results at 6 months were essentially the same except for the proportion of men in the normal PSA group who read articles about prostate cancer, which went from 45% at 6 months to 56% at 12 months. The difference between the groups on this item was significant at 6 months ($P=.01$). The biopsy group reported more often having talked with their wife or significant other about prostate cancer and more often reported thinking their wife or significant other was concerned about them developing prostate cancer. More men in the biopsy group compared with the normal PSA group reported seeking information about prostate cancer on the Internet, and this group reported feeling more informed about prostate cancer.

Medical Care

Medical record review revealed that over 1-year of follow-up the biopsied men had more follow-up PSA tests and prostate biopsies than the normal PSA group, and the survey revealed they had more office visits and calls to urologists (Table 3). Thirty-three percent (40/121) of men in the biopsy group had 2 or more follow-up PSA tests; of these, 25 men had 2, 13 men had 3, and 2 men had 4 additional PSA tests within the year.

We stratified by history of BPH and prostatitis to see if the additional follow-up may have resulted from these other urological problems. Among men in the biopsy group, there were no differences with respect to history of BPH in the number of additional PSA tests, prostate biopsies, or visits to the urologist; however, more men in the biopsy group with a history of prostatitis had visited the urologist more than 2 times (56% vs 31%, $P=.03$). There were no differences according to prostatitis status within the biopsy group in number of additional PSA tests or prostate biopsies.

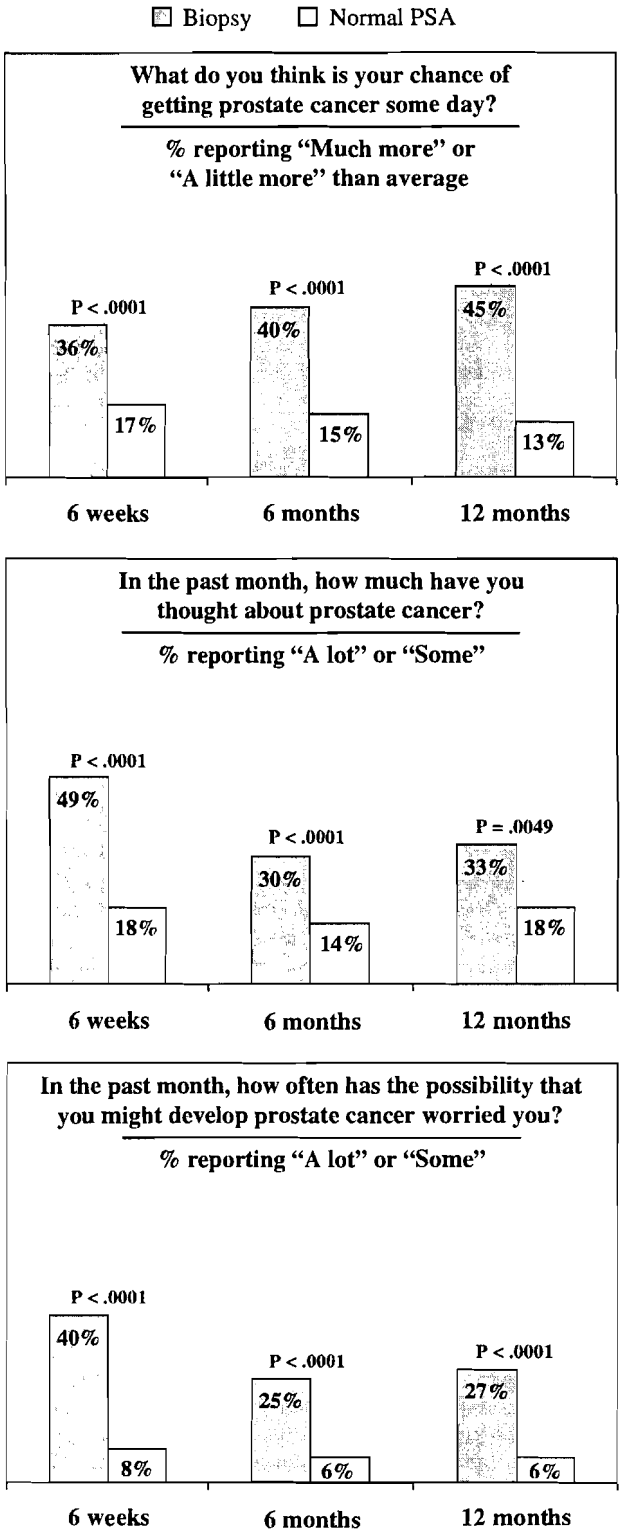


FIGURE 1. Psychological impact at 6 weeks, 6 months, and 12 months by group.

The vast majority of men (>90%) in both groups reported (at 1-year follow-up) planning to have regular PSA tests in the future, and that they would have a biopsy if their doctor recommended it.

DISCUSSION

We found that a considerable proportion of men with benign prostate biopsies after suspicious screening tests reported a negative psychological impact at 6 and 12 months, which extends our previous work showing a negative psychological impact at 6 weeks. Men with benign prostate biopsies reported substantial thinking and worrying about prostate cancer, even after the benign biopsy. In addition there appeared to be associated psychological, socio-behavioral, and medical utilization implications, demonstrating that the impact was an important one. Men in the biopsy group were more likely than men in the normal PSA group to report talking with their wife or significant other about prostate cancer, thinking their wife or significant other was worried about prostate cancer, searching on the Internet about prostate cancer, visiting the urologist, and undergoing additional PSA tests and prostate biopsies.

Investigators from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), have described the medical and nonmedical costs associated with false-positive prostate cancer screens.[13] The PLCO trial found that men with false-positive prostate cancer screening results were nearly twice as likely not to return for further prostate cancer screening, compared with men with normal prostate cancers screening results.[14] Our survey contrasted with this finding; the vast majority of men in our study reported being committed to having subsequent screening tests. In fact, our medical record review showed fully 71% of men in the biopsy group had at least 1 more screening PSA test in the subsequent year. The differences in the findings between our study and the PLCO substudy may be related to study design; the PLCO study was a population based screening trial, whereas our study was a nonrandomized, comparative study that recruited men from primary care physician offices. Although our study population and methodology differed, our findings are consistent with those of Schwartz et al.,[15] who found that the public is enthusiastic about cancer screening, and the commitment is not dampened by false-positive test results. The finding that about 1-third of men in our biopsy group had between 2 and 4 additional PSA tests and 15% had another prostate biopsy within the year suggests that physicians and patients do not view the initial benign biopsy result as entirely reassuring, and physicians continue closely monitoring these men. It is noteworthy that the biopsied men were regular patients of urologists over the year of follow-up; 71% had seen an urologist at least once and 38% had 2 or more urology visits. This ongoing surveillance and the possibility of a false negative biopsy may help propagate the anxiety we documented among these men.

Concern about false-negative biopsy results is fairly unique to prostate cancer screening. In addition to lower specificity (therefore more false positives) than breast and colorectal cancer screening,[1,16,17] the follow-up test, transrectal ultrasound-guided biopsy, involves random sampling of the prostate gland (in addition to targeted biopsies of suspicious areas), causing mounting concerns about false-negative biopsies.[18] Whereas a benign biopsy in response to an abnormal mammogram is fairly reassuring (because the abnormal area of the breast has been visualized and biopsied), the elevated screening PSA, a blood test, simply represents a general indictment of the prostate gland, and, because of the poor negative predictive value of the random biopsy, at least 10% of

Table 2. Sociobehavioral Impact at 12 Months*

Item	Number (%)		P Value
	Biopsy Group N=121	Normal PSA Group N=164	
Currently married/significant other	N=101 (84)	N=134 (83)	P=.9
In the past month, how much have you talked with your wife or significant other about prostate cancer? (for those married/have significant other*)			
A lot	0 (0)	0 (0)	P=.001
Some	11 (11)	4 (3)	
Only a little	35 (35)	27 (21)	
Not at all	55 (54)	100 (76)	
How much do you think your wife or significant other is concerned about your developing prostate cancer? (for those married/have significant other*)			
A lot	9 (9)	4 (3)	P<.0001
Some	37 (37)	15 (11)	
Only a little	35 (35)	52 (39)	
Not at all	19 (19)	62 (47)	
In the past 6 months, have you read any books about prostate cancer?			
Yes	8 (7)	7 (4)	P=.43
No	112 (93)	154 (96)	
In the past 6 months, have you read any articles in magazines or the newspaper about prostate cancer?			
Yes	71 (59)	90 (56)	P=.63
No	49 (41)	71 (44)	
In the past 6 months, have you gone on the Internet for information about prostate cancer?			
Yes	16 (13)	6 (4)	P=.006
No	104 (87)	155 (96)	
How well informed do you feel about prostate cancer?			
Very well	19 (16)	16 (10)	P=.01
Fairly well	90 (76)	111 (69)	
Not well at all	10 (8)	33 (21)	

*Number for individual items vary slightly because of nonresponse.
PSA, prostate-specific antigen.

Table 3. Medical Care at 12 Months*

Item	Number (%)		P value
	Biopsy Group N=121	Normal PSA Group N=164	
Number of times visited primary care physician over 12 months			
0 times	19 (16)	19 (12)	P=.6
1 time	35 (29)	49 (30)	
2 or more times	67 (55)	96 (58)	
Number of times called primary care physician over 12 months			
0 times	81 (67)	93 (57)	P=.1
1 time	21 (17)	30 (18)	
2 or more times	19 (16)	41 (25)	
Number of times visited urologist over 12 months			
0 times	35 (29)	142 (87)	P<.0001
1 time	40 (33)	10 (6)	
2 or more times	46 (38)	12 (7)	
Number of times called urologist over 12 months			
0 times	92 (76)	158 (96)	P<.0001
1 time	18 (15)	4 (2)	
2 or more times	11 (9)	2 (1)	
Number of PSA tests over 12 months			
0	32 (27)	93 (58)	P<.0001
1	46 (39)	66 (41)	
2 or more	40 (34)	2 (1)	
Number of biopsies over 12 months			
0	100 (85)	159 (99)	P<.0001
1 or 2	18 (15)	2 (1)	

*Number for individual items vary slightly because of nonresponse.
PSA, prostate-specific antigen.

men with a benign biopsy result will have prostate cancer detected on a subsequent biopsy.[19] Therefore, urologists are urged to perform repeat sets of biopsies[20] in men who have suspicious screening tests and initially benign biopsies.

The clinical significance of an elusive prostate cancer detected subsequent to a series of benign prostate biopsies has been questioned.[21] Djavan and colleagues prospectively examined the biochemical and pathological features of cancer detected on biopsies 1, 2, 3, and 4, as well as the biopsy-related morbidity. The investigators found that prostate cancers detected on biopsies 1 and 2 were similar, but that cancers detected on biopsies 3 and 4 had lower grade, stage, and volume compared with biopsies 1 and 2; moreover, the third and fourth biopsies were associated with higher complication rates. When to stop the biopsy cascade that has started, especially for men with conditions known to elevate the PSA level, such as BPH and prostatitis, deserves more attention. This is important because many of these men will have false-positive screening results, which may have psychological and socio-behavioral consequences. While only 1 biopsied man in our study had more than 1 subsequent biopsy during the follow-up year, 25% of the biopsy group reported at the baseline survey having already had 3 or more sets of biopsies, suggesting that the strategy of repeated sets of biopsies is not uncommon.

Our study had a number of limitations. The absence of pre-screening data precluded the determination of whether men in the 2 groups had equivalent baseline psychological profiles. In addition, the 2 groups were not comparable at baseline with regard to history of benign prostatic hyperplasia, prostatitis, previous PSA tests, and previous visits to urologists. However, adjustment for these factors in logistic regression models predicting key outcomes from group membership did not change our findings. Also, men who had a previous prostate biopsy were excluded from the normal PSA group, but not from the biopsy group. However, when we restricted our analyses to include only those men in the biopsy group without a previous biopsy, the findings were essentially unchanged, except that with the reduced power from the restricted sample the difference between the groups responses to the question "In the past month, how much have you thought about prostate cancer?" lost significance at 12 months ($P=.16$). Another potential limitation involves missing data; however, as the amount of missing data was small and the magnitude of the differences between groups was large it is unlikely that missing data made a difference in the findings from our study. Also, we obtained information about worry on the part of the spouse or significant other from the patient, rather than directly from the spouse or significant other[22,23]; however, we believe the perception of the patient regarding worry on the part of their intimate partner is an important issue. We limited our study population to men with a primary care physician at 1 of the 3 participating institutions, anticipating that most of the men would be receiving their health care in that setting. However, men were not asked whether they had any PSA tests or prostate biopsies performed elsewhere, and, therefore, it is conceivable that the number of follow-up PSA tests and prostate biopsies that we obtained from our electronic medical record review at the 3 participating institutions is an underestimation. Lastly, the sample primarily included well-educated white men, and the results may not be generalizable to other racial and ethnic groups, and men with less education. We recommend verification of the results of this study in other samples, particularly African Americans, who are at higher risk for prostate cancer.

In conclusion, we found that even benign prostate biopsy results have psychological, socio-behavioral, and medical consequences. For many men, the benign biopsy result does not put the question of prostate cancer to rest; but rather, is associated with additional urology visits, PSA testing, and prostate biopsies, all of which have consequences for the patient and his family. We do not know the relative contribution of patient and urologist concern to the patterns observed, but it is certainly clear that men with benign biopsies receive more follow-up medical care than those with normal PSA results. These hidden tolls associated with screening should be considered in the discussion about the benefits and risks of prostate cancer screening, particularly in men with benign prostatic hyperplasia or prostatitis, who are at higher risk of false-positive screening results. Although it may be the path of greater resistance,[24] physicians will better serve patients by acknowledging that screening for prostate cancer, although an attractive option for many, is not the best option for all.[25]

We are extremely grateful to the primary care physicians from Massachusetts General Hospital, Brigham and Women's Hospital, and Boston Medical Center who gave us permission to contact their patients.

REFERENCES

1. **Harris R, Lohr KN.** Screening for prostate cancer: an update of the evidence for the U.S. Preventive services task force. Ann Intern Med. 2002;137:917–29.
2. **Sirovich BE, Schwartz LM, Woloshin S.** Screening men for prostate and colorectal cancer in the United States: does practice reflect the evidence? JAMA. 2003;289:1414–20.
3. **Stamey T, Caldwell M, McNeal J, Nolley RMH, Downs J.** The prostate specific antigen era in the United States is over for prostate cancer: what happened in the last 20 years? J Urol. 2004;172:1297–301.
4. **McNaughton-Collins M, Fowler F, Caubet J, et al.** Psychological effects of a suspicious prostate cancer screening test followed by a benign biopsy result. Am J Med. 2004;117:719–25.
5. **Essink-Bot ML, de Koning HJ, Nijs HG, Kirkels WJ, van der Maas PJ, Schroder FH.** Short-term effects of population-based screening for prostate cancer on health-related quality of life. J Natl Cancer Inst. 1998;90:925–31.
6. **Gustafsson O, Theorell T, Norming U, Perski A, Ohstrom M, Nyman CR.** Psychological reactions in men screened for prostate cancer. Br J Urol. 1995;75:631–6.
7. **Taylor KL, DiPlacido J, Redd WH, Faccenda K, Greer L, Perlmutter A.** Demographics, family histories, and psychological characteristics of prostate carcinoma screening participants. Cancer. 1999;85:1305–12.
8. **Cantor SB, Volk RJ, Cass AR, Gilani J, Spann SJ.** Psychological benefits of prostate cancer screening: the role of reassurance. Health Expect. 2002;5:104–13.
9. **Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF.** Psychological and behavioral implications of abnormal mammograms. Ann Intern Med. 1991;114:657–61.
10. **Gram IT, Lund E, Slenker SE.** Quality of life following a false positive mammogram. Br J Cancer. 1990;62:1018–22.
11. **Lidbrink E, Elfving J, Frisell J, Jonsson E.** Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the stockholm trial [see comments]. BMJ. 1996;312:273–6.
12. **Ellman R, Angeli N, Christians A, Moss S, Chamberlain J, Maguire P.** Psychiatric morbidity associated with screening for breast cancer. Br J Cancer. 1989;60:781–4.
13. **Lafata J, Simpkins J, Lamerato L, Poisson L, Divine G, Johnson C.** The economic impact of false-positive cancer screens. Cancer Epidemiol Biomarkers Prev. 2004;13:2126–32.

14. **Ford M, Havstad S, Demers R, Johnson C.** Effects of false-positive prostate cancer screening results on subsequent prostate cancer screening behavior. Cancer Epidemiol Biomarkers Prev. 2005;14:190–4.
15. **Schwartz L, Woloshin S, Fowler F Jr, Welch H.** Enthusiasm for cancer screening in the United States. JAMA. 2004;291:71–8.
16. **Humphrey LL, Helfand M, Chan BK, Woolf SH.** Breast cancer screening: a summary of the evidence for the U.S. Preventive services task force. Ann Intern Med. 2002;137:347–60.
17. **Pignone M, Rich M, Teutsch SM, Berg AO, Lohr KN.** Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive services task force. Ann Intern Med. 2002; 137:132–41.
18. **Roehrborn CG, Pickens GJ, Sanders JS.** Diagnostic yield of repeated transrectal ultrasound-guided biopsies stratified by specific histopathologic diagnoses and prostate specific antigen levels. Urology. 1996; 47:347–52.
19. **Djavan B, Zlotta AR, Ekane S, et al.** Is one set of sextant biopsies enough to rule out prostate cancer? Influence of transition and total prostate volumes on prostate cancer yield [in process citation]. Eur Urol. 2000;38:218–24.
20. **Levine MA, Ittman M, Melamed J, Lepor H.** Two consecutive sets of transrectal ultrasound guided sextant biopsies of the prostate for the detection of prostate cancer. J Urol. 1998;159:471–5; Discussion 75–6.
21. **Djavan B, Ravery V, Zlotta AR, et al.** Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3, and 4: when should we stop? J Urol. 2001;166:1679–83.
22. **Volk R, Cantor S, Cass A, Spann S, Weller S, Krahn M.** Preferences of husbands and wives for outcomes of prostate cancer screening and treatment. J Gen Intern Med. 2004;19:339–48.
23. **Volk R, Cantor S, Spann S, Cass A, Cardenas M, Warren M.** Preferences of husbands and wives for prostate cancer screening. Arch Fam Med. 1997;6:72–6.
24. **Ransohoff D, McNaughton-Collins M, Fowler F Jr.** Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback. Am J Med. 2002;113:663–9.
25. **Wilt T, Partin M.** Reducing psanxiety: the importance of noninvasive chronic disease management in prostate cancer detection and treatment. Am J Med. 2004;117:796–8.

Health Status Among 28,000 Women Veterans

The VA Women's Health Program Evaluation Project

Susan M. Frayne, MD, MPH,[1,2,3] Victoria A. Parker, DBA,[3,4] Cindy L. Christiansen, PhD,[3,4] Susan Loveland, MAT,[3,4] Margaret R. Seaver, MD, MPH,[5,6] Lewis E. Kazis, ScD,[3,4] Katherine M. Skinner, PhD[3,4]✠

[1]Center for Health Care Evaluation, VA Palo Alto Health Care System, Menlo Park, CA, USA; [2]Division of General Internal Medicine; Stanford University School of Medicine, Stanford, CA, USA; [3]Center for Health Quality, Outcomes and Economic Research, VA Medical Center, Bedford, MA, USA; [4]Boston University School of Public Health, Boston, MA, USA; [5]VA Boston Healthcare System, Boston, MA, USA; [6]Boston University School of Medicine, Boston, MA, USA.

BACKGROUND: Male veterans receiving Veterans Health Administration (VA) care have worse health than men in the general population. Less is known about health status in women veteran VA patients, a rapidly growing population.

OBJECTIVE: To characterize health status of women (vs men) veteran VA patients across age cohorts, and assess gender differences in the effect of social support upon health status.

DESIGN AND PATIENTS: Data came from the national 1999 Large Health Survey of Veteran Enrollees (response rate 63%) and included 28,048 women and 651,811 men who used VA in the prior 3 years.

MEASUREMENTS: Dimensions of health status from validated Veterans Short Form-36 instrument; social support (married, living arrangement, have someone to take patient to the doctor).

RESULTS: In each age stratum (18 to 44, 45 to 64, and $\geq$65 years), Physical Component Summary (PCS) and Mental Component Summary (MCS) scores were clinically comparable by gender, except that for those aged $\geq$65, mean MCS was better for women than men (49.3 vs 45.9, $P<.001$). Patient gender had a clinically insignificant effect upon PCS and MCS after adjusting for age, race/ethnicity, and education. Women had lower levels of social support than men; in patients aged <65, being married or living with someone benefited MCS more in men than in women.

CONCLUSIONS: Women veteran VA patients have as heavy a burden of physical and mental illness as do men in VA, and are expected to require comparable intensity of health care services. Their ill health occurs in the context of poor social support, and varies by age.

KEY WORDS: women's health; veterans; health status; quality of life; social support.
DOI: 10.1111/j.1525-1497.2006.00373.x

Address correspondence and requests for reprints to Dr. Frayne: Center for Health Care Evaluation, 795 Willow Road (152-MPD), Menlo Park, CA 94025 (e-mail: sfrayne@stanford.edu).

✠*Deceased.*

J GEN INTERN MED 2006; 21:S40–46.

Robust empiric data and decades of clinical experience have established that male veterans receiving care in the Veterans Health Administration (VA) have worse health status than do men in the general population.[1–4] However, much less is known about the health status of *women* veterans in the VA health care system.

This group deserves attention: there were 1.6 million women veterans in the United States as of 2000,[5] and their ranks will continue to expand as women have an increasing presence in the military. With this ongoing demographic shift, VA clinicians and policy makers need to understand various health care issues unique to this population, including, at the most fundamental level, their overall health status.

Some known characteristics of women veteran VA patients would be expected to be associated with more favorable health status, others with less favorable health status. For example, women veterans are on average younger than their male counterparts,[6] and younger patients tend to be in better health. Similarly, women are less likely than men to have seen combat,[7] a potential source of physical injury, emotional trauma, and disability. Conversely, prior work has shown that the rate of military sexual trauma is substantially higher in women veterans than in male veterans[8–10]; sexual trauma is associated with decrements in health.[11–13] War zone exposure likewise correlates with poor physical health in women.[14] There is little available information about the net effect of these and other characteristics of women veterans upon health status.[15]

Contextual factors mediate the impact of illness upon functional health status. Social support is a particularly powerful contextual factor, affecting a range of health outcomes,[16–20] sometimes differently in woman than in men.[21] The typical woman veteran makes a nontraditional career choice in early adulthood; that decision could have a lifelong series of repercussions upon higher education, marriage, childbearing, employment, and connectedness to social networks.[22] Social ramifications of military service may be less pronounced for men. Thus, the health status of women veterans must be understood against the backdrop of their social support structures.

Therefore, we used existing national survey data to examine the health status of women veterans in VA, benchmarking

them against male veterans, whose health issues are much better understood. Our specific questions were:

1. Does the health status of women veterans differ from that of male veterans, across age strata?
2. What is the contribution of female gender to physical and mental health status after accounting for age, race/ethnicity, and education?
3. Does social support influence health status differently in women veterans than in men veterans?

METHODS

Data Source

Data came from the 1999 Large Health Survey of Veteran Enrollees, a cross-sectional, mailed survey assessing health and functional status in a national sample of VA enrollees. A random sample of 1.5 million veterans enrolled in VA as of March 1999 (from a sampling frame of 3.4 million) received a mailed questionnaire between July 1999 and January 2000, using a modified Dillman Approach with 4 carefully spaced mailings over 12 weeks.[23]

Patient Sample

From the 887,775 respondents (63% of the 1,406,049 surveyed who were living and had valid names/addresses),[24] we excluded patients for whom gender or age (ranging from 18 to 98 years) could not be ascertained (N=39,296, 4.4%), or who reported in the survey that they had received no VA care in the prior 3 years (N=168,620, 19.0%). Six hundred seventy-nine thousand eight hundred and fifty-nine patients (28,048 women and 651,811 men) constituted the analytic cohort for this study. This secondary data analysis was approved by the human studies committee at Edith Nourse Rogers Memorial Veterans Hospital.

Variables

Patient gender and age were available from VA administrative data. Because patients could select as many race/ethnicity categories as applied, we used a hierarchy to assign race/ethnicity so as to maximize identification of minority subgroups. Specifically, if a patient selected more than one race/ethnicity category (true for 4% of patients), she/he was assigned to the lowest prevalence race/ethnicity group selected. As "white" was the highest prevalence group, a patient was "white" only if she/he selected the category "white" and no other categories. Education and employment status likewise came from self-report.

The 1999 Large Health Survey of Veteran Enrollees included 3 indicators of social support, originally developed for the Veterans Health Study from established constructs of social support[25]: whether the patient was currently married, whether the patient was living alone, and whether the patient had someone who could take him/her to the doctor if needed. The items were completed by 96%, 92%, and 99% of subjects, respectively. These indicators of social support have important health correlates, and can influence health outcomes differently.[16,26–30] Survey items were also used to determine whether VA was the patient's exclusive source of care in the past year and in the past 3 years.

To assess health status, we used scales from the Veterans SF-36 health survey instrument, a version of the Short Form 36 (SF-36) modified for use in veteran populations. The SF-36 has been used in diverse patient populations, and has well-established validity and reliability.[31–34] The SF-36 examines 8 domains of health: physical functioning (PF), role limitations because of physical problems (RP), bodily pain (BP), general health (GH), role limitations because of emotional problems (RE), mental health (MH), energy/vitality (VT), and social function (SF), each scored from 0 to 100 where 100 denotes best health. The Physical Component Summary (PCS) and Mental Component Summary (MCS) are weighted summaries of physical and mental health status, respectively, using weights derived from a national probability sample of the U.S. population (weighted to a U.S. population mean of 50 and standard deviation of 10).[35] The Veterans SF-36 includes modifications to the role physical and role emotional scales, with improvements to the reliability and validity of the original version.[25,36–38] Among our subjects, 93% completed the Veterans SF-36 items in usable format.

Analyses

The characteristics of veterans using VA care vary by war era.[3,39] Therefore, to better understand the characteristics of various cohorts, we stratified the women and men in our cohort by age: $<$35, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and $\geq$75 years. Then, in age-stratified analyses, we compared socio-demographic characteristics, source of care (VA only or VA plus non-VA), and levels of social support for women versus men.

We used t tests to compare mean Veterans SF-36 scores (8 domains and 2 summary scores) by age, but for parsimony we collapsed age strata into 3 groups ($<$45, 45 to 64, and $\geq$65 years). Based on prior work,[1] we considered a small effect of at least 20% of 1 standard deviation for the 8 Veterans SF-36 scales and the PCS and MCS scores to be clinically significant. In a sensitivity analysis we used least squares means to adjust for age as a continuous variable within age strata, to account for differences in average age by gender. Next, in a multiple regression model we assessed the independent effect of patient gender upon PCS and MCS scores after controlling for age, race/ethnicity, and education.

Finally, we calculated mean PCS and MCS scores, stratified by gender and age, for patients with versus without indicators of social support (married, live with someone, have someone who could take patient to the doctor). We then used t tests to compare those means.

RESULTS

Among veterans using VA care, women tended to be younger than men (Table 1). Across age strata, women were more highly educated than men, and in the youngest cohort, a greater proportion of women than men were nonwhite. Among patients less than 65 years old, women were more likely than men to report that some of their care was provided outside of VA.

Levels of social support were almost universally lower in women than in men, across age strata (Table 1). Women were more likely to be unmarried and to have no one to take them to the doctor if unwell, and women over age 45 years were progressively more likely than men to live alone. For example, among patients 75 years and older, 71.4% of women (vs 30.9% of men) were not married, 52.5% of women (vs 22.1% of men)

Table 1. Sociodemographic, Utilization, and Social Support Characteristics of Female Veterans, Benchmarked Against Male Veterans, by Age Cohort*

Age Cohort	18 to 34		35 to 44		45 to 54		55 to 64		65 to 75		75+	
	F	M	F	M	F	M	F	M	F	M	F	M
N	4,033	14,926	6,956	38,107	6,641	126,252	3,010	116,881	2,574	197,983	4,834	157,662
%	14.4	2.3	24.8	5.9	23.7	19.4	10.7	17.9	9.2	30.4	17.2	24.2
Sociodemographics												
White (%)	59.6	64.5	58.9	58.8	66.2	67.8	79.8	74.8	84.6	80.0	91.9	83.8
Black (%)	26.1	18.1	27.6	25.4	20.7	17.5	9.9	12.8	7.5	10.0	3.2	8.0
Hispanic (%)	7.0	10.1	5.7	8.4	4.3	6.8	2.3	5.6	2.4	5.4	1.4	4.2
Other (%)	7.3	7.3	7.8	7.4	8.8	7.8	8.0	6.7	5.5	4.6	3.5	4.0
College grad (%)	17.7	13.2	25.1	12.4	32.2	14.8	25.1	15.0	19.9	12.3	19.4	11.3
Employed (%)	62.5	73.7	58.8	60.5	51.8	48.0	36.7	36.0	12.1	13.6	4.4	6.0
Source of health care												
VA only past 3 y (%)	25.8	36.2	33.7	42.7	41.1	48.2	48.3	50.2	45.4	44.3	38.9	38.5
VA only past 1 y (%)	31.6	36.9	39.6	44.7	47.1	51.8	55.3	55.7	49.1	48.7	42.6	41.6
Social support indicators												
Not married (%)	60.3	54.6	60.2	51.6	61.8	47.6	66.0	41.1	70.8	32.2	71.4	30.9
Live alone (%)	16.3	18.1	20.0	22.5	28.1	25.2	38.8	25.4	48.1	22.3	52.5	22.1
No one to take to doctor (%)	14.4	10.9	19.6	16.5	21.9	16.5	20.6	13.9	18.5	10.6	15.1	9.6

**This table is descriptive, so tests of statistical significance are not presented; F refers to females, M to males; VA is Veterans Health Administration.*

lived alone, and 15.1% of women (vs 9.6% of men) had no one to take them to the doctor.

In each age stratum (18 to 44, 45 to 64, 65 or higher), PCS and MCS scores were clinically comparable in women and men, except that for those aged $\geq$65 years, mean MCS score was better for women (49.3) than for men (45.9) (Table 2). For the 8 dimensions of the Veterans SF-36, scores of women tended to be comparable to or—in a few instances, for older women—better than those of men.

In a sensitivity analysis (data not shown in Tables; online Appendix 2), we modeled scores for the 8 dimensions and 2 summary scales of the Veterans SF-36 as a function of patient gender and patient age within age strata. The magnitude of difference in scores for women versus men was generally comparable to that shown in Table 2.

Next, we examined PCS and MCS scores as a function of patient gender, age category, race/ethnicity, and education, stratified based on whether or not the patient had received any non-VA care in the prior 3 years (Table 3). Findings were generally similar for patients receiving care in VA only and in those receiving both VA and non-VA care. Patient gender had a clinically insignificant (but positive) effect upon PCS and MCS scores when these other factors were taken into account, certainly much less pronounced than the effect of age or education.

Finally, as Table 4 shows, indicators of social support had a complex effect upon health status. Being married had a paradoxical adverse effect on PCS score in men under age 45; the effect of social support upon PCS was not clinically significant in other groups. Being married or living with someone had a beneficial effect on MCS in men (but not women) under age 65. For example, among patients <45 years old, men living with someone had a 3.4 point better MCS score than did men living alone, whereas women living with someone had only a 1.1

Table 2. Mean SF-36 Scores for Physical and Mental Health Status of Women Versus Men Across Age Groups

	Age 18–44			Age 45–64			Age 65–98		
N	Women 10,989	Men 53,033	Delta*	Women 9,651	Men 243,133	Delta*	Women 7,408	Men 355,645	Delta*
Physical									
PF	66.8	66.0	0.8†	56.4	53.2	3.2‡	45.5	46.2	−0.7‡
RP	53.3	50.7	2.5‡	43.4	35.5	7.9‡	33.0	27.5	5.5‡
BP	50.8	50.3	0.6	46.7	45.0	1.7‡	49.6	47.8	1.8‡
GH	53.7	51.6	2.1‡	51.2	43.8	**7.4**‡	53.7	45.2	**8.5**‡
Emotional/social									
RE	62.2	61.6	0.6	58.9	50.2	8.8‡	60.3	49.0	**11.3**‡
MH	61.8	61.6	0.2	62.2	58.9	3.4‡	72.0	66.9	**5.2**‡
VT	41.5	45.5	−4.0‡	40.3	39.8	0.5	45.3	41.8	3.5‡
SF	58.9	59.1	−0.2	57.7	54.4	3.2‡	64.0	59.9	4.1‡
Summary scores									
PCS (physical)	40.7	40.2	0.6‡	37.1	35.5	1.6‡	33.6	33.0	0.5‡
MCS (mental)	42.8	43.4	−0.6‡	43.8	42.1	1.8‡	49.3	45.9	**3.4**‡

**Delta: Mean score for women minus mean score for men; differences in bold are of clinically meaningful magnitude (i.e., >20% of 1 standard deviation)*
†P<.05 *(Note: exact P-values are presented in online Appendix 1).*
‡P<.001.
PF, Physical Function; RP, Role Functioning, Physical; BP, Bodily Pain; GH, General Health Perception; RE, Role Functioning, Emotional; MH, Mental Health Index; VT, Energy and Vitality; SF, Social Function; PCS, Physical Component Summary; MCS, Mental Component Summary.

Table 3. Parameter Estimates From Regression Models on Physical Component Summary (PCS) and Mental Component Summary (MCS) Scores in Veterans Who Used Veterans Health Administration (VA) Services Only or VA Plus non-VA Services During the Past 3 Years*

	Used VA Care Only in Past 3 y N=300,261	P	Used VA+non-VA Care in Past 3 y N=379,598	P
PCS				
Intercept	40.3	†	39.1	†
Female	0.5	†	0.6	†
Age 65+	−7.7	†	−6.7	†
Age 45 to 64	−5.1	†	−4.4	†
African American	1.1	†	0.0	.64
Hispanic	1.2	†	0.2	.014
Other race	−1.0	†	−1.6	†
College graduate	4.0	†	4.5	†
MCS				
Intercept	42.5	†	43.4	†
Female	1.0	†	0.6	†
Age 65+	2.7	†	2.6	†
Age 45 to 64	−0.8	†	−1.4	†
African American	0.2	.03	−0.4	†
Hispanic	−2.6	†	−3.5	†
Other race	−1.8	†	−1.9	†
College graduate	3.7	†	4.4	†

**The intercept represents the mean PCS (or MCS) score for a white male age <45 years who did not graduate from college. Estimates for other groups can be calculated from these scores. For example, the mean PCS score for a white woman age 65+who graduated from college in the "VA only" group would be: 40.3+(0.5)+(−7.7)+(4.0)=37.1.*

†P<.001.

point better MCS score than did women living alone; this difference in women did not reach the threshold for clinical significance. Having someone who could take the patient to the doctor when needed had a large beneficial effect on MCS in all gender-age groups and especially in patients less than 65 years old.

DISCUSSION

Among veterans using VA care, physical and mental health status are comparable across genders in each age stratum, except that mental health status is better for elderly women than for elderly men. Health status is also comparable across genders after controlling for age, race/ethnicity, and education. The association between indicators of social support and health status is complex; being married or living with someone benefits mental health status in men but not in women (among patients less than 65 years old), whereas having instrumental support benefits mental health status in both men and women.

Overall, the health status of women veterans is comparable to the health status of male veterans, who represent the bulk of VA clinicians' practices, and who are well known to be much sicker, on average, than the general population.[1–4] Our findings are generally similar to prior VA work by Skinner et al.[15] except that in that study, women veterans had even lower mental health function than did a comparison group of men; however, unlike our national study, theirs was conducted at a single tertiary care VA facility which was a referral center for women with posttraumatic stress disorder. Therefore, our findings extend this line of inquiry by being nationally representative.

Comparing mean SF-36 subscale scores of the women veterans in our cohort (who had a mean age of 52 years) to those of women in the general population[31] (who had a mean age of 46 years[40]), women veterans have consistently and markedly worse scores in every domain of physical and mental health (see Fig. 1). Likewise, comparing mean SF-36 subscale scores of the women veterans in our cohort to care-seeking private sector patients in the Medical Outcomes Study (who were 53% female and had a mean age of 55 years),[32] scores in the general population of women veteran VA patients were comparable to or worse than scores of the subset of private sector patients who had "serious chronic medical conditions" (symptomatic congestive heart failure patients, myocardial infarction survivors with recurring angina and/or severe congestive heart failure symptoms, hypertension patients with severe congestive heart failure symptoms and/or history of stroke, and diabetes patients with severe complications). Differences were particularly marked for Bodily Pain (mean score 49.1 vs 65.1 in women veteran VA patients vs care-seeking private sector patients with serious medical conditions, respectively), Role Emotional (60.6 vs 76.2,

Table 4. Difference in Mean Physical Component Summary (PCS) and Mental Component Summary (MCS) Health Status Scores in Patients With Versus Without Indicators of Social Support, by Gender and Age

	Domain of Social Support											
	Married				Live With Someone				Someone Could Take to Doctor			
	Female	P	Male	P	Female	P	Male	P	Female	P	Male	P
PCS												
Age 18 to 44	−1.9	*	**−2.9**	*	−1.4	*	−1.6	*	1.8	*	1.0	*
Age 45 to 64	0.1	.85	−1.2	*	0.0	.95	−0.8	*	2.0	*	0.1	.14
Age 65 to 98	2.0	*	−0.8	*	0.0	.85	−1.1	*	0.9	.02	0.0	.86
MCS												
Age 18 to 44	1.8	*	**3.2**	*	1.1	.005	**3.4**	*	**6.2**	*	**6.6**	*
Age 45 to 64	2.1	*	**3.0**	*	1.6	*	**3.0**	*	**5.7**	*	**5.4**	*
Age 65 to 98	0.6	.06	1.1	*	−0.2	.54	1.1	*	**3.2**	*	**3.8**	*

**Difference significant at P<.001; a positive difference means that patients with more social support have better health status than patients with less social support. Note: Differences in bold face are of clinically meaningful magnitude (i.e., >20% of 1 standard deviation).*
PCS, physical component summary; MCS, mental component summary.

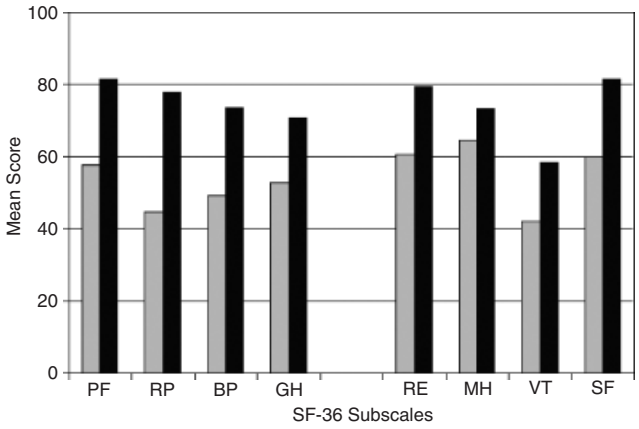


FIGURE 1. Mean Short Form 36 subscale scores (unadjusted) in women veterans in the current study (gray bars) compared with Medical Outcomes Study data[31] for women in the general population (black bars). PF, Physical Function; RP, Role Functioning, Physical; BP, Bodily Pain; GH, General Health Perception; RE, Role Functioning, Emotional; MH, Mental Health Index; VT, Energy and Vitality; SF, Social Function.

respectively) and Social Function (59.8 vs 80.0, respectively).[32] Thus, like male veteran patients, women veterans in VA have particularly poor health status.

The reasons for the ill health of women veterans in VA are unknown; indeed, the finding is somewhat surprising given that good health is a prerequisite to entering the military. It could be that exposures during military service (such as physical injuries, toxic exposures, combat exposure, or military sexual trauma) distinguish women veterans from non-veteran women, adversely affecting their health.[8–10,41–43] High-risk behaviors, such as smoking or substance use, could begin during military service and persist after discharge from the armed services.[44,45] The vast majority of the 1.6 million women veterans in the United States do not receive VA care. It could be that women who elect to use VA services are sicker than other women veterans, perhaps related in part to economic disadvantage.[46]

Another possibility is that the markedly low levels of indicators of social support that we documented among women veterans—which is consistent with prior VA work (where gender comparisons were not available)[22]—contribute to their ill health. While our cross-sectional study could not test this directly, it is well established that low levels of social support are associated with adverse health outcomes.[16–20] Even when married, women tend to bear a heavier burden of care giving and may receive less support themselves than do men.[47,48] This is consistent with our finding that being married or living with someone appears to benefit mental health status in men but not women. In contrast, instrumental support (having someone who could take the patient to the doctor when needed) does benefit mental health status in women. Therefore, VA's efforts to outreach to vulnerable populations—e.g., with in-home care, transportation benefits, and satellite primary care centers—may prove of particular value to women.

Gender differences varied across age cohorts. In particular, while health status summary scores were mostly comparable in women and men, an exception was that women over age 65 years had better mental health summary scores than did men. In this group the effect of social support upon mental health status was also less pronounced. Therefore, it is possible that women in this cohort had access to some types of social support (e.g., networks of friends or qualitatively different types of relationships) less available to men. Alternatively, the women in this oldest group (who represent mostly World War II and Korean era veterans) may have acquired some specific patterns of coping which distinguish them from men of their era.

With ongoing growth of women's representation in the armed services, the health care needs of VA's youngest female enrollees (who may receive VA care for the remainder of their lives) also require special scrutiny. Like their older female counterparts, young women tend to be more highly educated than men in VA, yet are less likely than men to be employed or married or to use VA as their exclusive source of care. VA needs to take possible economic hardship into account when planning care for this emerging population, given the established connections between poverty and ill health.[49,50] Clinicians will also need to monitor the degree to which their care is coordinated across health care systems. Women veterans in the youngest age group did not exceed men in dimensions of the Veterans SF-36 (unlike older women). Therefore, the possibility that, as they age, the newest cohort's health care needs will prove to be greater than those of current cohorts of older women veterans deserves exploration.

Our study must be interpreted subject to several considerations. While the response rate was high for a large national survey, the characteristics of VA patients who responded to the survey could differ from those of veterans who did not. It is also important to recognize that women veterans who use VA services may be different from women veterans who do not use VA; our findings cannot necessarily be extrapolated to the latter population. Finally, because of the cross-sectional nature of our data, causal conclusions about the association between social support and health status cannot be drawn.

Despite these limitations, there are major strengths of our study. We sampled a large proportion of the women veterans who use VA services, maximizing our ability to represent the health status of this special population. To do so, we used a well-validated measure of health status known to correlate with objective outcomes such as severity of medical conditions and mortality.[32,51] We also had the opportunity to assess indicators of social support, a strong but often neglected predictor of health.

Our study has important implications for policy makers, researchers, and health care providers. Strong age cohort effects are seen, suggesting that approaches to providing care for older women veterans may not apply to recent military discharges. Caring for the large subgroup of women with low levels of social support will require interventions sensitive to social context; to compensate for gender role differences in our society, the nature of such interventions may need to be different in women than in men. It is well known that male VHA patients have worse health status than men in the general population; our work demonstrates that female VHA patients are not substantially better off, suggesting they will require comparable intensity of services.

Financial Support: Work for this article was financially supported by the Veterans Health Administration through the Office of

Public Health and Environmental Hazards (Project XVA-65003), by Department of Veterans Affairs Health Services Research & Development grant RCD 98-312, and by Department of Veterans Affairs' Office of Research & Development Project CSF 04-376. Data used in this manuscript were originally collected as part of the 1999 Large Health Survey of Veteran Enrollees, funded by the Department of Veterans Affairs Office for Quality and Performance. The SF-36 is a registered trademark of the Medical Outcomes Trust. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

The authors gratefully acknowledge the VA Office of Public Health and Environmental Hazards, and in particular Susan Mather, MD, MPH, Carole Turner, RN, MN, CNAA, and the National Women Veterans Health Strategic Work Group, for their support for this project. They also appreciate contributions of Rudolf Moos, PhD, to portions of an earlier version of the manuscript. Ms. Ann Walsh provided invaluable technical expertise.

REFERENCES

1. **Kazis LE, Miller DR, Clark J, et al.** Health-related quality of life in patients served by the Department of Veterans Affairs: results from the Veterans Health Study. Arch Intern Med. 1998;158:626–32.
2. **Hankin CS, Spiro A III, Miller DR, Kazis L.** Mental disorders and mental health treatment among U.S. Department of Veterans Affairs outpatients: the Veterans Health Study. Am J Psychiatry. 1999;156:1924–30.
3. **Rosenheck R, Massari L.** Wartime military service and utilization of VA health care services. Mil Med. 1993;158:223–8.
4. **Randall M, Kilpatrick KE, Pendergast JF, Jones KR, Vogel WB.** Differences in patient characteristics between Veterans Administration and community hospitals. Implications for VA planning. Med Care. 1987;25:1099–104.
5. **Richardson C, Waldrop J.** (U. S. Census Bureau, U.S. Department of Commerce). Veterans, 2000, May 2003.
6. **Skinner K, Christiansen C, Frayne S, et al.** (VA Women Veterans Health Program and National Women Veterans Health Strategic Work Group, Office of Public Health and Environmental Hazards). Final Report: Project #XVA 65–003: Status of VA Women Veterans Health Program: Perceptions of Women Veterans Coordinators, Results of a National Survey. 2002.
7. **Friedman MJ, Schnurr PP, McDonagh-Coyle A.** Post-traumatic stress disorder in the military veteran. Psychiatr Clin North Am. 1994;17: 265–77.
8. **Skinner K, Kressin N, Frayne S, et al.** The prevalence of military sexual assault among female Veterans Administration outpatients. J Interpers Violence. 2000;15:289–309.
9. **Murdoch M, Nichol KL.** Women veterans' experiences with domestic violence and with sexual harassment while in the military. Arch Fam Med. 1995;4:411–8.
10. **Turner C, Frayne S.** eds. Veterans Health Initiative: Military Sexual Trauma. Available at: http://www.va.gov/VHI/.Washington, DC: Department of Veterans Affairs Employee Education System and the Center for Women's Health; 2004.
11. **Frayne SM, Skinner KM, Sullivan LM, et al.** Medical profile of women Veterans Administration outpatients who report a history of sexual assault occurring while in the military. J Womens Health Gender Based Med. 1999;8:835–45.
12. **Hankin CS, Skinner KM, Sullivan LM, Miller DR, Frayne S, Tripp TJ.** Prevalence of depressive and alcohol abuse symptoms among women VA outpatients who report experiencing sexual assault while in the military. J Trauma Stress. 1999;12:601–12.
13. **Kimerling R, Clum GA, Wolfe J.** Relationships among trauma exposure, chronic posttraumatic stress disorder symptoms, and self-reported health in women: replication and extension. J Trauma Stress. 2000;13:115–28.
14. **Wolfe J, Schnurr PP, Brown PJ, Furey J.** Posttraumatic stress disorder and war-zone exposure as correlates of perceived health in female Vietnam War veterans. J Consulting Clin Psych. 1994;62:1235–40.
15. **Skinner K, Sullivan LM, Tripp TJ, et al.** Comparing the health status of male and female veterans who use VA health care: results from the VA Women's Health Project. Women Health. 1999;29:17–33.
16. **Ren XS, Skinner K, Lee A, Kazis L.** Social support, social selection and self-assessed health status: results from the Veterans Health Study in the United States. Soc Sci Med. 1999;48:1721–34.
17. **Williams RB, Barefoot JC, Califf RM, et al.** Prognostic importance of social and economic resources among medically treated patients with angiographically documented coronary artery disease. JAMA. 1992;267: 520–4.
18. **Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM Jr.** Social ties and susceptibility to the common cold. JAMA. 1997;277:1940–4.
19. **Case RB, Moss AJ, Case N, McDermott M, Eberly S.** Living alone after myocardial infarction. Impact on prognosis. JAMA. 1992;267: 515–9.
20. **Sherbourne CD, Hays RD, Wells KB.** Personal and psychosocial risk factors for physical and mental health outcomes and course of depression among depressed patients. J Consult Clin Psychol. 1995;63: 345–55.
21. **Berkman LF, Syme SL.** Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol. 1979;109:186–204.
22. **Cotten SR, Skinner KM, Sullivan LM.** Social support among women veterans. J Women Aging. 2000;12:39–62.
23. **Dillman D.** Mail and Telephone Surveys: The Total Design Method. New York: John Wiley and Sons; 1978.
24. **Perlin J, Kazis L, Skinner K, et al.** (Department of Veterans Affairs). Health status and outcomes of veterans: Physical and mental component summary scores, Veterans SF36 1999 Large Health Survey of Veteran Enrollees Executive Report. May 2000.
25. **Kazis LE, Miller DR, Skinner KM, et al.** Patient-reported measures of health: the Veterans Health Study. J Ambul Care Manage. 2004;27: 70–83.
26. **Sherbourne CD, Hays RD.** Marital status, social support, and health transitions in chronic disease patients. J Health Soc Behav. 1990;31: 328–43.
27. **Berkman LF, Leo-Summers L, Horwitz RI.** Emotional support and survival after myocardial infarction. A prospective, population-based study of the elderly. Ann Intern Med. 1992;117:1003–9.
28. **Gove WR, Hughes M, Style CB.** Does marriage have positive effects on the psychological well-being of the individual? J Health Soc Behav. 1983;24:122–31.
29. **Blazer DG.** Social support and mortality in an elderly community population. Am J Epidemiol. 1982;115:684–94.
30. **Thompson MG, Heller K.** Facets of support related to well-being: quantitative social isolation and perceived family support in a sample of elderly women. Psychol Aging. 1990;5:535–44.
31. **Ware JE.** SF-36 Health Survey: Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center; 1993.
32. **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: 247–63.
33. **Ware JE Jr., Sherbourne CD.** The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care. 1992;30:473–83.
34. **Ware JE Jr., Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A.** Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care. 1995;33(4 suppl):264–79.
35. **Ware J, Kosinski M, Keller S.** SF-36 Physical and Mental Health Summary Scales: A Users Manual. Boston, MA: The Health Institute, New England Medical Center; 1994.
36. **Kazis LE, Miller DR, Clark JA, et al.** Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the Veterans Health Study. J Ambul Care Manage. 2004;27:263–80.
37. **Kazis L, Lee A, Spiro III A, et al.** Measurement comparisons of the medical outcomes study and the Veterans SF-36® Health Survey. Health Care Financ Rev. 2004;25:43–58.
38. **Jones D, Kazis L, Lee A, et al.** Health status assessments using the Veterans SF-12 and SF-36: methods for evaluating outcomes in the Veterans Health Administration. J Ambul Care Manage. 2001;24: 68–86.
39. **Villa VM, Harada ND, Washington D, Damron-Rodriguez J.** Health and functioning among four war eras of US veterans: examining the impact of war cohort membership, socioeconomic status, mental health, and disease prevalence. Mil Med. 2002;167:783–9.
40. **McHorney C, Kosinsky M, Ware J.** Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey. Medical Care. 1994;32: 551–67.

41. **Bullman TA, Kang HK.** The effects of mustard gas, ionizing radiation, herbicides, trauma, and oil smoke on US military personnel: the results of veteran studies. Annu Rev Public Health. 1994;15:69–90.
42. **Coyle B, Wolan D.** The prevalence of physical and sexual abuse in women veterans seeking care at a Veterans Affairs Medical Center. Mil Med. 1996;161:588–93.
43. **Proctor SP, Harley R, Wolfe J, Heeren T, White RF.** Health-related quality of life in Persian Gulf War Veterans. Mil Med. 2001;166: 510–9.
44. **Klevens RM, Giovino GA, Peddicord JP, Nelson DE, Mowery P, Grummer-Strawn L.** The association between veteran status and cigarette-smoking behaviors. Am J Prev Med. 1995;11:245–50.
45. **Bradley KA, Bush KR, Davis TM, et al.** Binge drinking among female Veterans Affairs patients: prevalence and associated risks. Psychol Addict Behav. 2001;15:297–305.
46. **Wilson NJ, Kizer KW.** The VA health care system: an unrecognized national safety net. Health Aff (Millwood). 1997;16:200–4.
47. **Wilkins K.** Social support and mortality in seniors. Health Rep. 2003;14:21–34.
48. **Allen SM.** Gender differences in spousal caregiving and unmet need for care. J Gerontol. 1994;49:S187–95.
49. **Epstein AM, Stern RS, Weissman JS.** Do the poor cost more? A multihospital study of patients' socioeconomic status and use of hospital resources. New Engl J Med. 1990;322:1122–8.
50. **Haan M, Kaplan GA, Camacho T.** Poverty and health. Prospective evidence from the Alameda County Study. Am J Epidemiol. 1987;125: 989–98.
51. **Idler EL, Benyamini Y.** Self-rated health and mortality: a review of twenty-seven community studies. J Health and Soc Behav. 1997;38: 21–37.

Supplementary Material

The following supplementary material is available for this article online at www.blackwell-synergy.com

Appendix 1. Standard Deviations and *P*-values for Mean SF-36 Scores Presented in Table S2.

Appendix 2. Mean SF-36 scores in Women and Men by Age Strata, Adjusting for Age as a Continuous Variable Using Least Square Means (Sensitivity Analyses).

ELSEVIER
SAUNDERS

Clin Geriatr Med 22 (2006) 155–166

CLINICS IN
GERIATRIC
MEDICINE

Antithrombotic Therapy in Atrial Fibrillation

David A. Garcia, MD[a,*], Elaine M. Hylek, MD, MPH[b]

[a]*Department of Internal Medicine, University of New Mexico, MSC 10 5550, Albuquerque, NM 87131, USA*
[b]*Department of Medicine, General Internal Medicine Research Unit, Boston University School of Medicine, 91 East Concord Street, Suite 200, Boston, MA 02118, USA*

Atrial fibrillation (AF) is a common dysrhythmia and its prevalence, especially among the elderly, is expected to increase significantly in the coming decades [1]. For men and women 40 years of age and older, the lifetime risk of developing AF is one in four [2]. Because disorganized electromechanical activity can lead to thrombosis within the left atrium, patients with AF at any age have a fivefold increased risk for stroke. An estimated 15% of all strokes occur in individuals with AF [3]. Cerebrovascular accidents related to AF have 25% 30-day mortality and are more likely to result in significant disability than are noncardioembolic strokes [4–6]. Warfarin has been shown to reduce the risk of stroke for patients with AF. Despite its proved efficacy, warfarin continues to be underused particularly among elderly patients who are at the highest risk of stroke. Weighing the benefits of anticoagulation against the risks is more difficult among patients 80 years of age and older because few octogenarians have been enrolled in clinical trials.

Efficacy of warfarin

During the late 1980s and early 1990s, five primary prevention trials and one secondary prevention study yielded consistent results supporting the hypothesis that warfarin can reduce the risk of stroke among patients with AF [7–12]. In a meta-analysis of these studies, Hart and colleagues [13] determined that, compared with placebo, anticoagulation with a vitamin K antagonist, such as

* Corresponding author.
E-mail address: davgarcia@salud.unm.edu (D.A. Garcia).

0749-0690/06/$ – see front matter
doi:10.1016/j.cger.2005.09.011

warfarin, can effect a 62% reduction in the relative risk for stroke among patients with AF. A significant proportion of the strokes reported among the patients assigned to receive warfarin in these trials occurred among patients whose anticoagulation was subtherapeutic. Because trial results are derived from intention-to-treat analyses, it is likely that the relative risk reduction calculated by Hart and colleagues [13] underestimates the power of warfarin to protect AF patients from stroke.

Safety of warfarin

Pooled analysis of the primary stroke prevention trials demonstrates that the annual rate of major hemorrhage among AF patients treated with warfarin is 2.3% (annual rate of intracranial hemorrhage, 0.3%) [14]. Major hemorrhage was defined slightly differently in these trials and could be represented by a bleeding event that required a blood transfusion or an emergency procedure, led to admission, involved the central nervous system, or resulted in prominent residual impairment. Intracranial hemorrhage, because it produces sequelae that are often at least as devastating as ischemic stroke, may be a more important clinical end point. A meta-analysis of six randomized clinical trials indicates that compared with placebo, oral anticoagulation is associated with an absolute risk increase of 0.2% per year for intracranial hemorrhage [13]. This is consistent with the report from a large observational cohort study that the rate of intracranial hemorrhage (per 100 person-years) increased from 0.23 among patients not taking warfarin to 0.46 among patients who were taking warfarin [15]. These findings (ie, that 1 year of warfarin therapy produces an estimated one to two additional intracranial hemorrhages per 1000 patients) have strongly supported the hypothesis that, for most patients with AF, the benefits of warfarin substantially outweigh the risks.

Translating the results of randomized trials into clinical practice

Despite the proved benefit of warfarin and low rates of major hemorrhage, warfarin therapy remains underused in clinical practice [16–22]. The authors of a study assessing the quality of care received by Medicare beneficiaries during the period 1998 to 1999 reported that warfarin was prescribed at hospital discharge to only 42% to 65% of patients with documented AF [23]. There may be several reasons that high-quality evidence of the efficacy of warfarin has not produced more widespread clinical practice change; concerns have been raised about whether the findings of randomized controlled trials (that enrolled highly selected patients who were closely monitored) can be generalized. Indeed, the relatively low enrollment rate among patients screened for the landmark primary prevention studies raises concerns about the external validity of the results of such trials (Table 1). The paucity of elderly participants included in placebo-controlled studies of vitamin K antagonists (see Table 1) is also important because older age has

Table 1
Randomized controlled trials evaluating primary stroke prevention in atrial fibrillation[a]

Study	Design	Randomized / screened	Age comment
AFASAK	warfarin versus ASA versus placebo	1007 / 2546	Median age = 74.2
BAATAF	warfarin versus no warfarin (ASA permitted)	NR	32 / 420 pts. >80
CAFA	warfarin versus placebo	NR	Median age = 68 (warfarin), 67.4 (placebo)
SPAF	Group 1: warfarin versus ASA versus placebo Group 2: ASA versus placebo	1330 / 18376	278 / 1330 pts. >75
SPINAF	Warfarin versus placebo	538 / 7982	88 / 538 pts. >75
SPORTIF III	Warfarin versus ximelagatran (open-label)	3410 / 5188	1146 / 3410 pts. ≥75
SPORTIF V	Warfarin versus ximelagatran (double-blind)	3922 / 4763	1658 / 3922 pts. ≥75

Abbreviations: ASA, acetylsalicylic acid; NR, not reported; pts, patients.

[a] The five primary prevention studies (no shading) that established the efficacy and safety of warfarin and two recent noninferiority studies (shaded) comparing warfarin with ximelagatran are shown.

repeatedly been shown to be an independent risk factor for major bleeding on warfarin [14,24–29]. Some reassurance is provided by the low rates of hemorrhagic stroke (0.1% and 0.4%, respectively) reported among the patients assigned to receive warfarin in two recently published large clinical trials designed to evaluate ximelagatran, Stroke Prevention using an Oral Thrombin Inhibitor in Atrial Fibrillation (SPORTIF V and SPORTIF III) [30,31]. In SPORTIF V, 42% (N = 820) of patients randomized to warfarin were age 75 or greater and 33% (N = 565) were in this age range in SPORTIF III. It is important to note, however, that for SPORTIF V and III, 84% and 74% of patients, respectively, had been taking an oral vitamin K antagonist at the time of randomization. Thus most patients included in these trials were already proved to be at lower risk of hemorrhage.

Like the randomized controlled trials, many observational studies of AF populations have included relatively few patients over the age of 80. A notable exception is the Anticoagulation and Risk Factors In Atrial Fibrillation study, an observational cohort study involving over 11,500 adults with nonvalvular AF. The mean age of enrolled patients was 71 years and 2211 patients taking warfarin were age 75 years or greater. Treatment with warfarin was associated with a 51% lower risk of thromboembolism compared with no warfarin therapy (either no antithrombotic therapy or aspirin) and the rate of intracranial hemorrhage was 0.46% [15]. Although reassuring, studies of prevalent warfarin use may underestimate the rate of major hemorrhage because the early phase of therapy, which is reported to convey the highest risk, is often not included. Observational studies are also subject to selection bias resulting from the physician's initial assessment of an individual's candidacy for long-term anticoagulation. More data are needed to elucidate better the risks and benefits of anticoagulation therapy among patients over age 80 in real-world practice.

Antiplatelet agents

Aspirin is an inexpensive, widely available, and relatively safe medication that has several advantages over warfarin: substantially less potential for drug-drug or drug-diet interactions, wider therapeutic index, and no need for coagulation monitoring. Although a meta-analysis of six randomized controlled trials suggests that aspirin therapy does reduce the risk for ischemic stroke among patients with AF, the protective effect associated with aspirin use is substantially less powerful than that observed with full-intensity warfarin therapy (pooled relative risk reductions compared with placebo are 22% and 62%, respectively) [13]. All six of the individual trials included in the meta-analysis demonstrated a trend favoring aspirin over placebo but only one of these studies (the SPAF study) [9] reported a statistically significant difference. It is noteworthy that in the SPAF study, 52% of the strokes were nondisabling. When only the 12 patients who suffered more severe stroke are considered, the difference between aspirin and placebo in the SPAF study is not statistically significant. Aspirin therapy probably does reduce the risk of stroke in patients with AF but the evidence to support this conclusion is weaker than the evidence for warfarin.

Whether a thienopyridine derivative, such as clopidogrel (either added to or prescribed instead of aspirin), might reduce the risk of stroke for patients with AF is unknown. This class of medications inhibits platelet function by a mechanism different from that of aspirin, and the combination of clopidogrel with aspirin has been shown to be of significant benefit for patients with ischemic heart disease [32]. A large, randomized clinical trial (the ACTIVE study) designed to evaluate the role of clopidogrel for stroke prevention in AF is underway.

Restoring sinus rhythm

Several nonpharmacologic strategies to prevent stroke in AF patients have been proposed; a comprehensive discussion of these is beyond the scope of this article, but important results from trials that examined the use of a strategy to restore and maintain sinus rhythm in AF patients are worthy of mention. Cardioversion for AF patients has several theoretical benefits, including the possibility that if normal atrial electromechanical activity can be re-established, the risk of cardioembolism might be eliminated and antithrombotic therapy would be unnecessary. The strategy of rhythm-control has now been directly compared with rate-control in several randomized, clinical trials that enrolled AF patients at risk for stroke [33–37]. In a pooled analysis that included three of these trials, the frequency of ischemic stroke in patients assigned to rate-control versus the frequency among patients assigned to rhythm-control was comparable (3.5% versus 3.9%, respectively; odds ratio, 0.50; 95% confidence interval, 0.14–1.83; $P = .30$) [38]. Based on these results, the hope that restoring sinus rhythm might obviate the need to anticoagulate AF patients has greatly diminished.

Optimal target international normalized ratio range

The currently recommended anticoagulation intensity for stroke prevention in AF is an international normalized ratio (INR) of 2 to 3 [39]. Numerous studies have documented an increased risk of bleeding with an INR of 4 or greater (Fig. 1) [5,40]. Compared with patients whose INR is >2, AF patients whose INR value is <2 are at increased risk to suffer a stroke; furthermore, the strokes

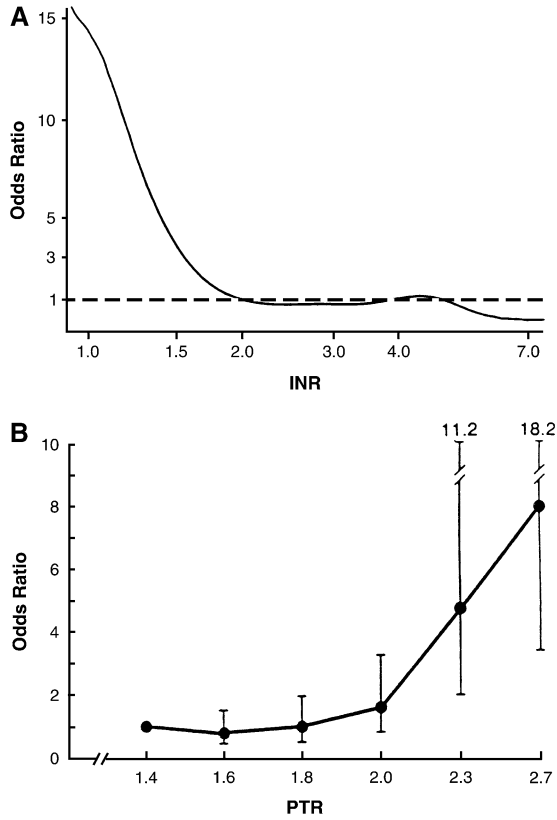


Fig. 1. (*A*) Relationship of odds ratio for ischemic stroke versus International Normalized Ratio (INR) value at, or closest to, the time of the event. Reference value is INR of 2.0. Dashed line corresponds to an odds ratio of 1.0. All cases and control had atrial fibrillation and were treated with warfarin. (*From* Haylek EM, Skates SJ, Sheehan MA, et al. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. N Engl J Med 1996;335:540–6; with permission.) (*B*) Relationship of odds ratio for intracranial hemorrhage versus prothrombin time ratio (PTR) value at, or closest to, the time of the event. In this display the PTR values for the data points are the median values for the following intervals: 1.0–1.5, 1.6–1.7, 1.8–1.9, 2.0–2.1, 2.2–2.3, and 2.4–3.5. The reference interval is 1.0–1.5 (PTR 1.4 median). All cases and controls were taking warfarin. INR equivalent can be roughly approximated as the square of the PTR value. (*From* Haylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med 1994;120:897–902; with permission.)

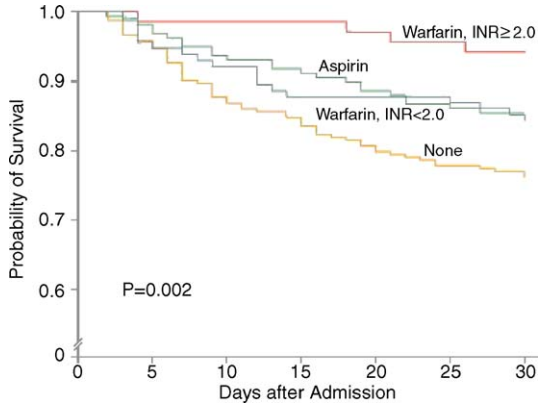


Fig. 2. Kaplan-Meier estimate of survival among nonvalvular AF patients during the 30 days after an ischemic stroke. The patient groups are separated according to medication status at the time of admission. INR, international normalized ratio. (*From* Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. N Engl J Med 2003;349:1019–26; with permission. Copyright 2003 Massachusetts Medical Society. All rights reserved.)

experienced by AF patients with INR values <2 are more likely to result in death or disability (Fig. 2) [5].

Stroke risk assessment for individual patients

Several factors should be considered when determining whether a particular patient with AF should receive warfarin therapy to prevent stroke: their baseline risk for stroke, their risk of bleeding on warfarin therapy, the overall burden of their INR monitoring, and their personal preferences. A number of models and risk classification schemes are now available to assist clinicians in estimating an individual patient's annual risk of stroke [41–45]. Although these models were derived or validated in different populations, they have consistently identified important risk factors that, for patients with AF, are independently associated with an increased risk of stroke. Advancing age, prior stroke, hypertension, heart failure, diabetes, and female sex are examples of such risk factors. A useful resource for estimating an individual patient's risk of stroke was derived by Wang and coworkers from the Framingham Heart Study [45]. The tool is easy to use and can be found at http://www.nhlbi.nih.gov/about/framingham/stroke.htm. An adapted "point-based risk estimate" for the 5-year risk of stroke is reproduced in Fig. 3. Using this model, an 84-year-old woman with a history of diabetes and prior ischemic stroke has an estimated 5-year risk of stroke of 48%. In contrast, a 70-year-old man with well-controlled hypertension has a 5-year stroke risk closer to 7%.

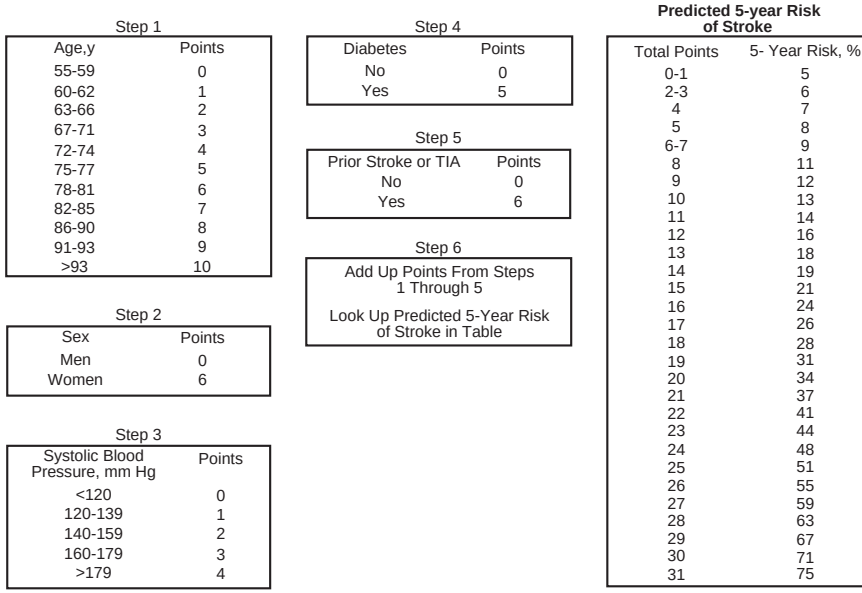


Fig. 3. This point-based scoring system approximates the predicted 5-year risk of stroke for an individual with nonvalvular AF. A more precise risk function is available at http://www.nhlbi.nih.gov/about/framingham/stroke.htm. TIA, transient ischemic attack. (*From* Wang TJ, Massaro JM, Levy D, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. JAMA 2003;290:1049–56; with permission.)

Another recently published risk classification scheme, $CHADS_2$, estimates an AF patient's annual risk of stroke based on the presence or absence of five risk factors (Table 2) [42]. The external validity of the $CHADS_2$ scheme is good because the scoring system was derived from a cohort of 1733 Medicare patients with AF. Although the simplicity of the mnemonic makes it easy to remember, the $CHADS_2$ scoring system may provide less precise risk estimates than the Framingham Model because $CHADS_2$ treats age and blood pressure as dichotomous variables.

Improving the safety margin of anticoagulant therapy in the elderly population

Older age is associated with lower maintenance doses of warfarin. Large initiating doses of warfarin should be avoided in older patients. The warfarin dose schedule should be kept as consistent as possible to minimize dosing confusion. Clinicians should warn patients about (and remain vigilant for) medications known to interact with warfarin, especially amiodarone. A consistent amount of foods rich in vitamin K should be consumed because this may reduce

Table 2
Risk of stroke in National Registry of Atrial Fibrillation participants, stratified by $CHADS_2$ Score[a]

$CHADS_2$ score	No. of patients (N = 1733)	No. of strokes (N = 94)	NRAF crude stroke rate per 100 patient-years	NRAF adjusted stroke rate, (95% CI)[b]
0	120	2	1.2	1.9 (1.2–3)
1	463	17	2.8	2.8 (2–3.8)
2	523	23	3.6	4 (3.1–5.1)
3	337	25	6.4	5.9 (4.6–7.3)
4	220	19	8	8.5 (6.3–11.1)
5	65	6	7.7	12.5 (8.2–17.5)
6	5	2	44	18.2 (10.5–27.4)

Abbreviations: CI, confidence interval; NRAF, National Registry of Atrial Fibrillation.

[a] $CHADS_2$ score is calculated by adding 1 point for recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus, and adding 2 points for having had a prior stroke or transient ischemic attack.

[b] The adjusted stroke rate is the expected stroke per 100 patient-years from the exponential survival model, assuring that aspirin was not taken.

From Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA 2001;285:2864–70.

INR variability. Circumstantial evidence suggests that, for patients treated with warfarin who must take concomitant aspirin, doses $\leq$100 mg may have the most acceptable bleeding risk [46]. Anticoagulated patients who require analgesia should be counseled about the risks of combining certain pain-relieving medications with warfarin. Nonsteroidal anti-inflammatory medications, regardless of their selectivity for cyclooxygenase inhibition, seem to increase the risk of hemorrhage among warfarin users. This association is probably related to some combination of these drugs' effects on both the gastric mucosa and platelet function [47,48]. In the case of acetaminophen, augmentation of warfarin's anticoagulant effect through interference with the enzymes of the vitamin K cycle has been reported [49,50]. Finally, it is important to warn patients about the risk of falling while taking warfarin; measures to optimize a patient's balance should be considered.

Newer antithrombotic strategies

Because warfarin has many negative attributes (narrow therapeutic window, drug-diet interactions, the need for INR monitoring), several clinical trials examining alternative pharmacologic agents have been undertaken in recent years. Ximelagatran, an oral anticoagulant (direct thrombin inhibitor) that does not require coagulation monitoring, has been compared with warfarin therapy (target INR range 2–3) for the prevention of stroke in patients with AF. Although no statistically significant difference in the rate of stroke or major bleeding was observed, ximelagatran is not currently available on the United States market because of other safety concerns that are being addressed (Food and Drug Administration decision letter released October 2004).

Idraparinux, an injectable inhibitor of factor Xa, is another anticoagulant under study among patients with AF. Because it has a prolonged half-life, this agent needs to be administered only once per week. Because its bioavailability is highly predictable, it is believed that coagulation monitoring will be unnecessary in patients using idraparinux. A large phase III study Evaluating the Use of SR34006 Compared to Warfarin or Acenocoumarol in Patients with Atrial Fibrillation (AMADEUS) comparing this drug with warfarin has recently been stopped. Finally, clopidogrel in combination with aspirin is being studied head-to-head with warfarin Atrial Fibrillation Clopidogrel Trial With Irbesartan for Prevention of Vascular Events (ACTIVE) for the prevention of ischemic stroke in patients with AF.

Nonpharmacologic approaches

Other, nonpharmacologic strategies for protecting AF patients from stroke (eg, pulmonary vein isolation, occlusion or removal of the left atrial appendage, and deployment of a polytetrafluoroethylene membrane) are being studied and have been described elsewhere [51–55]. To the authors' knowledge, there is no published high-quality evidence demonstrating that any of these approaches reduces the risk of stroke in an unselected population with AF.

Summary

Warfarin is highly effective at reducing the risk of stroke in AF. The benefit of oral anticoagulant therapy strongly outweighs the risk in most patients with AF. More data are needed to define better the overall risk-to-benefit ratio for patients age 80 years and greater. Because a significant proportion of elderly individuals may not be optimal candidates for anticoagulant therapy, clinicians must continue to evaluate alternative stroke prevention strategies while redoubling efforts to understand the mechanisms underlying AF and thrombogenesis.

References

[1] Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. JAMA 2001;285:2370–5.
[2] Lloyd-Jones DM, Wang TJ, Leip EP, et al. Lifetime risk for development of atrial fibrillation: the Framingham Heart Study. Circulation 2004;110:1042–6.
[3] Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991;22:983–8.
[4] Lin HJ, Wolf PA, Kelly-Hayes M, et al. Stroke severity in atrial fibrillation: the Framingham Study. Stroke 1996;27:1760–4.
[5] Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. N Engl J Med 2003;349:1019–26.

[6] Spratt N, Wang Y, Levi C, et al. A prospective study of predictors of prolonged hospital stay and disability after stroke. J Clin Neurosci 2003;10:665–9.
[7] The European Atrial Fibrillation Trial Study Group. Optimal oral anticoagulant therapy in patients with nonrheumatic atrial fibrillation and recent cerebral ischemia. N Engl J Med 1995; 333:5–10.
[8] The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. N Engl J Med 1990;323:1505–11.
[9] The Stroke Prevention in Atrial Fibrillation Investigators. Stroke Prevention in Atrial Fibrillation Study. Final results. Circulation 1991;84:527–39.
[10] Connolly SJ, Laupacis A, Gent M, et al. Canadian Atrial Fibrillation Anticoagulation (CAFA) Study. J Am Coll Cardiol 1991;18:349–55.
[11] Ezekowitz MD, Bridgers SL, James KE, et al. Warfarin in the prevention of stroke associated with nonrheumatic atrial fibrillation. Veterans Affairs Stroke Prevention in Nonrheumatic Atrial Fibrillation Investigators. N Engl J Med 1992;327:1406–12.
[12] Petersen P, Boysen G, Godtfredsen J, et al. Placebo-controlled, randomised trial of warfarin and aspirin for prevention of thromboembolic complications in chronic atrial fibrillation. The Copenhagen AFASAK study. Lancet 1989;1:175–9.
[13] Hart RG, Benavente O, McBride R, et al. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 1999;131:492–501.
[14] The Stroke Prevention in Atrial Fibrillation Investigators. Bleeding during antithrombotic therapy in patients with atrial fibrillation. Arch Intern Med 1996;156:409–16.
[15] Go AS, Hylek EM, Chang Y, et al. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? JAMA 2003;290: 2685–92.
[16] Gage BF, Boechler M, Doggette AL, et al. Adverse outcomes and predictors of underuse of antithrombotic therapy in Medicare beneficiaries with chronic atrial fibrillation. Stroke 2000; 31:822–7.
[17] Fang MC, Stafford RS, Ruskin JN, et al. National trends in antiarrhythmic and antithrombotic medication use in atrial fibrillation. Arch Intern Med 2004;164:55–60.
[18] Go AS, Hylek EM, Borowsky LH, et al. Warfarin use among ambulatory patients with nonvalvular atrial fibrillation: the anticoagulation and risk factors in atrial fibrillation (ATRIA) study. Ann Intern Med 1999;131:927–34.
[19] 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;159:1574–8.
[20] White RH, McBurnie MA, Manolio T, et al. Oral anticoagulation in patients with atrial fibrillation: adherence with guidelines in an elderly cohort. Am J Med 1999;106:165–71.
[21] Cohen N, Almoznino-Sarafian D, Alon I, et al. Warfarin for stroke prevention still underused in atrial fibrillation: patterns of omission. Stroke 2000;31:1217–22.
[22] Bungard TJ, Ghali WA, Teo KK, et al. Why do patients with atrial fibrillation not receive warfarin? Arch Intern Med 2000;160:41–6.
[23] Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. JAMA 2000;284:1670–6.
[24] The Stroke Prevention in Atrial Fibrillation Investigators. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation: Stroke Prevention in Atrial Fibrillation II Study. Lancet 1994;343:687–91.
[25] Pengo V, Legnani C, Noventa F, et al. Oral anticoagulant therapy in patients with nonrheumatic atrial fibrillation and risk of bleeding: a Multicenter Inception Cohort Study. Thromb Haemost 2001;85:418–22.
[26] Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med 1994;120:897–902.
[27] Steffensen FH, Kristensen K, Ejlersen E, et al. Major haemorrhagic complications during oral anticoagulant therapy in a Danish population-based cohort. J Intern Med 1997;242:497–503.

[28] 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;153:1557–62.
[29] Landefeld CS, Rosenblatt MW, Goldman L. Bleeding in outpatients treated with warfarin: relation to the prothrombin time and important remediable lesions. Am J Med 1989;87:153–9.
[30] Albers GW, Diener HC, Frison L, et al. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. JAMA 2005;293:690–8.
[31] Olsson SB. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. Lancet 2003;362:1691–8.
[32] Wodlinger AM, Pieper JA. The role of clopidogrel in the management of acute coronary syndromes. Clin Ther 2003;25:2155–81.
[33] Opolski G, Torbicki A, Kosior DA, et al. Rate control vs rhythm control in patients with nonvalvular persistent atrial fibrillation: the results of the Polish How to Treat Chronic Atrial Fibrillation (HOT CAFE) Study. Chest 2004;126:476–86.
[34] Wyse DG, Waldo AL, DiMarco JP, et al. A comparison of rate control and rhythm control in patients with atrial fibrillation. N Engl J Med 2002;347:1825–33.
[35] Van Gelder IC, Hagens VE, Bosker HA, et al. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. N Engl J Med 2002;347:1834–40.
[36] Hohnloser SH, Kuck KH, Lilienthal J. Rhythm or rate control in atrial fibrillation. Pharmacological Intervention in Atrial Fibrillation (PIAF): a randomised trial. Lancet 2000;356: 1789–94.
[37] Carlsson J, Miketic S, Windeler J, et al. Randomized trial of rate-control versus rhythm-control in persistent atrial fibrillation: the Strategies of Treatment of Atrial Fibrillation (STAF) study. J Am Coll Cardiol 2003;41:1690–6.
[38] de Denus S, Sanoski CA, Carlsson J, et al. Rate vs rhythm control in patients with atrial fibrillation: a meta-analysis. Arch Intern Med 2005;165:258–62.
[39] Singer DE, Albers GW, Dalen JE, et al. Antithrombotic therapy in atrial fibrillation: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004;126(3 Suppl): 429S–56S.
[40] 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;348:423–8.
[41] Hart RG, Pearce LA, McBride R, et al. Factors associated with ischemic stroke during aspirin therapy in atrial fibrillation: analysis of 2012 participants in the SPAF I–III clinical trials. The Stroke Prevention in Atrial Fibrillation (SPAF) Investigators. Stroke 1999;30:1223–9.
[42] Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA 2001;285: 2864–70.
[43] Atrial Fibrillation Investigators. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation: analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994;154:1449–57.
[44] van Walraven C, Hart RG, Wells GA, et al. A clinical prediction rule to identify patients with atrial fibrillation and a low risk for stroke while taking aspirin. Arch Intern Med 2003;163:936–43.
[45] Wang TJ, Massaro JM, Levy D, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. JAMA 2003;290:1049–56.
[46] Massel D, Little SH. Risks and benefits of adding anti-platelet therapy to warfarin among patients with prosthetic heart valves: a meta-analysis. J Am Coll Cardiol 2001;37:569–78.
[47] Battistella M, Mamdami MM, Juurlink DN, et al. Risk of upper gastrointestinal hemorrhage in warfarin users treated with nonselective NSAIDs or COX-2 inhibitors. Arch Intern Med 2005; 165:189–92.
[48] Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. N Engl J Med 1999;340:1888–99.

[49] Mahe I, Bertrand N, Drouet L, et al. Paracetamol: a haemorrhagic risk factor in patients on warfarin. Br J Clin Pharmacol 2005;59:371–4.
[50] Thijssen HH, Soute BA, Vervoort LM, et al. Paracetamol (acetaminophen) warfarin interaction: NAPQI, the toxic metabolite of paracetamol, is an inhibitor of enzymes in the vitamin K cycle. Thromb Haemost 2004;92:797–802.
[51] Pappone C, Rosanio S, Augello G, et al. Mortality, morbidity, and quality of life after circumferential pulmonary vein ablation for atrial fibrillation: outcomes from a controlled nonrandomized long-term study. J Am Coll Cardiol 2003;42:185–97.
[52] Blackshear JL, Odell JA. Appendage obliteration to reduce stroke in cardiac surgical patients with atrial fibrillation. Ann Thorac Surg 1996;61:755–9.
[53] Sueda T, Imai K, Ishii O, et al. Efficacy of pulmonary vein isolation for the elimination of chronic atrial fibrillation in cardiac valvular surgery. Ann Thorac Surg 2001;71:1189–93.
[54] Oral H, Knight BP, Tada H, et al. Pulmonary vein isolation for paroxysmal and persistent atrial fibrillation. Circulation 2002;105:1077–81.
[55] Sievert H, Lesh MD, Trepels T, et al. Percutaneous left atrial appendage transcatheter occlusion to prevent stroke in high-risk patients with atrial fibrillation: early clinical experience. Circulation 2002;105:1887–9.

Journal of the American College of Cardiology
© 2006 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 47, No. 4, 2006
ISSN 0735-1097/06/$32.00
doi:10.1016/j.jacc.2005.09.058

Coagulation

The Risk of Hemorrhage Among Patients With Warfarin-Associated Coagulopathy

David A. Garcia, MD,* Susan Regan, PhD,† Mark Crowther, MD,‡ Elaine M. Hylek, MD, MPH§
Albuquerque, New Mexico; Boston, Massachusetts; and Hamilton, Ontario, Canada

OBJECTIVES Among warfarin-treated patients with international normalized ratio (INR) >5, we sought to determine the risk of major bleeding within 30 days.

BACKGROUND For warfarin-treated patients, the risk of bleeding increases as the INR rises, particularly if the INR exceeds 4. The 30-day risk of hemorrhage among outpatients with excessively prolonged INR values is unknown.

METHODS To assess anticoagulation care in the U.S., a cohort of 6,761 patients taking warfarin was prospectively assembled from 101 participating sites (43% were community-based cardiology practices). From this cohort, 1,104 patients were identified with a first episode of INR >5.

RESULTS A total of 979 met eligibility criteria; complete follow-up information was available for 976 (99.7%). Ninety-six percent (n = 937) of patients had an INR value between 5 and 9; 80% of INR values were <7. Thirteen patients (1.3%) experienced major hemorrhage during the 30-day follow-up period; among patients whose INR was >5 and <9, 0.96% experienced major hemorrhage. None of the bleeding events was fatal. Intervention with vitamin K was uncommon (8.7%). Warfarin doses were withheld for the majority of patients. Fifty percent of patients who were managed conservatively and retested on day 4 or 5 had an INR of 2.0 or less.

CONCLUSIONS For warfarin-treated outpatients presenting with an INR >5 and <9, the 30-day risk of major bleeding is low (0.96%). Intervention with vitamin K among asymptomatic patients presenting with an INR <9 is not routine practice in the U.S. (J Am Coll Cardiol 2006; 47:804–8) © 2006 by the American College of Cardiology Foundation

Warfarin is a commonly prescribed anticoagulant used to treat or prevent arterial and venous thrombosis. Because warfarin has a narrow therapeutic index, patients routinely undergo phlebotomy or capillary blood sampling to measure a prothrombin time, the laboratory assessment of anticoagulant effect. The prothrombin time is standardized between laboratories as the international normalized ratio (INR). For most indications, an INR value between 2.0 and 3.0 is targeted.

The narrow therapeutic index of warfarin is highlighted by the marked increase in the risk of major hemorrhage associated with INR values that exceed 4.0 (1–4). Although the association between an INR >4 and increased risk of hemorrhage is well documented, published evidence documenting the absolute 30-day bleeding risk for an individual, asymptomatic patient who presents with a supratherapeutic INR is limited. This uncertainty fuels patient fear and physician anxiety. In one prospective study of 114 asymptomatic patients presenting with INR values >6.0, 5 experienced a major hemorrhage during 14 days of follow-up (4.4%, 95% confidence interval [CI] 1.4% to 9.9%) compared to none of 268 patients whose INR was in the target range (5). In a retrospective case series, 3 of 23 patients (12.5%) whose INR values exceeded 10.0 had clinically important bleeding (follow-up interval was not specified) (6). In contrast, two other retrospective studies (in which length of follow-up was not specified) have suggested that for asymptomatic patients whose INR value is >6.0, the risk of major hemorrhage is <1% (7,8). All of the aforementioned studies were limited by small size.

It is widely accepted that patients taking warfarin who present with evidence of active bleeding and INR prolongation should be given coagulation factor replacement (most often in the form of plasma products) and vitamin K (9). For asymptomatic patients with warfarin-associated coagulopathy, however, management strategies vary. Although low-dose oral vitamin K has been shown to return an elevated INR to the therapeutic range more rapidly than placebo (10,11), many providers choose simply to withhold warfarin from such patients, allowing the INR to decline spontaneously (12). This inconsistency of vitamin K use is likely explained by the uncertainty surrounding the short-term risk of bleeding faced by these patients.

To estimate the 30-day risk of hemorrhage among patients with an excessively prolonged INR, we collected data prospectively from a large observational cohort. By better defining the bleeding risk in this population, we hoped to provide guidance to physicians caring for patients with excessive prolongation of their INR. Our secondary aim was to describe the management practices (e.g., frequency of

From the *Department of Medicine, University of New Mexico, Albuquerque, New Mexico; †Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ‡Department of Medicine, McMaster University, Hamilton, Ontario, Canada; and the §Department of Medicine, General Internal Medicine Research Unit, Boston University School of Medicine, Boston Medical Center, Boston, Massachusetts. This study was supported by a grant from Bristol-Myers Squibb (formerly DuPont Pharmaceuticals). Dr. Hylek has received research support from Bristol-Myers Squibb and Pfizer, and lecture honoraria from AstraZeneca. Dr. Garcia has received consulting and lecture honoraria from AstraZeneca. Dr. Crowther has received consulting honoraria from AstraZeneca.

Manuscript received July 1, 2005; revised manuscript received September 8, 2005, accepted September 13, 2005.

Abbreviations and Acronyms
AF = atrial fibrillation
CI = confidence interval
INR = international normalized ratio

vitamin K use and time to follow-up INR testing) employed by the anticoagulation providers who enrolled patients in this study.

METHODS

Site recruitment. A prospective cohort was assembled in order to assess anticoagulation care in the U.S. (13). Registered users of CoumaCare software (Bristol-Myers Squibb [formerly DuPont Pharmaceuticals], Princeton, New Jersey) were invited to participate by letter and through a study website. CoumaCare is a software program that was freely available and used by many anticoagulation management services for clinical purposes such as patient tracking, data entry, and record keeping. The program does not make dosing or follow-up interval recommendations. A total of 174 individual site registrations were received by McKesson HBOC BioServices (Rockville, Maryland), the designated registry specialists for the study. Of these, 101 sites had the technologic capability and the review board approval necessary to participate. All sites had at least one dedicated provider managing warfarin, usually within the setting of a community-based, physician group practice. Patients were invited to participate in the registry either by letter or in person (at the time of a routine appointment). To be eligible for inclusion in the registry, patients had to be 18 years of age or older and provide written informed consent. The participation of sites and patients was voluntary. For all centers, McKesson HBOC provided individual on-site training about how to recruit patients, obtain consent, and transmit data. Adverse event reporting was mandatory, and study personnel were trained to carry out such reporting with rigor sufficient to meet federal regulatory requirements. Enrollment began in April 2000 and ended in February 2002.

Data management. Encrypted data from each site were transmitted to the independent data-coordinating center weekly (McKesson HBOC BioServices). Missing data fields and data entry errors were flagged and resolved directly with the sites before data were transferred to the study investigators. A direct query from the data-coordinating center to the reporting site was triggered by any interval in INR testing that exceeded 45 days *or* any INR value >10 or <0.8. Resolution of the flag relating to the INR testing interval required validation of warfarin status and confirmation that the gap was not related to an adverse event. Study investigators were blinded to the identification and location of participating practices and patients.

Patients and outcomes. Patients whose INR measurement was ≥5.0 were identified. For patients who had more than one such qualifying INR value, only the first occurrence was considered. To be eligible, the index INR value had to: 1) be accompanied by a progress note (to classify provider management); and 2) occur more than 60 days before the site's final study data transmission (to enable complete follow-up by McKesson HBOC).

For each eligible patient, an investigator (D.A.G. or E.M.H.) reviewed the anticoagulation progress notes and follow-up INR values recorded during the 60 days after the index INR measurement. The primary outcome of interest was major hemorrhage, defined as bleeding that was fatal, led to hospitalization with transfusion of at least 2 U of packed red blood cells, or occurred at a critical site (e.g., intracranial, retroperitoneal). All other bleeding was considered minor. Occurrence of an arterial or venous thromboembolic event in the 30-day follow-up period was also recorded. All major events were validated by an investigator (D.A.G. or E.M.H.) and directly verified with the site director by McKesson HBOC. For each patient, the following data were also recorded: number of days warfarin was withheld, number of days until the first repeat INR was performed, number of days until an INR value <4.0 was documented, and the presence or absence of documented vitamin K use.

Statistical analysis. The effect of vitamin K on the risk of major hemorrhage was assessed with logistic regression using a general estimating equation model to account for variation within sites. The model included terms for index INR, age, and vitamin K (14).

The study protocol was approved by Western Institutional Review Board (WIRB), Olympia, Washington, the institutional review board at Massachusetts General Hospital, and local review boards where they existed.

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

In the anticoagulation management study, 6,761 patients were enrolled from 38 states providing 5,961 person-years of observation. Of the 101 clinical sites that participated, 98 were community-based physician office practices, and 3 were academic practices. Forty-three percent of the community-based sites were cardiology group practices. The most common indications for warfarin therapy included atrial fibrillation (AF) (52%), prosthetic heart valve (15%), and venous thromboembolism (14%). Overall, 83% of patients were 60 years of age or greater; 22% were age 80 years or greater. Among patients with AF, 46% had hypertension, 26% coronary artery disease, 15% diabetes mellitus, and 8% history of stroke.

From the overall cohort, 1,104 patients with a first observed INR ≥5 were identified; 979 (89%) met the eligibility criteria. Nearly all of the excluded patients were deemed ineligible because they had an elevated INR recorded fewer than 60 days before the site's final data

Table 1. Characteristics of Patients With an INR ≥5.0 (n = 979)

Mean age	69.5 ± 13 (range = 20–94)
Gender	
Female	494 (50.5)
Race	
White	883 (90.2)
Black	28 (2.9)
Hispanic	16 (1.6)
Asian	7 (0.7)
American Indian	4 (0.4)
Unknown/other	41 (4.2)
Indication for warfarin	
Atrial fibrillation	386 (39.4)
Prosthetic heart valve(s)	280 (28.6)
Venous thromboembolism	135 (13.8)
Cerebrovascular accident	78 (8.0)
CHF/cardiomyopathy/LV thrombus	26 (2.7)
Atherosclerosis, peripheral vascular disease, CAD, myocardial infarction	25 (2.6)
Systemic arterial embolism	11 (1.1)
Other	38 (3.9)
Mean index INR	6.5
Number of patients with index INR >5, <7	783 (80%)
Number of patients with index INR ≥7, ≤9	154 (15.7%)
Number of patients with index INR >9	42 (4.3%)
Patients with documented vitamin K use	85 (8.7%)

CAD = coronary artery disease; CHF = congestive heart failure; INR = international normalized ratio; LV = left ventricular.

transmission. Of the 979 included patients, 39% were receiving warfarin for AF and 29% had a prosthetic heart valve. The mean age was 69 years (range 20 to 94), and 50% were women. Sixty-two patients (6%) had been taking warfarin for <3 months, 96% (n = 937) of the patients had an index INR value between 5 and 9, 80% of index INR values were <7 (Table 1).

Outcomes. ADVERSE EVENTS AND DOCUMENTED VITAMIN K USE. Of the 979 patients, follow-up was complete for 976 (99.7%). Overall, 13 patients (1.3% [95% CI 0.6 to 2.1]) experienced a major hemorrhage during the first 30 days after the index INR; the majority occurred within one week (Table 2). Most of these events were gastrointestinal in origin, and none was fatal. Among the 934 patients with an INR <9, 9 (0.96%) sustained a major hemorrhage. Four (9.5%) of the 42 patients whose INR was 9.0 or greater experienced a major bleed. Thirty-four patients (3.5%) reported minor bleeding; all such cases were self-limited except two patients whose epistaxis required cautery in the emergency department. One patient with a very elevated INR (but no sign of bleeding) was admitted to the hospital for observation and warfarin reversal. Thromboembolic events were infrequent. Three patients (0.3%) sustained an ischemic event during the 30-day follow-up period (two strokes and one peripheral arterial embolism).

Most patients were managed by withholding subsequent warfarin doses. For different subgroups, the mean number of days without warfarin is shown in Table 3. Vitamin K use, as documented in the progress notes in the Coumacare database, was not routine. The progress notes of 85 patients (8.7%) contained evidence that vitamin K was given. The dose of vitamin K was not available for five patients. Compared to patients with an index INR between 5 and 9, vitamin K use was more common among patients whose index INR was >9 (62% vs. 7%). When given, vitamin K was prescribed orally more than 90% of the time; the most common doses were 2.5 mg (60%), 5 mg (24%), and 1.25 mg (11%). In the U.S., vitamin K for oral administration is available only in 5-mg tablets. Of the 85 patients who were treated with vitamin K, one patient who received a 10-mg dose experienced an arterial embolus requiring embolectomy.

Table 2. Age, Gender, Index INR Value, and Site of Hemorrhage for Each of the 13 Individuals With Major Bleeding*

Age (yrs), Gender	INR	Site of Bleed
63, F	13.1	Gastrointestinal
89, M	7.0	Retroperitoneal
69, F	7.7	Gastrointestinal
53, F	8.0	Gastrointestinal
78, M	16.1	Gastrointestinal
74, M	7.0	Gastrointestinal
72, M	11.0	Gastrointestinal
80, F	9.2	Gastrointestinal
79, M	5.6	Gastrointestinal
81, M†	6.4	Soft tissue
86, M†	6.7	Gastrointestinal
60, F†	8.0	Gastrointestinal
66, F†	5.8	Unknown

*The first 9 patients listed experienced major hemmorrhage 0–14 days after the index international normalized ratio (INR) value was recorded. †Patients who experienced bleeding 15–30 days after the index INR was recorded.

The effect of vitamin K on the risk of major hemorrhage was not statistically significant in the univariate analysis (odds ratio 0.85 [95% CI 0.11 to 6.67]; p = 0.88) or in the logistic model adjusting for index INR and age (odds ratio 0.76 [95% CI 0.10 to 6.00]; p = 0.80). The small number of major bleeding events in our study precludes definitive assessment of the relationship between vitamin K use and the risk of hemorrhage.

PATIENT FOLLOW-UP AND SUBSEQUENT INR MEASUREMENT. Of the 979 patients with an INR of 5.0 or greater, 12 patients had no subsequent INR measurements, 6 patients were taken off warfarin, 3 died of causes unrelated

Table 3. Mean Number of Days Warfarin Was Withheld by Index INR and Target INR Range*

	Target INR Range			
	2.0–3.0 (n = 558)		2.5–3.5 (n = 304)	
Index INR	% of Total	Mean Days Held	% of Total	Mean Days Held
5–<7	86% (482)	1.9	88% (268)	1.5
7–9	12% (67)	2.6	11% (33)	2.3
≥9	1.6% (9)	4.2	1% (3)	2.7

*Excluded: 85 patients who were given vitamin K, 16 patients with missing vitamin K data, and an additional 16 patients for whom the number of days held could not be determined.
INR = international normalized ratio.

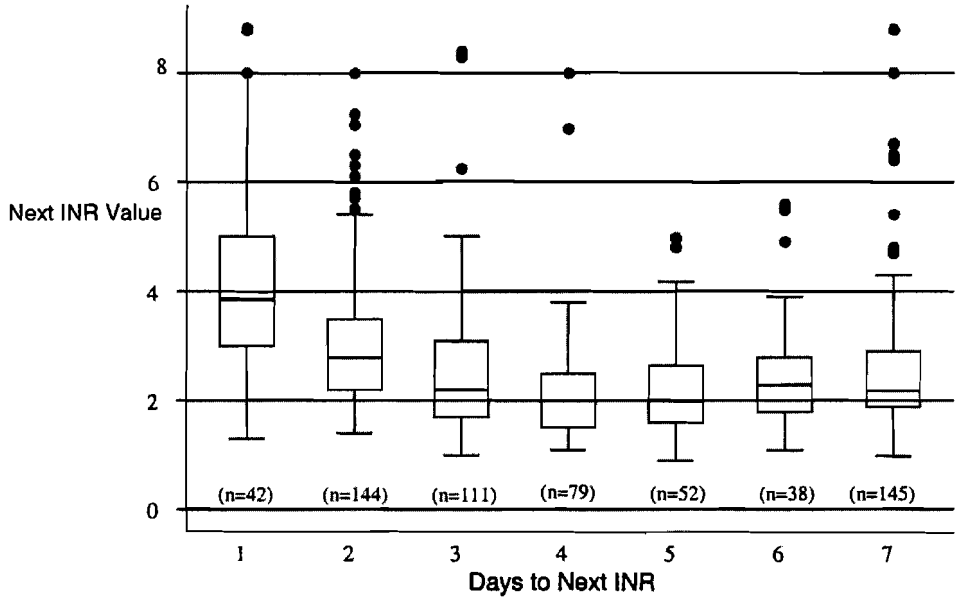


Figure 1. Box and whisker plot of the next consecutive international normalized ratio (INR) value obtained after the index INR, plotted by the number of days between the index and next INR date. The **boxes** delineate the interquartile range, extending from the 25th to the 75th percentile. The **bar across the middle of the box** represents the median. The **whiskers** reach to the most extreme value that is within 1.5 times the interquartile range beyond the box. Points more extreme than that are individually plotted. Patients for whom vitamin K use was documented and patients whose warfarin therapy was not interrupted have been excluded from this analysis.

to warfarin, 2 transferred their warfarin management to another provider, and 1 patient was lost to follow-up. The time elapsed before the first repeat INR varied: 39% of patients were retested within three days, and 75% were retested within the first week. Of the 239 patients whose first follow-up INR occurred 8 or more days after the index INR, 120 (50%) were scheduled for follow-up >7 days after the index INR, 50 (21%) missed an earlier scheduled appointment, 50 (21%) had no documented reason for the follow-up interval chosen, 14 (6%) were hospitalized, and 5 (2%) had other reasons provided. For the patients whose warfarin was held (but for whom no vitamin K was prescribed), Figure 1 shows the median INR result among patients tested on each of the first seven days after the index measurement. Only the first follow-up INR value for any patient was used for this analysis. Of the patients retested on day 4 or 5, nearly 50% had an INR of 2.0 or less.

DISCUSSION

The principal finding of our study is that for warfarin-treated asymptomatic outpatients with an INR value ≥5, the risk of major hemorrhage within 30 days is low (1.3%). It is noteworthy that these INR results were obtained in routine practice, and that 96% of index INR values were <9.0. The likelihood that our results fairly estimate the bleeding risk among such patients is high because our study included nearly 1,000 individuals identified from a cohort of 6,761 patients, 22% of whom were age 80 years or greater. Furthermore, our finding is consistent with two previously reported, smaller, single-center studies of 248 patients and 51 patients, respectively (7,8). The participation of 101 predominantly community-based anticoagulation sites in the present study enhances the generalizability of our findings.

The difference between the proportion of patients suffering major hemorrhage in our cohort and the higher proportion previously reported by Hylek et al. (5) may be explained by several important differences in the two studies. First, the previous study included only 114 patients, and the 95% CI surrounding the point estimate of major hemorrhage risk was wide. Second, the previous study included a population with characteristics that might be expected to increase the overall rate of hemorrhage: 100% of index INR values were >6 (14% were >10), the mean index INR was 8.1, the mean age was 71 years, and 13% of patients had been taking warfarin for <3 months. By comparison, the present study included a population in which only 3% of index INR values were >10, the mean index INR was 6.5, the mean age was 69, and only 6% of patients had been taking warfarin for <3 months. International normalized ratio, age, and the early phase of therapy have all been shown to be risk factors for major hemorrhage among patients taking warfarin (15–19); INR prolongation conveys the highest risk. During 14 days of follow-up in a previous study by Hylek et al. (5), there were no major hemorrhagic events among 268 patients whose INR was in the target range.

Administration of vitamin K, as documented in the CoumaCare database, was uncommon in our cohort despite evidence that low-dose oral vitamin K returns the INR to the normal range more quickly than placebo without causing "over-reversal" of warfarin's anticoagulant effect (10,11,20–22). Published recommendations propose that

low-dose oral vitamin K should be considered for patients with an INR >5 who are at risk for hemorrhage (23,24). However, the absence of high-quality evidence (coupled with the concern that vitamin K use may increase the risk of thromboembolism) likely explains the low frequency of vitamin K use among patients whose INR was <9.

We acknowledge that the patients at highest risk for bleeding may have avoided hemorrhagic events because they received vitamin K. However, because 894 individuals (91% of our cohort) were managed without vitamin K, the potential impact of any such selection bias on our results would be limited, particularly among patients whose INR was <9, for whom documented vitamin K use was particularly uncommon. In our analyses, vitamin K was not found to have a significant effect on the risk of major hemorrhage, but the small number of bleeding events precludes definitive assessment. Because participation in our study was voluntary and required written informed consent, the potential for selection bias does exist. However, our cohort is representative of community-based practice based on patient age and a prevalence of medical conditions that is very similar to that reported from other large ambulatory patient populations. To be eligible, patients had to be 18 years of age or older. Twenty-two percent of the 979 patients in our study cohort were age 80 years or greater. Patients with a prior history of bleeding would have been eligible, if they were still taking warfarin. The risk of major hemorrhage on warfarin is largely driven by anticoagulation intensity, and all patients in our study had an INR of 5.0 or greater. Because the CoumaCare software program does not incorporate decision aids or management tools, it is unlikely that use of this tracking system would have affected an individual's risk of bleeding after an INR of ≥5.

In conclusion, this large cohort study indicates that for asymptomatic outpatients presenting with an INR between 5 and 9, the risk of major hemorrhage over the next 30 days is low (0.96%). The small number (n = 42) of patients whose INR was >9 precludes the extrapolation of our findings to this subgroup. Individuals presenting with INR values over 9 (and perhaps other subpopulations) may indeed have a higher 30-day risk of bleeding. Appropriate assessment of the potential benefit of selected (or routine) vitamin K administration will only be accomplished with randomized, placebo-controlled trials.

Reprint requests and correspondence: Dr. David A. Garcia, 2211 Lomas Northeast, MSC10 5550, Albuquerque, New Mexico 87131. E-mail: davgarcia@salud.unm.edu.

REFERENCES

1. Cannegieter SC, Rosendaal FR, Wintzen AR, van der Meer FJ, Vandenbroucke JP, Briet E. Optimal oral anticoagulant therapy in patients with mechanical heart valves. N Engl J Med 1995;333:11–7.
2. Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med 1994;120:897–902.
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;349:1019–26.
4. Palareti G, Leali N, Coccheri S, et al. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study. Italian Study on Complications of Oral Anticoagulant Therapy (ISCOAT). Lancet 1996;348:423–8.
5. Hylek EM, Chang YC, Skates SJ, Hughes RA, Singer DE. Prospective study of the outcomes of ambulatory patients with excessive warfarin anticoagulation. Arch Intern Med 2000;160:1612–7.
6. Gunther KE, Conway G, Leibach L, Crowther MA. Low-dose oral vitamin K is safe and effective for outpatient management of patients with an INR >10. Thromb Res 2004;113:205–9.
7. Glover JJ, Morrill GB. Conservative treatment of overanticoagulated patients. Chest 1995;108:987–90.
8. Lousberg TR, Witt DM, Beall DG, Carter BL, Malone DC. Evaluation of excessive anticoagulation in a group model health maintenance organization. Arch Intern Med 1998;158:528–34.
9. Ansell J, Hirsh J, Poller L, Bussey H, Jacobson A, Hylek E. The pharmacology and management of the vitamin K antagonists: the seventh ACCP conference on antithrombotic and thrombolytic therapy. Chest 2004;126:204S–33S.
10. Crowther MA, Julian J, Douketis JD, et al. Treatment of warfarin-associated coagulopathy with oral vitamin K: a randomized clinical trial. Lancet 2000;356:1551–3.
11. Crowther MA, Douketis JD, Schnurr T, et al. Oral vitamin K lowers the international normalized ratio more rapidly than subcutaneous vitamin K in the treatment of warfarin-associated coagulopathy. A randomized, controlled trial. Ann Intern Med 2002;137:251–4.
12. Libby EN, Garcia DA. A survey of oral vitamin K use by anticoagulation clinics. Arch Intern Med 2002;162:1893–6.
13. Hylek EM. Trends in anticoagulation management across community-based practices in the United States: the Anticoagulation Consortium to Improve Outcomes Nationally (ACTION) study. J Thromb Thrombolysis 2003;16:83–6.
14. Liang K, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
15. Gullov AL, Koefoed BG, Petersen P. Bleeding during warfarin and aspirin therapy in patients with atrial fibrillation: the AFASAK 2 study. Atrial Fibrillation Aspirin and Anticoagulation. Arch Intern Med 1999;159:1322–8.
16. Turpie AG, Gunstensen J, Hirsh J, Nelson H, Gent M. Randomised comparison of two intensities of oral anticoagulant therapy after tissue heart valve replacement. Lancet 1988;1:1242–5.
17. Hull R, Hirsh J, Jay R, et al. Different intensities of oral anticoagulant therapy in the treatment of proximal-vein thrombosis. N Engl J Med 1982;307:1676–81.
18. Landefeld CS, Goldman L. Major bleeding in outpatients treated with warfarin: incidence and prediction by factors known at the start of outpatient therapy. Am J Med 1989;87:144–52.
19. Landefeld CS, Rosenblatt MW, Goldman L. Bleeding in outpatients treated with warfarin: relation to the prothrombin time and important remediable lesions. Am J Med 1989;87:153–9.
20. Weibert RT, Le DT, Kayser SR, Rapaport SI. Correction of excessive anticoagulation with low-dose oral vitamin K1. Ann Intern Med 1997;126:959–62.
21. Patel RJ, Witt DM, Saseen JJ, Tillman DJ, Wilkinson DS. Randomized, placebo-controlled trial of oral phytonadione for excessive anticoagulation. Pharmacotherapy 2000;20:1159–66.
22. Crowther MA, Donovan D, Harrison L, McGinnis J, Ginsberg J. Low-dose oral vitamin K reliably reverses over-anticoagulation due to warfarin. Thromb Haemost 1998;79:1116–8.
23. Gage BF, Fihn SD, White RH. Management and dosing of warfarin therapy. (review). Am J Med 2000;109:481–8.
24. Ansell J, Hirsh J, Dalen J, et al. Managing oral anticoagulant therapy. Chest 2001;119:22S–38S.

AIDS PATIENT CARE and STDs
Volume 20, Number 5, 2006

Approaching the CDC's Guidelines on the HIV Testing of Inpatients: Physician-Referral versus Nonreferral-Based Testing

JEFFREY L. GREENWALD, M.D.,[1] JONATHAN HALL, L.A.D.C.,[2] and PAUL R. SKOLNIK, M.D.[2]

ABSTRACT

Despite ongoing evidence that one quarter of HIV-infected people in the United States are unaware of their infection, widespread implementation of the Centers for Disease Control and Prevention's 1993 recommendations regarding routine inpatient HIV testing has not occurred. This study compares two HIV testing strategies: the initial phase of inpatient HIV testing (1999–2001) utilized a physician-referral–based system. The second phase (2001–2003), occurring after a patient attitude survey, demonstrated favorable responses to being approached in an unsolicited fashion regarding HIV testing, included the first 2 years' experience with having trained HIV counselors directly approach inpatients regarding their willingness to undergo voluntary HIV counseling and testing (VCT) without physician referral. Barriers to implementing the latter strategy are discussed and initial experience with rapid HIV testing on this service is also presented. Referral-based testing yielded 2.3 patient referrals (6.4% of total admissions) resulting in 1.2 HIV tests and 0.7 counseling only sessions per day. Nonreferral based testing resulted 6.2 HIV tests and another 3.0 counseling-only sessions per day. HIV VCT on an inpatient service is feasible but challenging. Most patients respond favorably to being approached for VCT. Routinely offering HIV tests to inpatients yields higher testing rates than physician referral-based systems and increases the number of patients who know their HIV status. Recommendations for implementing routing HIV testing on an inpatient service are made.

INTRODUCTION

DESPITE SIGNIFICANT ADVANCES in human immunodeficiency virus (HIV) testing and treatment, the Centers for Disease Control and Prevention (CDC) reports that the incidence of new HIV cases in the United States has remained stable at approximately 40,000 cases annually.[1] It estimates that 252,000–312,000 of the 1,039,000–1,185,000 people in the United States with HIV infection do not know their serostatus.[2] To identify this large proportion of the population who are unaware of their status and to provide access to appropriate clinical services, the CDC published recommendations that the medical community routinely incorporate HIV testing during patient medical screening, especially in high-prevalence areas or with high-risk individuals. These recommendations outline specific guidelines for expanding HIV testing.[1]

A decade earlier, the CDC made a number of recommendations about new potential test-

[1]Hospital Medicine Unit, [2]Center for HIV/AIDS Care and Research, Boston Medical Center, Boston, Massachusetts.

ing sites, and stated that "hospitals with an HIV seroprevalence rate of at least 1% or an AIDS diagnosis rate ≥ 1.0 per 1,000 discharges should strongly consider adopting a policy of offering HIV counseling and testing routinely to patients ages 15–54 years."[3] Despite other advances in HIV care, routine HIV testing in hospitals has not occurred (B. Branson, personal communication).

This paper describes the development, implementation, and experience of the HIV Inpatient Testing Service (HITS) at Boston Medical Center (BMC), a large, urban, academic medical center during its first 2 years offering non-physician-referral–based testing and compares it to a similar period prior to HITS when physicians were encouraged to offer HIV testing to all new admissions. It will also address the early phase of rapid HIV test implementation on HITS. This experience may help inform similar hospital-based testing initiatives to increase the number of patients who know their serostatus, facilitate their linkage to care, and guide public health programs toward decreased transmission.

MATERIALS AND METHODS

The initial inpatient testing program was begun in 1999 on a physician-referral basis with a single inpatient medical team and subsequently expanded across the medical service at the hospital. The teams were instructed to ask all of their inpatients whose HIV status was unknown (to the patient or provider) whether they were interested in voluntary HIV counseling and testing (VCT). If the patient was interested, the physician called the BMC HIV VCT program, Project TRUST (Teaching, Referral, Understanding, Support, and Testing), a Massachusetts Department of Public Health (DPH)-funded program. A counselor would come to the patient's bedside to provide VCT. Testing included venipuncture and oral mucosal HIV testing with OraSure® (OraSure Technologies, Inc., Bethlehem, PA). All samples for HIV testing were processed by the Massachusetts State Laboratory free of charge to the patient. All testing results were confidential. Because results often took 7 to 14 days to obtain, patients who had been discharged were given their results either at Project TRUST's office or by special arrangement upon return for other medical follow-up at the medical center.

Based on the favorable results of a quality improvement inpatient attitude survey regarding their feelings about being approached in an unsolicited manner regarding HIV testing,[4] the requirement for physician referral was removed in 2001. Thereafter, the program adopted the name HIV Inpatient Testing Service (HITS) and hired a full-time staff member to approach as many adult patients admitted to the medical service within 24 hours of admission (Monday through Friday) as possible. Although the HITS counselor attempted to see all admissions, the program developed a prioritization scheme to assist the counselor if there were too many new patients to be seen in a day. The priority from highest to lowest was physician referrals, then patients who had infectious or drug-related admitting diagnoses, then other patients under the age of 55.[3] All additional patients were screened as time permitted.

Patients identified as HIV-positive were immediately connected to the BMC Center for HIV/AIDS Care and Research intake services, a dedicated system that addresses medical, social, mental health, and insurance issues on short notice.

With additional funding in 2003, HITS added 1.5 additional VCT staff and subsequently rapid HIV testing with OraQuick® Rapid HIV-1/2 Antibody Test (OraSure Technologies, Inc.) was added. This test, which offers results in 20 minutes, has sensitivity and specificity characteristics similar to commercially available HIV enzyme-linked immunosorbent assay (ELISA) tests, and requires confirmation of all reactive samples by Western blot or IFA testing.[5] HITS was then able to offer VCT to all medicine admissions, and expand services to the surgery and gynecology services, no longer requiring the prioritization scheme.

RESULTS

During the physician-referral–based pilot period (1999–2001), the HIV seroprevalence

averaged 5.5%. An average of 6.4% of total daily admissions (2.3 patients/day) were referred for VCT (Table 1). Counselors met with 83% of these patients; the rest were unavailable, were discharged before the counselor arrived, or refused to speak with the counselor. Of the patients seen, 62% participated in a full VCT session, yielding 1.2 patients per day who underwent testing. Additional information about this initial period has been published previously.[6]

After the 2001 patient survey (Fig. 1) and the creation of HITS, the dedicated full-time counselor was able to increase testing to 6.2 patients per day with an additional 3 patients per day receiving counseling only in the first 12 months of the program. Despite this increase, only one third to one half of all the new admissions were approached because of time and patient availability. As a result of staff changes in the second 12-month period, HITS tested and counseled fewer patients; however, the HIV seroprevalence rates in the first 2 years of HITS were similar: 2.5% and 2.0%, respectively (p = 0.46) (Fig. 1).

TABLE 1. DEMOGRAPHICS OF HITS PATIENTS

	Percent
Gender	
Male	57
Female	43
Race	
White	28
African American	37
Latino	21
Haitian	4
Other[a]	10
Age	
< 20	1
20–29	17
30–39	20
40–49	30
50–59	22
60–69	8
70–79	1
Risk[b]	
MSM	1
IDU	6
Sex w/HIV-positive	1
Sex w/IDU	2
Multiple partners	30
Heterosexual	47
NAR	4
Other	8

[a]Other, Asian, Brazilian, Cape Verdian, Portuguese, Asian, and not recorded.

[b]Risk: MSM, men who have unprotected sex with men; IDU, injection drug user; sex w/HIV+, unprotected sex with an HIV-positive person; sex w/IDU, unprotected sex with an injection drug user; multiple partners, ≥ 10 unprotected sex partners (heterosexual or homosexual); heterosexual, unprotected heterosexual sex; NAR, no acknowledged risk.

HITS, HIV Inpatient Testing Service.

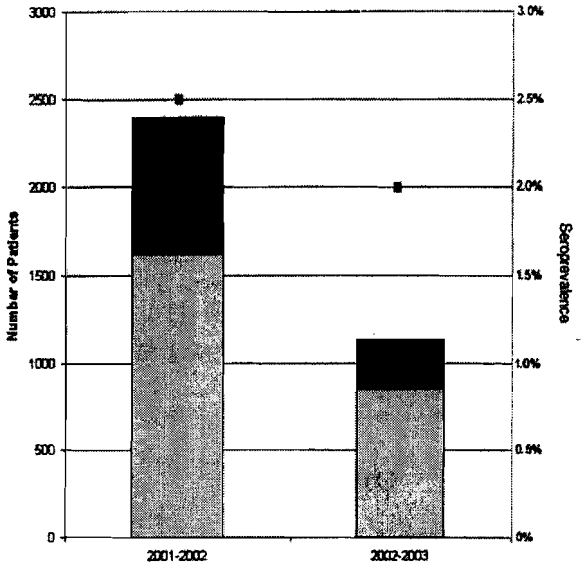


FIG. 1. HIV Inpatient Testing Service (HITS) volume and seroprevalence in the first 2 years. Bars represent HIV voluntary counseling and testing (VCT) for the first 2 years of HITS. Light gray shaded areas of the bars represent the number of VCT sessions that included testing. Dark gray shaded areas represent additional VCT sessions where testing was refused. The black square dots represent the annual HIV seroprevalence for the period represented by the bars.

Seventy-seven percent of patients accepting testing received their test results (range, 43%–90% per quarter). The average inpatient length of stay during this period was 4.4 days; only 31% of patients received their test results while still in the hospital. The rest returned to Project TRUST's outpatient site to obtain their results or were met at subsequent medical center appointments. For those patients who never returned, HITS sought assistance from the patient's primary care providers when subjects had signed the consent to disclose their HIV status to the medical care team. Once HITS began offering rapid HIV tests routinely on the inpatient service, 100% of medical admissions were approached and 100% of those tested (over 4000 tests to date) received their rapid test results the same day they were performed.

An average of 14% of patients per quarter who tested HIV-positive knew previously that they were HIV infected but had not disclosed this fact to the counselor until the posttest session or in follow-up care. Of patients identified to be HIV-infected prior to initiating rapid HIV testing, 82% were successfully linked to HIV care at BMC. Of the remaining 18%, 5% chose to follow-up at another hospital, 5% died before follow-up, 4% informed HITS that they were moving out of the area, and 4% refused care or were lost to follow-up.

DISCUSSION

The 1993 CDC guidelines[3] for implementing HIV testing in areas of high prevalence have not been widely implemented, and limited literature exists on HIV testing in inpatient settings,[7–10] although more than one quarter of patients with HIV are diagnosed while inpatients.[11] How a program should implement HIV testing on an inpatient service will largely depend upon the resources available, however, physician-referral–based systems do not foster broad screening and can skew prevalence estimates because of selection bias. Although the HIV prevalence identified in during the physician-referral–based testing phase (1999–2001) of the program (5.5%) was higher than under HITS (2.0%–2.5%), the absolute number of patients learning their HIV status and those being found to be HIV positive were substantially lower. The elevated seroprevalence in the early phase clearly illustrates this selective referral bias. In the latter phase, this bias may have existed to a lesser degree as a prioritization scheme was occasionally needed to manage workload. Routinely approaching all patients on the inpatient service was not achieved during the primary periods examined in this paper but has subsequently occurred since HITS has expanded its staff and begun offering rapid HIV tests.

As illustrated by the patient survey (Fig. 1), patients have been quite receptive to being approached regarding HIV VCT in an unsolicited manner. Concern over patient reactions to this style of HIV testing program should not be considered a significant barrier to program institution. While we do not have information on why patients refused HIV testing, the recent study by Wurcel et al.[10] suggests that prior testing and self-perception of being at low risk for HIV were the most common reasons.

A number of factors may contribute to the lack of implementation of programs in this area, including funding, logistic barriers, and postdischarge follow-up. Start-up and subsequent financial support for inpatient testing at BMC was facilitated by the Department of Public Health (DPH) and the state laboratory. Although diagnostic testing for HIV is reimbursed as other laboratory services are, HIV counseling services for inpatients are not be reimbursed by private insurance plans or Medicare. Grant funding and a strong relationship with local governmental agencies such as the state DPH and the state or municipal laboratory may help defray costs for such projects.

Another concern is inpatient confidentiality. The physical space of the hospital room is not ideal for HIV counseling. Privacy is a major issue because of visitors, patients in adjacent beds, and hospital staff who might interrupt or overhear the VCT session. Care must be taken to preserve patient confidentiality as much as possible including asking visitors to leave, drawing curtains, keeping counselor voices soft during discussions and deferring testing if conditions do not permit adequate privacy. It is very important that the VCT staff be aware of the events inpatients undergo (e.g., rounds, medication delivery, therapy visits, etc.) so they may tailor their sessions accordingly. Orientation of hospital personnel to the VCT service is equally useful to avoid inopportune comments or questions and to facilitate referrals. Timely communication of results with patients' care providers is also crucial for good patient care.

Because of the complexity of the VCT process, it is important to have an identifiable medical director, program manager, and clinical supervisor. The medical director should have broad knowledge of HIV testing and extensive experience on the inpatient service. The director should be able to address issues such as new testing modalities, training in universal precautions, including various types of isolation precautions; and developing medical and coun-

seling continuing education for staff counselors. Additionally, the medical director needs to coordinate communications with other providers, collect program data for QA, maintain a policies and procedures manual, and facilitate assessment of clinical competency for the hospital and funding agencies. The program manager may share many of these tasks and additionally faces other logistical issues, including obtaining funding, managing grants, hiring staff, coordinating staff schedules, and ensuring service coverage. Staff support issues may be addressed by a clinical supervisor, who should assess counselors' levels of stress and create a forum for case discussion with regular feedback sessions that allow staff to raise issues or make suggestions.

Not all counselors who function well in a walk-in VCT site are suited for an inpatient service. Inpatient counselors must be comfortable with walking into a patient room and engaging the patient in discussion. Counselors who do this process well are outgoing, respectful, and have excellent communication skills. Multilingual staff is a significant benefit if the population served has limited English proficiency. It is also helpful if the counselors come from the population being served. We estimate that hospitals require approximately one full-time VCT staff member for every 80 beds, depending on planned operating hours, the types of testing being offered (standard, rapid, or both), whether posttest counseling is anticipated in the hospital or postdischarge, the acuity of the inpatient population, and the anticipated HIV seroprevalence.

Delays between HIV diagnosis and entry into appropriate care are common. Samet et al.[8] noted 39% of patients did not enter care within 1 year of diagnosis and 18% waited over 5 years. When establishing a VCT service, identifying resources to which patients may be immediately linked is critical. Mental health, medical, social, financial, and other issues must be anticipated and resources allocated to provide rapid interventions for new patients.

HITS found that 14% of patients identified as HIV-positive knew previously that they were infected. Some of these patients may have been looking for a way to get into the medical system. Others may not understand HIV sufficiently to recognize the improper nature of retesting once seropositive. Exploring why retesting occurs offers counselors the opportunity to educate patients further about their disease and, most importantly, get them into care if they are not already linked to an appropriate provider.

Among patients undergoing HIV testing nationally, published reports show that 12% to 39% of patients who are tested for HIV never learn their results.[1,9,12,13] In 2000, the CDC estimated that 31% of patients who tested positive for HIV did not return to receive their test results. State-funded, community-based programs also have highly variable rates of return, with published reports of 25% to 48% of patients never receiving their results.[14–17] Rapid HIV testing resolves some issues surrounding inpatient follow-up for results. A thorough review of rapid HIV testing has been published elsewhere.[18] Confirming reactive tests remains a critical aspect of rapid HIV testing that must be considered when instituting such a program. Many ambulatory sites have had the experience of achieving a greater than 97% follow-up with rapid HIV tests.[11,16,19] To date, 100% of patients tested through HITS with the rapid HIV test have received their rapid test results.

Working with the hospital laboratory staff is crucial in assisting the point-of-care, rapid HIV testing program with set-up and training on the needed procedures. Initial laboratory assistance requirements can be extensive (training, competency assessment for running rapid HIV tests and controls, color blindness testing, universal precautions review, QA monitoring, etc.), and subsequent QA and oversight are important.

Running the rapid HIV test poses other logistical issues. While refrigeration is required only for storing controls, the test requires a flat surface with good lighting. HITS devised a rolling cart to store supplies and to act as a dedicated testing space. The cart has a door that is closed when samples are being processed to avoid the counselor or patient watching the testing process rather than attending to the counseling. The cart also has locking wheels to keep it stationary while the sample is processing. Disinfectant spray is used to clean the test-

ing surface of the cart after each test is completed. The advantage of this system is that it is self-contained and moveable. The disadvantages are that the carts are expensive and, for patients in isolation rooms, HITS leaves the cart just outside the patient's door.

Additionally, point-of-care rapid HIV testing generates no laboratory charges and therefore the cost of the test supplies is not covered by insurance. Program grants and relationships with state and hospital laboratories may help reduce this issue.

CONCLUSIONS

The experience of the BMC HITS program strongly supports routine testing over targeted (physician-referral) based systems and has illuminated a number of specialized issues concerning inpatient HIV VCT. Barriers to implementation of inpatient VCT programs, such as poor follow-up rates for results, are remediable with the advent of point-of-care, rapid HIV tests. Linkage to follow-up care is substantially easier for inpatients since the full resources of the hospital's HIV services may be mobilized for newly diagnosed patients. Referral bias can be overcome with hospital approval to approach all inpatients for HIV VCT without physician referral. Expanding routine testing to inpatient venues offers excellent opportunity to identify HIV patients and to bring them into care expeditiously.

ACKNOWLEDGMENTS

Special thanks to Kenneth Freedberg, M.D., M.Sc., for starting inpatient testing at Boston Medical Center. Additionally, thanks to C. Robert Horsburgh, M.D., M.U.S., Jonathan Pincus, M.D., Michael Paasch-Orlow, M.D., and Carly Bridden, M.P.H., for their editorial assistance in the preparation of this manuscript.

REFERENCES

1. Advancing HIV prevention: New strategies for a changing epidemic—United States, 2003. MMWR Morb Mortal Wkly Rep 2003;52:329–332.
2. Glynn M, Rhodes P. Estimated HIV Prevalence in the United States at the end of 2003. [Abstract T1-B1101]. 2005 National HIV Prevention Conference. Atlanta, GA: June 12–15, 2005.
3. Recommendations for HIV testing services for inpatients and outpatients in acute-care hospital settings. Center for Disease Control and Prevention. MMWR Recomm Rep 1993;42:1–6.
4. Greenwald JL. Routine rapid HIV testing in hospitals: another opportunity for hospitalists to improve care. J Hosp Med 2006;1. [in press]
5. OraQuick Advance Rapid HIV-1/2 Antibody Test [package insert]. Bethlehem, Pa.: OraSure Technologies, Inc. www.orasure.com (Last accessed November 16, 2005).
6. Walensky RP, Losina E, Steger-Craven KA, Freedberg KA. Identifying undiagnosed human immunodeficiency virus: The yield of routine, voluntary inpatient testing. Arch Intern Med 2002;162:887–892.
7. Irwin K, Olivo N, Schable CA, Weber JT, Janssen R, Ernst J. Performance characteristics of a rapid HIV antibody assay in a hospital with a high prevalence of HIV infection. CDC-Bronx-Lebanon HIV Serosurvey Team. Ann Intern Med 1996;125:471–475.
8. Samet JH, Freedberg KA, Stein MD, et al. Trillion virion delay: Time from testing positive for HIV to presentation for primary care. Arch Intern Med 1998;158:734–740.
9. HIV counseling with rapid tests. Center for Disease Control and Prevention. www.cdc.gov/hiv/pubs/rt-counseling.htm. (Last accessed February 17, 2005).
10. Wurcel A, Zaman T, Zhen S, Stone D. Acceptance of HIV antibody testing among inpatients and outpatients at a public health hospital. A study of rapid versus standard testing. AIDS Patient Care STDs 2005;19:499–505.
11. HIV/AIDS Special Surveillance Report. Number 2. Demographic and behavioral data from a supplemental HIV/AIDS behavioral surveillance project 1997–2000. Table 9. Testing location and main reason for HIV test, by race/ethnicity and sex. Centers for Disease Control and Prevention. www.cdc.gov/hiv/stats/special-reportVol1No2/table9.htm (Last accessed March 21, 2005).
12. Tao G, Branson BM, Kassler WJ, Cohen RA. Rates of receiving HIV test results: data from the U.S. National Health Interview Survey for 1994 and 1995. J Acquir Immune Defic Syndr. 1999;22:395–400.
13. National HIV Testing Day at CDC-funded HIV counseling, testing, and referral sites—United States, 1994–1998. MMWR Morb Mortal Wkly Rep 2000; 49:529–532.
14. Keenan PA, Keenan JM. Rapid hiv testing in urban outreach: a strategy for improving posttest counseling rates. AIDS Educ Prev 2001;13:541–550.
15. HIV counseling and testing in publicly funded sites. Annual report, 1997 and 1998. Centers for Disease Control and Prevention. www.cdc.gov/hiv/pubs/cts98.pdf (Last accessed February 17, 2005).

16. Update: HIV counseling and testing using rapid tests—United States, 1995. MMWR Morb Mortal Wkly Rep 1998;47:211–215.
17. Pugatch DL, Levesque BG, Lally MA, et al. HIV testing among young adults and older adolescents in the setting of acute substance abuse treatment. J Acquir Immune Defic Syndr. 2001;27:135–142.
18. Greenwald JL, Pincus J. Rapid HIV testing in the era of OraQuick®. Today's Therapeutic Trends 2003;21: 307–344.
19. Neal JJ, Fleming PL. Frequency and predictors of late HIV diagnosis in the United States, 1994 through 1999. Poster 474M. Presented at: 9th Conference on Retroviruses and Opportunistic Infections. Seattle, WA: February 24–28, 2002.

Address reprint requests to:
Jeffrey L. Greenwald, M.D.
Hospital Medicine Unit, Boston Medical Center
850 Harrison Ave, NIF 6W
Boston, MA 02118

E-mail: Jeffrey.Greenwald@bmc.org

JEFFREY L. GREENWALD

AU1
Is "QA" "quality assurance"?

Routine Rapid HIV Testing in Hospitals:

Another Opportunity for Hospitalists to Improve Care

Jeffrey L. Greenwald, MD[1,2]

[1] Boston University School of Medicine, Boston, Massachusetts

[2] Hospital Medicine Unit, Boston Medical Center, Boston, Massachusetts

BACKGROUND: The Centers for Disease Control and Prevention recommends routinely offering HIV testing to inpatients at hospitals with an HIV seroprevalence rate of greater than 1% or an AIDS diagnosis rate of greater than 1.0 per 1000 discharges. This recommendation has not been widely adopted, perhaps because of one of several barriers: the cost of implementing a counseling and testing program; the logistics of HIV counseling and testing on a hospital ward particularly with respect to privacy; concern about the follow-up of HIV test results necessitating patients to return after discharge; and the cultural mindset of screening as an outpatient modality complicated by the fear of raising the possibility of HIV testing and therefore eliciting a negative reaction from a patient who has not requested it.

PURPOSE: This article focuses on these barriers and some possible solutions, emphasizing the role of FDA-approved rapid HIV tests, which may decrease follow-up issues for HIV testing programs. It also considers hospitalists, given their frontline status and ability to coordinate the multidisciplinary services and systemwide approach required to implement such a program, as leaders in this area.
***Journal of Hospital Medicine* 2006;1:106–112.**

KEYWORDS: rapid HIV test, routine testing, hospitalist, quality.

Despite more than 2 decades of significant advances in human immunodeficiency virus (HIV) testing and treatment and major HIV-oriented public health initiatives, the Centers for Disease Control and Prevention (CDC) reports that the incidence of new HIV cases in the United States has remained stable at about 40 000 cases annually.[1] CDC estimates indicate that 252 000–312 000 of the 1 039 000–1 185 000 people in the United States with HIV infection do not know their serostatus,[2] and it appears that these unaware individuals may play a significant role in HIV transmission to others.[3,4] In an effort to promote testing for HIV, the CDC initiated a program called "Advancing HIV Prevention: New Strategies for a Changing Epidemic" in 2003.[1] This program recommends incorporating HIV testing into routine medical care.

A decade before "Advancing HIV Prevention" was published, the CDC directly addressed the issue of HIV testing of hospitalized patients by recommending that "hospitals with an HIV seroprevalence rate of at least 1% or an AIDS diagnosis rate $\geq$ 1.0 per 1000 discharges should strongly consider adopting a policy of offering HIV counseling and testing routinely to patients ages 15–54 years."[5] Despite the information on discharge diagnosis rates often being easily available from hospital databases, even if seroprevalence rates may not, routine HIV testing of hospitalized patients has not occurred.

DOI 10.1002/jhm.66
Published online in Wiley InterScience (www.interscience.wiley.com).

In 2005 the United States Preventive Services Taskforce (USPSTF) recommendations stated that there was "fair evidence that screening adolescents and adults not known to be at increased risk for HIV can detect additional individuals with HIV."[6] Their statement reflects data from Chen et al., who identified that self-reported risk factor–directed testing strategies would have missed nearly three quarters of the HIV infections in their clinic setting,[7] and from Peterman et al., who demonstrated that 20–26% of HIV-positive patients acknowledged no HIV-associated risk factors.[8]

Despite the prior CDC recommendations,[1,5] Chen and Peterman's data,[7,8] and acknowledgment of the high accuracy of the new HIV antibody tests, making false-positive test results quite rare, the published recommendations of the USPSTF do not support routinely testing individuals who are not at increased risk for acquiring the infection because of the relatively low yield and concern about anxiety and related consequences of HIV testing.

Hospitalists are poised to offer inpatient HIV testing to all inpatients at hospitals that meet the CDC guidelines in an effort to reduce the numbers of patients who have undiagnosed HIV infection. This article examines inpatient HIV testing including barriers that may exist to routine testing and reviews the available rapid HIV tests, which may assist in overcoming some of these barriers.

HIV Testing in the Hospital

Patients diagnosed with HIV infection often have had multiple contacts with the medical community, both inpatient and outpatient, prior to their HIV diagnosis, during which HIV testing had not been offered, thus delaying diagnosis.[9] Though clinicians often identify and document triggers that should prompt HIV testing, patients with HIV infection are still not diagnosed in a timely manner. In addition, according to previously published data on inpatient testing from urban institutions, the targeted testing of patients based on traditional risk factors also misses a large proportion of HIV-infected patients.[10] Thus, routine nontargeted inpatient testing, as the CDC suggests, is the preferred strategy.

More than a quarter of patients with HIV in the United States are diagnosed in hospital settings, often in conjunction with an illness that prompts specific testing.[11] An important recent study by Brady evaluated the HIV seroprevalence on the medicine and trauma medicine services of 2 hospitals during 2 seasons. The study was blinded and used leftover blood samples taken for other reasons. It found seroprevalence rates varying between 1.4% and 3.7%.[12] Two points are noteworthy about this study. First, having excluded those from patients with known HIV disease, a significant proportion of the samples identified as seropositive likely represented unidentified HIV cases. Second, although the seroprevalence varied depending on the season during which testing was done and the service from which blood was obtained, even the lower percentage (1.4%) is higher than the CDC's threshold for offering routine HIV testing.[5]

With the average length of a hospital stay declining to less than 5 days,[13] many patients who undergo nonrapid HIV testing while hospitalized will not receive their results prior to discharge. Though no data specifying the rates of HIV test result follow-up after hospital discharge have been published, the experience in the outpatient setting suggests a significant number of patients never receive their test results. The CDC estimates that 31% of patients who tested positive for HIV did not return to receive their test results.[14] State-funded, community-based programs also have highly variable rates of return, with published reports of 25–48% of patients never receiving their results.[15–17] Fortunately, new and highly accurate rapid HIV tests are now available in the United States, almost eliminating the problem of loss to follow-up[18] (see Rapid HIV Antibody Tests, below).

Barriers to Implementing HIV Testing

There are numerous potential barriers to instituting broad-based screening of hospitalized patients for HIV in addition to the follow-up issues with standard HIV tests illustrated above. These include the cost and cost effectiveness of the program; the logistics of test performance and counseling on the ward; the risk of offending patients; and the culture changes required of inpatient caregivers and hospital administrators. Each of these is addressed briefly.

Cost

Two cost effectiveness analyses examining routine HIV testing have been published recently. The first, by Sanders,[20] assumed a 1% seroprevalence of undiagnosed HIV infection in accordance with CDC recommendations[5] and found a one-time testing cost of $15 078 (2004 dollars) per quality-adjusted life-year (QALY) including the benefit accrued to sexual partners of the tested patient. This cost/

QALY rose to nearly $40 000/QALY with a seroprevalence of only 0.1%. The second study, by Paltiel,[21] demonstrated that the cost/QALY of one-time testing of patients with a 1% seroprevalence to be $38 000.

A few points must be noted about these studies. First, they are not based on inpatient testing specifically. Nonetheless, the Brady study, above,[12] as well as our own experience with routine inpatient testing (unpublished data), suggests that the prevalence may be similar in many inpatient populations. Second, the cost/QALY is very consistent with other routine screening efforts broadly accepted.[22] Finally, although both analyses cited moderately to significantly higher costs/QALY for recurrent (eg, every 3–5 years) routine testing, the relevance of this to routine inpatient testing is less clear.

Another study compared hospitalized patients newly testing HIV positive with a rapid HIV test kit, performed in an emergency department, with those testing HIV positive with conventional HIV tests performed on an inpatient unit.[23] Though it was not designed as a cost analysis, the length of stay of the group that received the rapid test was 7 days shorter than that of the group that received the conventional test (6 vs. 13 days; $P < .001$), with type of HIV testing used identified as an independent effect on length of stay in multivariate regression analysis.

Despite what these analyses reported, start-up costs for HIV testing services can be substantial, and, at present, insurance reimbursement for HIV counseling does not exist. If physicians offer HIV counseling, they may bill for their time as an extended service, when appropriate. Laboratory fees can be billed, which may help to cover materials and processing costs. Grants through the CDC or the Department of Public Health may be available to support programs that operationalize routine HIV testing.

Logistics of Routine Testing on the Ward

An inpatient unit is a difficult place to do HIV counseling. Issues of patient privacy are substantial, especially in shared rooms or when family or friends are present. Physicians and counselors must be cognizant of these issues and be flexible in the timing and structure of the counseling offered to maximize patient comfort and minimize interruptions. Educating inpatient staff about HIV counseling may help to avoid embarrassing situations and interruptions.

In addition, the time required to do HIV testing properly could significantly slow a busy physician's work flow if offered to every patient. Dedicated HIV counseling and testing staff members can be of great assistance in the process and can remove the time barrier from the physician by performing the tests themselves. Such staff members require training in HIV testing procedures if they are to perform point-of-care tests at the bedside. This type of program, coordinated with the leadership of the inpatient service, is ideal for providing routine screening of all admissions as recommended by the CDC.[5] In addition, considerations about minimizing or eliminating pretest counseling are ongoing, with counseling only offered during the posttest phase.[1,24] This plan would also reduce the impact of this process on work flow.

An advantage of using an inpatient service as a site for HIV testing is the ability to mobilize a hospital's resources should a patient be diagnosed as HIV positive. Addressing the medical, psychological, and psychosocial needs of newly diagnosed (or previously diagnosed but medically disconnected) patient requires using a multidisciplinary team approach, including inpatient caregivers, social workers, case managers, mental health providers, and HIV specialists.

Avoiding Offending Patients and Changing Hospital Culture

An inpatient unit is an unusual place for routine screening, which usually is relegated to the ambulatory setting. Moreover, with the stigma of HIV still present, despite efforts to quell it,[25] inpatient caregivers and hospital administrators may be uncomfortable in approaching or having a trained counselor approach all patients on an inpatient service to discuss HIV counseling and testing.

No studies have been published on inpatient attitudes toward routinely being offered HIV testing. Our HIV testing service faced this question when we wanted to expand our inpatient testing from risk-factor-directed and physician-referral-based testing to routine testing. To assess patient responses, we asked 72 medical inpatients how they would feel about an unsolicited offer to be tested for HIV while they were inpatients. The results, displayed in Figure 1, demonstrated that only 11% of the patients had an unfavorable response. Of note, the study did not permit further explanations to be given to dispel the concerns of those whose response was unfavorable. With this information,

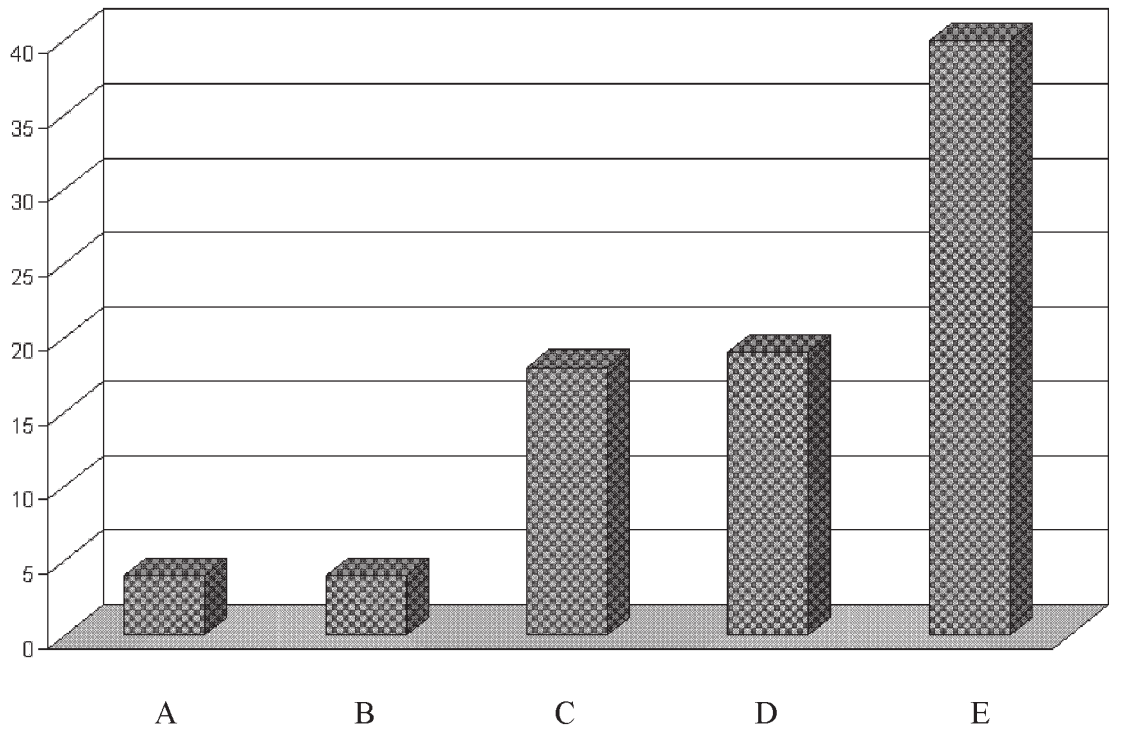


FIGURE 1. Patient attitude survey.

our administration permitted expanded testing to commence.

From the experiences of our testing program, with several thousand patients having been approached, we have found that patients are very rarely offended or upset by being offered HIV testing.

Rapid HIV Antibody Tests in the United States

As noted, a substantial proportion of patients fail to return to obtain results.[15–17] As with other posthospitalization test follow-ups,[26] significant complications may occur if follow-up of HIV test results is inadequate. Rapid HIV antibody tests may offer programs a way to ensure that the vast majority of patients learn their test results.

There are currently 4 rapid HIV tests that have been approved for use in the United States by the Food and Drug Administration (FDA). Two of these, the OraQuick *ADVANCE* Rapid HIV-1/2 Antibody Test® (OraSure Technologies, Inc., Bethlehem,

TABLE 1
United States Food and Drug Administration-Approved Rapid HIV Antibody Tests Performance for HIV-1 Detection*

Rapid HIV Test†	Specimen Type	Sensitivity (95% CI)	Specificity (95% CI)	CLIA Category	Cost
OraQuick Advance Rapid HIV–1/2 Antibody Test	Oral fluid	99.3% (98.4–99.7)	99.8% (99.6–99.9)	Waived	$17.50
	Whole blood (finger stick or venipuncture)	99.6% (98.5–99.9)	100% (99.7–100)	Waived	
	Plasma	99.6% (98.9–99.8)	99.9% (99.6–99.9)	Moderate complexity	
Reveal G-2 Rapid HIV-1 Antibody Test	Serum	99.8% (99.5–100)	99.1% (98.8–99.4)	Moderate complexity	$14.50
	Plasma	99.8% (99.5–100)	98.6% (98.4–98.8)	Moderate complexity	
Uni-Gold Recombigen HIV Test	Whole blood (finger stick or venipuncture)	100% (99.5–100)	99.7% (99.0–100)	Waived	$15.75
	Serum and plasma	100% (99.5–100)	99.8% (99.3–100)	Moderate complexity	
Multispot HIV-1/HIV-2 Rapid Test	Serum	100% (99.94–100)	99.93% (99.79–100)	Moderate complexity	$25.00
	Plasma	100% (99.94–100)	99.91% (99.77–100)	Moderate complexity	

*Modified from Health Research and Education Trust (HRET). Available at http://www.hret.org/hret/programs/hivtransmrpd.html. Accessed May 3, 2005.

PA)[27] and the Uni-Gold Recombigen HIV Test® (Trinity Biotech, Bray, County Wicklow, Ireland),[28] have received a waiver from the Clinical Laboratories Improvement Amendment (CLIA), which means they may be used outside a laboratory setting.[29] Such a waiver means these tests may be used at the bedside of a patient in a point-of-care (POC) fashion similar to that of blood sugar monitoring.

It must be noted, however, that extensive quality assurance and quality control are involved with the use of these POC tests.[30] Despite the CLIA waiver, a relationship with the hospital laboratory is required, as the test kits may only be used by an "agent" of the laboratory. An agent is an individual who the laboratory deems capable and qualified to perform the test competently.

Two additional rapid HIV tests are FDA approved but not CLIA waived. These tests, the Reveal G2 Rapid HIV-1 Antibody Test® (MedMira, Bayers Lake Park, Halifax, Nova Scotia)[31] and the Multispot HIV-1/HIV-2 Rapid Test® (Bio-Rad Laboratories, Redmond, Washington),[32] must be performed in a laboratory (see Table 1).

All 4 tests have sensitivities and specificities similar to those of commercially available standard HIV enzyme immunosorbent assays (EIA) for HIV. As the tests are extremely sensitive, no confirmatory testing is required for nonreactive rapid test results. These tests should be considered negative. False negatives may occur if the patient has had a recent HIV exposure. Thus, as with standard EIA tests, it is important to recommend retesting in 6 weeks for all patients who test HIV negative but who have had a high-risk exposure in the last 3 months. Also, very rarely, patients receiving antiretroviral therapy who have successfully suppressed their viral replication below detectable limits for long periods may also have false-negative results. Therefore, with all patients, it is important to reinforce the idea that it is not appropriate to retest for HIV if a patient already knows he or she is HIV positive.

All reactive rapid HIV tests require confirmation. This process is most commonly done with a Western Blot assay and must be completed before a patient is told that he or she has confirmed HIV infection. Although uncommon, false-positive rapid tests do occur, reinforcing the need for confirmatory testing before a formal diagnosis of HIV infection can be made. Currently, no FDA-approved rapid confirmatory HIV test is available, so standard laboratory delays may be unavoidable for these patients. It is therefore critical that hospitals providing rapid HIV testing have access to medical and social support systems that may be rapidly mobilized for patients with reactive and confirmed positive tests.

Hospitalists at the Helm of Routine Inpatient HIV Testing

Putting a hospitalist in charge of implementing inpatient HIV testing has several advantages. First, as experts in the hospital systems in which they work, hospitalists are prime candidates to organize a multidisciplinary team involving those from nursing, laboratory medicine, mental health, and social

work, as well as HIV specialists. If dedicated HIV counselors are available to participate, they, too, should be included. A hospitalist with an interest in HIV makes an ideal director of such a multidisciplinary program.

Second, hospitalists are on the front line of clinical care and see patients during the earliest hours of their clinical evaluation. By making HIV testing a routine part of all admissions, the hospitalist may act as a role model in the process and will also be able to explain to patients that they are not being singled out, as all patients are encouraged to undergo testing.

Finally, with the demonstrated added value of hospitalist programs[33] and the recent literature demonstrating the cost effectiveness of routine HIV testing,[20,21] hospitalists are well suited to demonstrate leadership in the acquisition of the resources required to make routine inpatient HIV testing possible.

Future Directions

To make routine testing a broadly accepted reality, several developments must begin to take place. These include: increasing education about HIV disease as a chronic disease rather than a rapidly terminal illness;[34] reducing the stigma of HIV disease (a stigma that has impaired testing rates),[25] which should include discussions of eliminating the need for separate HIV test consent forms, not required for testing for other sexually transmitted diseases (eg, syphilis) or life-threatening diseases (eg, hepatitis C);[1] examining the experience and impact of the universal HIV testing recommendations for pregnant women;[35,36] reducing[1,24] or entirely eliminating[37] the requirements for extensive pretest counseling—which may be a low-yield[38] time barrier—with a greater focus on case-specific post-test risk reduction;[1] and broadening the realization that targeted testing based on traditional HIV risk factors fails to identify a significant number of HIV cases.[10,39]

CONCLUSIONS

Though it has been more than a decade since the original CDC recommendations on inpatient HIV testing were released,[5] it remains quite clear that routine inpatient HIV testing can and should be a reality in many hospitals in the United States. As the literature[12] and our institution's experience suggest, those in an inpatient service may be a population with a higher prevalence of HIV disease, and as such, an inpatient service should be a venue where routine HIV testing is offered. The U.S. Preventive Services Taskforce's conclusion that "the benefit of screening adolescents and adults without risk factors for HIV is too small relative to potential harms to justify a general recommendation"[6] may not apply to the inpatient services where HIV disease may be more common than in the general population. However, because of time constraints, busy clinicians may require the assistance of an HIV counseling and testing service to make this kind of program a reality.

Clearly, using targeted testing strategies based on traditional HIV risk factors fails to identify a significant proportion of undiagnosed HIV cases.[7,8] New, FDA-approved rapid HIV antibody tests can help to reduce the issue of loss to follow-up as a barrier to having successful testing programs, and the cost effectiveness of such HIV testing programs has been suggested in recent literature. Although studies are needed to elucidate the differences between routinely tested inpatients and those tested in more traditional ambulatory sites, hospitalists have the opportunity to take the lead in dramatically increasing testing and in substantially decreasing the number of patients unaware of their HIV status.

Address for correspondence and reprints requests: Jeffrey Greenwald, MD, Director, Hospital Medicine Unit, Boston Medical Center, 850 Harrison Ave.–NIF—6W, Boston, MA 02118; Fax: (617) 414-4377; E-mail: Jeffrey.Greenwald@bmc.org

Received 5 October 2005; revision received 25 October 2005; accepted 27 October 2005.

REFERENCES

1. Centers for Disease Control and Prevention. Advancing HIV prevention: new strategies for a changing epidemic—United States, 2003. *MMWR Morb Mortal Wkly Rep.* 2003;52:329–332.
2. Glynn M, Rhodes P. Estimated HIV prevalence in the United States at the end of 2003. 2005 National HIV Prevention Conference; June 12–15, 2005; Atlanta, Ga. Abstract T1–B110.
3. Samet JH, Freedberg KA, Savetsky JB, et al. Understanding delay to medical care for HIV infection: the long-term nonpresenter. *AIDS* 2001;15:77–85.
4. Valleroy LA, MacKellar DA, Karon JM, et al. HIV prevalence and associated risks in young men who have sex with men. Young Men's Survey Study Group. *JAMA.* 2000;284:198–204.
5. Centers for Disease Control and Prevention. Recommendations for HIV testing services for inpatients and outpatients in acute-care hospital settings. *MMWR Recomm Rep.* 1993; 42(RR-2):1–6.

6. US Preventive Services Taskforce. Screening for HIV: recommendation statement. *Ann Intern Med.* 2005;143(1):32–37.
7. Chen Z, Branson B, Ballenger A, Peterman TA. Risk assessment to improve targeting of HIV counseling and testing services for STD clinic patients. *Sex Transm Dis.* 1998;25: 539–543.
8. Peterman TA, Todd KA, Mapanduki I. Opportunities of targeting publicly funded human immunodeficiency virus counseling and testing. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1996;12:69–74.
9. Liddicoat RV, Horton NJ, Urban R, Maier E, Christiansen D, Samet JH. Assessing missed opportunities for HIV testing in medical settings. *J Gen Intern Med.* 2004;19:349–356.
10. Walensky RP, Losina E, Steger-Craven KA, Freedberg KA. Identifying undiagnosed human immunodeficiency virus: the yield of routine, voluntary, inpatient testing. *Arch Intern Med.* 2002;162:887–892.
11. Kates J. Learning more about the HIV-infected population not in care in the US. Poster TuPeG 5690, presented at: XIV International AIDS Conference; July 2002; Barcelona, Spain.
12. Brady KA, Berry AA, Gupta R, et al. Seasonal variation in undiagnosed HIV infection on the general medicine and trauma services of two urban hospitals. *JGIM.* 2005;20:324–330.
13. Hall MJ, DeFrances CJ. 2001 National Hospital Discharge Survey. Advance data from vital and health statistics; no 332. Hyattsville, Md: National Center for Health Statistics; 2003.
14. HIV counseling and testing in publicly funded sites. Annual report, 1997 and 1998. Centers for Disease Control and Prevention [CDC Web site]. Available at: http://www.cdc.gov/hiv/pubs/cts98.pdf. Accessed February 17, 2005.
15. Keenan PA, Keenan JM. Rapid hiv testing in urban outreach: a strategy for improving posttest counseling rates. *AIDS Educ Prev.* Dec 2001;13(6):541–550.
16. Update: HIV counseling and testing using rapid tests—United States, 1995. *MMWR Morb Mortal Wkly Rep.* 1998; 47:211–215.
17. Pugatch DL, Levesque BG, Lally MA, et al. HIV testing among young adults and older adolescents in the setting of acute substance abuse treatment. *J Acquir Immune Defic Syndr.* 2001;27:135–142.
18. Greenwald JL, Pincus J. Rapid HIV testing in the era of OraQuick®. *Todays Ther Trends.* 2003;21:307–344.
19. Greenwald JL, Burstein GR, Pincus J, Branson B. A rapid review of rapid HIV antibody tests. *Curr Inf Dis Repts.* 2006; 8:125–131.
20. Sanders GD, Bayoumi AM, Sundaram V, et al. Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. *New Eng J Med.* 2005;352:570–585.
21. Paltiel AD, Weinstein MC, Kimmel AD, et al. Expanded screening for HIV in the United States—an analysis of cost effectiveness. *New Eng J Med.* 2005;352:586–595.
22. Harvard Center for Risk Analysis: The CEA Registry. Cost-utility analyses published from 1976 to 2001, with ratios converted to 2002 US dollars. Available at: http://www.hsph.harvard.edu/cearegistry/data/1976-2001_CEratios_comprehensive_4-7-2004.pdf. Accessed August 15, 2005.
23. Lubelchek R, Kroc K, Hota B, et al. The role of rapid vs conventional human immunodeficiency virus testing for inpatients: effects on quality of care. *Arch Intern Med.* 2005; 165:1956 The role of rapid vs. conventional Human Immunodeficiency Virus testing for inpatients 1960.
24. CDC. Revised guidelines for HIV counseling, testing, and referral. *MMWR Recomm Rep.* 2001;50(RR19);1–58.
25. Health Resources and Services Administration. Stigma and HIV/AIDS: a review of the literature. Available at: http://hab.hrsa.gov/publications/stigma/introduction.htm. Accessed August 15, 2005.
26. Roy CL, Poon EG, Karson AS, et al. Patient safety concerns arising from test results that return after hospital discharge. *Ann Intern Med.* 2005;143:121–128.
27. Orasure Technologies, Inc. Bethlehem, Pa. OraQuick Advance rapid HIV 1/2 rapid antibody test [package insert]. Available at: http://www.orasure.com/uploaded/398.pdf?138&sec=2&subsec=2. Accessed August 16, 2005.
28. Trinity Biotech. Bray, County Wicklow, Ireland. Uni-Gold Recombigen HIV [package insert]. Available at: http://www.trinitybiotech.com/EN/1206506-09-04.pdf. Accessed August 16, 2005.
29. Centers for Medicare and Medicaid Services. CLIA Program Clinical Laboratory Improvement Amendments. Available at: http://www.cms.hhs.gov/clia/. Accessed August 16, 2005.
30. CDC. Quality assurance guidelines for testing using the OraQuick Rapid HIV-1 antibody test. Available at: http://www.cdc.gov/hiv/rapid_testing/materials/QA-Guide.htm. Accessed August 16, 2005.
31. MedMira Laboratories, Inc. Halifax, Nova Scotia, Canada. Reveal Rapid HIV-1 Antibody Test [package insert]. Available at: http://www.reveal-hiv.com/pdf/PI.pdf. Accessed August 16, 2005.
32. Bio-Rad Laboratories, Inc. Redmond, Wa. Multispot HIV-1/HIV-2 Rapid Test. Available at: http://www.bio-rad.com/B2B/BioRad/product/br_category.jsp?BV_ SessionID=@@@@1896810173. 1124136663@@@@&BV_ EngineID=cccdaddfhedflidcfngcfkmdhkkdfll.0& categoryPath=Catalogs%2fClinical+Diagnostics% 2fBlood+Virus%2fBlood+Virus+(US+Market)% 2fReagents%2fRetrovirus+HIV% 2fRapid+Tests% 2fMultispot+HIV-1%7cHIV-2+Rapid+Test&divName=Clinical+Diagnostics& loggedIn=false&lang=English&country=US&catOID=-30657&isPA=false&catLevel =8&serviceLevel=Lit+Request. Accessed August 16, 2005.
33. How hospitalists add value. *The Hospitalist.* 2005;9(suppl 1).
34. Siegel K, Lekas HM. AIDS as a chronic illness: psychosocial implications. *AIDS.* 2002;16(suppl 4):S69–S76.
35. Chou R, Smits AK, Huffman LH, Fu R, Korthuis PT. Prenatal screening for HIV: a review of the evidence for the U.S. Preventive Services Taskforce. *Ann Intern Med* 2005;143:38–54.
36. CDC. Revised recommendations for HIV screening of pregnant women. *MMWR Recomm Rep.* 2001;50(RR19):59–86.
37. Manavi K, Welsby PD. HIV testing should no longer be accorded any special status. *BMJ.* 2005;330:492–493.
38. The EXPLORE Study Team. Effects of a behavioral intervention to reduce acquisition of HIV infection among men who have sex with men: the EXPLORE randomized controlled study. *Lancet.* 2004;364:41–50.
39. Freedberg KA, Samet JH. Think HIV. Why physicians should lower their threshold for HIV testing. *Arch Intern Med.* 1999; 159:1994–2000.

A Rapid Review of Rapid HIV Antibody Tests

Jeffrey L. Greenwald, MD, Gale R. Burstein, MD, MPH, FAAP, Jonathan Pincus, MD, and Bernard Branson, MD

Corresponding author
Jeffrey L Greenwald, MD
Boston Medical Center, Hospital Medicine Unit,
850 Harrison Avenue, NIF 6W, Boston, MA 02118, USA.
E-mail: Jeffrey.Greenwald@bmc.org

Current Infectious Disease Reports 2006, 8:125–131
Current Science Inc. ISSN 1523-3847

Rapid HIV antibody tests recently approved by the Food and Drug Administration can help reduce unrecognized infections by improving access to testing in both clinical and nonclinical settings and increase the proportion of those tested who learn their results. Four rapid HIV antibody tests are now available in the United States; two are approved for use at point-of-care sites outside a traditional laboratory. All four tests are interpreted visually. Sites offering rapid HIV testing must periodically run external controls (known HIV-positive and HIV-negative specimens) and provide persons who undergo rapid testing a subject information sheet. This paper reviews the operating and performance characteristics, quality assurance and laboratory requirements, and HIV counseling implications of the currently available rapid HIV tests.

Introduction

Despite ongoing prevention and education efforts, an estimated 40,000 new HIV infections have occurred annually in the United States since the early 1990s. Of the estimated 1,039,000 to 1,185,000 persons living with HIV, approximately 252,000 to 312,000 (25%) persons are unaware they are infected [1]. Available evidence suggests that many new infections are caused by persons unaware of their HIV infection [2,3].

HIV Testing

Many persons with HIV do not get tested until late in their infection. Approximately 40% to 50% of patients with HIV infection are diagnosed with AIDS within 1 year of first testing HIV-positive [2,4–6].

Many persons who are tested do not return to learn their test results. The National Health Interview Survey found that 12.5% of persons tested in 1994 and 13.3% in 1995 did not receive their results [7], and the Centers for Disease Control and Prevention (CDC) estimates that in 2000, 31% of patients who tested HIV-positive at public-sector testing sites did not return to receive their results [8].

To reduce barriers to early diagnosis of HIV infection and increase access to treatment and prevention services, the CDC announced a new initiative, "Advancing HIV Prevention: New Strategies for a Changing Epidemic" (AHP) [8]. This multifaceted program stresses the importance of routinely offering HIV testing as part of the medical visit and expands on the 1993 recommendations for testing inpatients and outpatients in acute-care hospital settings [9]. Additionally, AHP stresses the importance of using rapid HIV tests to facilitate access to early diagnosis in high prevalence areas, for high-risk individuals, and for women during labor and delivery who have not previously been tested and in nontraditional testing settings.

Rapid HIV tests can play an important role in HIV prevention activities and expand access to testing in both clinical and nonclinical settings. They can help overcome some of the barriers to early diagnosis and improve linkage to care of infected persons. This paper will review the operating and performance characteristics, quality assurance (QA) and laboratory requirements for currently available rapid HIV tests, and counseling implications.

The Tests

Four rapid HIV tests have been approved by the US Food and Drug Administration (FDA): OraQuick® (and its newer version OraQuick® Advance) Rapid HIV-1/2 Antibody Test (OraSure Technologies, Inc., Bethlehem, PA); Reveal™ (and its newer version Reveal™ G2) Rapid HIV-1 Antibody Test (MedMira, Halifax, Nova Scotia); Uni-Gold Recombigen® HIV Test (Trinity BioTech, Bray, Ireland); and Multispot HIV-1/HIV-2 Rapid Test (Bio-Rad Laboratories, Redmond, WA). Like conventional HIV enzyme immunoassays (EIAs), rapid HIV tests are screening tests that require confirmation if reactive. Though each of these rapid HIV tests has unique characteristics, they share many common features, including how the tests work, the use of external controls, and other requirements such as the product information sheets that are provided to patients.

Table 1. US Food and Drug Administration–approved rapid HIV antibody tests for HIV-1 detection

Rapid HIV test*	Specimen type	Sensitivity†	Specificity†	CLIA category
OraQuick® Advance Rapid HIV-1/2 Antibody test	Oral fluid	99.3% (98.4–99.7)	99.8% (99.6–99.9)	Waived
	Whole blood (fingerstick or venipucture)	99.6% (98.5–99.9)	100% (99.7–100)	Waived
	Plasma	99.6% (98.9–99.8)	99.9% (99.6–99.9)	Moderate complexity
Reveal™ G-2 Rapid HIV-1 Antibody test	Serum	99.8% (99.5–100)	99.1% (98.8–99.4)	Moderate complexity
	Plasma	99.8% (99.5–100)	98.6% (98.4–98.8)	Moderate complexity
Uni-Gold Recombigen® HIV test	Whole blood (fingerstick or venipucture)	100% (99.5–100)	99.7% (99.0–100)	Waived
	Serum and plasma	100% (99.5–100)	99.8% (99.3–100)	Moderate complexity
Multispot HIV-1/HIV-2 Rapid test	Serum	100% (99.94–100)	99.93% (99.79–100)	Moderate complexity
	Plasma	100% (99.94–100)	99.91% (99.77–100)	Moderate complexity

*Trade names are for identification purposes only and do not imply endorsement by the US Department of Health and Human Services or the Centers for Disease Control and Prevention.
†95% CI
CLIA—the Clinical Laboratory Improvement Amendments of 1988.
Modified from Health Research and Education Trust available at http://www.hret.org/hret/programs/hivtransmsp.html.

All four tests are interpreted visually and require no instrumentation. HIV antigens are affixed to the test strip or membrane. If HIV antibodies are present in the specimen being tested, they bind to the affixed antigen. The test kit's colorimetric reagent binds to these immunoglobulins creating an indicator that is visually detectable.

External controls

All four rapid HIV tests require the periodic use of external controls (known HIV-positive and -negative specimens). External controls must be run 1) by each new operator prior to performing the test on patients, 2) when a new lot of test kits is used, 3) upon receipt of a new shipment of test kits, 4) when the temperature of the storage or testing area falls outside the recommended range, and 5) at periodic intervals determined by the testing facility, usually based on their volume of testing.

Subject information sheets

The FDA requires that persons who undergo rapid testing receive a subject information sheet. This sheet, provided by each manufacturer with its rapid HIV test kits, includes basic information about HIV/AIDS, HIV testing, how the test works, what the tests results mean, and specifies that reactive rapid test results need to be confirmed.

The Clinical Laboratory Improvement Amendments of 1988

All laboratory testing is regulated under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), which classifies tests according to their complexity. To receive a CLIA waiver, tests must use direct, unprocessed specimens (such as whole blood or oral fluid) and be easy to perform with a negligible chance of error. Waived tests can be performed by persons without formal laboratory training outside traditional laboratories. Waived tests, suitable for use at the point-of-care, make it easier for nonclinical testing sites to offer rapid HIV tests. In order to purchase CLIA-waived rapid HIV tests, a facility must register as a laboratory with the CLIA program and adhere to the manufacturer's instructions for performing the tests.

OraQuick® and Uni-Gold are CLIA-waived; Reveal™ and Multispot are categorized as moderate complexity (Table 1). Laboratories that perform moderate complexity testing must meet more stringent standards for personnel, supervision, quality assurance, and proficiency testing than laboratories that perform waived testing.

OraQuick® Advance Rapid HIV-1/2 Antibody Test

On November 7, 2002, the FDA approved the OraQuick® Rapid HIV-1 Antibody Test for use on fingerstick blood samples. It received its CLIA waiver in January 2003. Subsequently, OraQuick® received approval for use with venipuncture whole blood and plasma (though OraQuick® used with plasma is classified as moderate complexity under CLIA). In 2004, OraQuick® Advance received FDA approval for use with oral fluid and for detection of both HIV-1 and HIV-2.

The OraQuick® test device is shown in Figure 1. The paddle-shaped device contains a nitrocellulose strip, upon which a stripe of synthetic gp41 peptides represent-

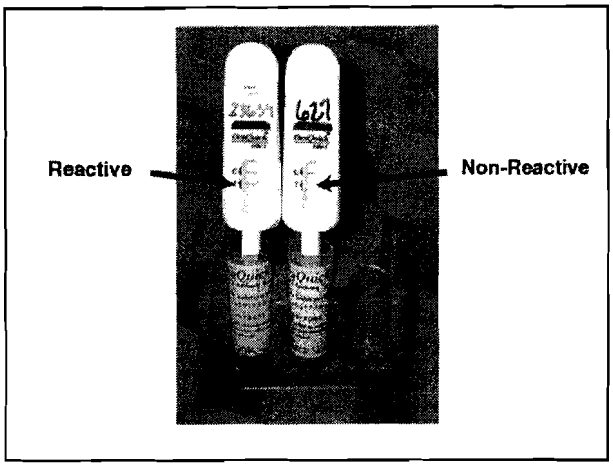


Figure 1. OraQuick® Advance Rapid HIV Antibody test.

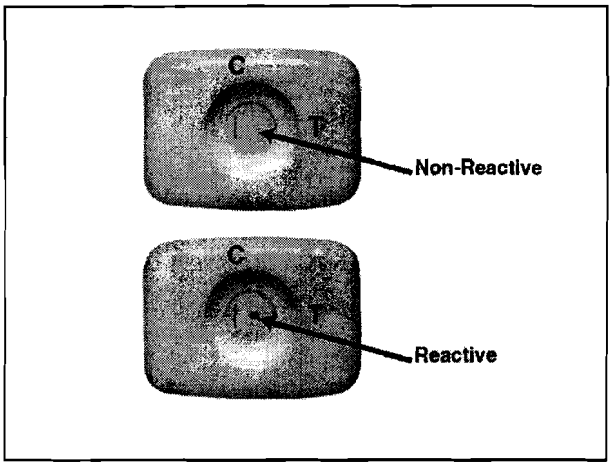


Figure 2. Reveal™ G2 Rapid HIV-1 Antibody test.

ing the HIV-1 envelope and the gp36 region of the HIV-2 envelope have been applied in the "T" (test) location, and a stripe of goat antihuman IgG in the "C" (control) location. The specimen of blood or plasma is added directly to the developer vial. For oral fluid testing, the oral fluid sample is collected by swabbing the gums with the paddle-shaped device. The test device is then added to the developer vial. If HIV antibodies are present in the specimen, they bind to the peptides causing a red line to appear in the test location. As the solution migrates further, it encounters the antihuman IgG control, and if an adequate specimen was added, a red line appears in the control location.

The test result should be read no sooner than 20 minutes and no later than 40 minutes after the test device is inserted into the developer vial. A red line at both the test and control location indicates a valid reactive test result; a red line only in the control location indicates a valid negative test result. The test is invalid and should be repeated with a new device if no line appears at the control location or if lines appear outside the areas indicated by the triangles [10].

Designed as a point-of-care HIV test, OraQuick® has been used in numerous settings including labor and delivery [11•], ambulatory clinical sites [12], emergency departments [13,14], hospital inpatient services [15] (Greenwald JL, unpublished data), correctional facilities [16], and for occupational exposures [17–19]. Additionally, OraQuick® has also been used by the military in battlefield operations [20].

Reveal™ G2 Rapid HIV-1 Antibody Test

On April 17, 2003, the FDA approved the Reveal™ Rapid HIV-1 Antibody Test to detect HIV antibodies in serum or plasma. In June 2004, it was superseded by the second generation Reveal™ G2 test, which incorporates an internal control [21]. Reveal™ G2 consists of a test cartridge and a proprietary colorimetric detection agent. Positive and negative external controls, which must also be reconstituted, are supplied with the kit.

Reveal™ is considered reactive if both the red control line and central red test dot appear, negative if only the control line appears, and invalid if the control line does not appear (Fig. 2). The Reveal™ G2 only takes 3 minutes to run [22]. However because it requires serum or plasma from centrifuged blood samples and several reagent steps, it is classified as a moderate complexity test under CLIA and is usually performed in a clinical laboratory.

Uni-Gold Recombigen® HIV Test

The Uni-Gold Recombigen® HIV Test received FDA approval in December 2003 for testing whole blood, serum, and plasma for antibodies to HIV-1. It was waived under CLIA in 2004 for use with venipuncture and fingerstick whole blood specimens [23]. The device consists of a rectangular plastic test cartridge and a dropper bottle of buffer solution (Fig. 3). Peptides from the immunodominant region of the HIV-1 envelope are immobilized on a nitrocellulose strip in the test region. Reagents are also bound at the control region to indicate whether the test is functioning correctly, but these do not detect IgG and thus appearance of the control line does not validate that adequate patient specimen has been added. One drop of specimen is added to the specimen well on the test cartridge followed by four drops of wash buffer. The specimen combines with the colorimetric reagent and migrates along the nitrocellulose strip past the test and control regions. The test is read 10 to 12 minutes after specimen is added. A line in both the test and control regions indicates a reactive test; a line in only the control region indicates a negative test. When used with whole blood, the test is valid only if the control line is present and the sample well is red, indicating that an adequate blood sample has been added [24].

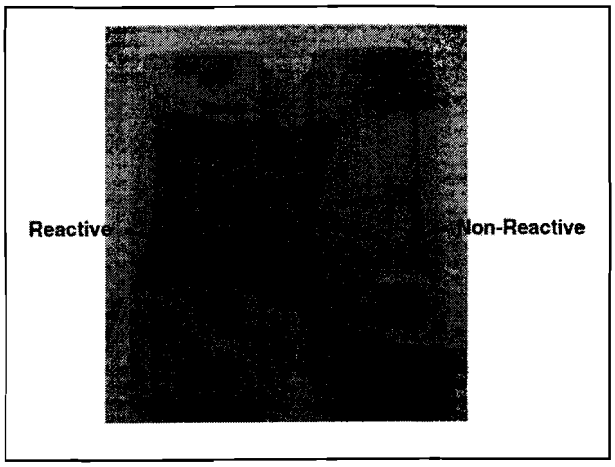


Figure 3. Uni-Gold Recombigen® HIV test.

Multispot HIV-1/HIV-2 Rapid Test

The Multispot HIV-1/HIV-2 Rapid Test received FDA approval in November, 2004 [25]. Multispot is classified as a moderate complexity under CLIA, approved for use on fresh or frozen serum and plasma to both detect and distinguish HIV-1 from HIV-2.

Multispot consists of a test cartridge and five reagents: specimen diluent, wash solution, conjugate, development reagent, and stop solution. The cartridge contains a membrane on which microparticles have been immobilized in four spots. Two of the spots consist of recombinant and synthetic gp41 peptides to detect HIV-1 antibodies; one consists of synthetic gp36 peptides to detect antibodies to HIV-2; and the fourth spot consists of goat antihuman IgG as the internal control.

The test is considered positive for HIV-1 if the control spot and either or both of the HIV-1 spots turn purple, and positive for HIV-2 if the control and HIV-2 spots appear (Fig. 4). If purple appears in the control spot, the HIV-2 spot, and one or both of the HIV-1 spots, the test is considered HIV reactive (undifferentiated). In this case, the specimen may be tested by additional methods which allow differentiation between HIV-1 and HIV-2. The test is negative when only the control spot appears. The absence of the control spot indicates an invalid result, regardless of any other spot pattern.

Rapid HIV Antibody Test Performance and Interpretation of Test Results

Like conventional EIAs, rapid HIV tests are screening tests. If performed correctly, they detect HIV antibodies with sensitivities similar to currently available EIAs [10,22,24–29] (Table 1). A negative rapid HIV test result requires no further confirmatory testing. False negative results, though rare, may occur in a person who has been acutely infected but who has not yet developed HIV antibodies. Therefore, any patient testing negative who has had known or suspected exposure to HIV within 3 months should be instructed to retest 3 months after the exposure date [30]. Additionally, false-negative rapid HIV test results have been observed in some patients receiving highly active antiretroviral therapy with undetectable virus in whom levels of HIV antibody have waned below the level of detection by the rapid HIV test [31].

A reactive result from any of the four rapid HIV tests is interpreted as a "preliminary positive" and requires confirmation by a more specific assay, typically a Western Blot (WB) or immunofluorescent assay (IFA) [10,22,24,25]. Performing a standard EIA screening prior to confirmatory testing is not required. However, if an EIA is performed, the specimen must still proceed to WB or IFA testing regardless of the EIA result. A positive WB or IFA confirms the diagnosis of HIV infection. If the confirmatory test yields negative or indeterminate results, follow-up HIV testing should be performed on a blood specimen collected 4 weeks after the initial reactive rapid HIV test result [32•] as some patients newly infected with HIV may not have developed antibody levels sufficient to produce a positive WB or IFA [33].

Table 1 presents the test performances of US FDA-approved rapid HIV tests. It is important to note that because the test specificities are less than 100%, false positive rapid test results are an expected but rare event. When testing low seroprevalence areas, a higher proportion of reactive tests will be false positives because there are few true positives in low-prevalence populations. The causes of falsely positive rapid HIV tests (ie, a reactive rapid HIV test with a negative or indeterminate confirmatory test) are poorly understood. Certain medical conditions may be associated with a slightly increased risk for false-positive OraQuick® rapid HIV tests (eg, hepatitis A and B viruses, Epstein-Barr virus, multiparity, and the serologic presence of rheumatoid factor) [10].

Quality Assurance for CLIA-waived Rapid HIV Antibody Tests

Although CLIA-waived rapid HIV test devices are easy to use and can provide reliable results when the manufacturer's directions are followed, mistakes can occur at any point in the testing process, including storage and testing area temperature, test kit shelf-life, specimen collection, test performance and results interpretation, referring specimens for confirmatory testing, managing confirmatory test results, etc. To reduce mistakes and to ensure that the FDA restrictions for sale of the test are followed, a site that performs rapid HIV tests must have a QA program in place before offering these tests. In January 2003, the CDC convened a panel of experts including laboratory scientists and individuals from the FDA and the Centers for Medicare and Medicaid Services to develop guidelines that outline the basic parts of a rapid HIV test QA program [32•]. The *Quality*

Assurance Guidelines for Testing Using the OraQuick® Rapid HIV-1 Antibody Test are intended to assist a range of providers in developing policies, processes and procedures to ensure high quality HIV testing services. These guidelines include 1) the basics of a QA program for testing using OraQuick®, 2) an overview of government rules that apply to using this test, and 3) examples of forms/checklists that can be used to keep track of QA outcomes.

Counseling with Rapid HIV Antibody Tests

Counseling for patients choosing rapid HIV testing involves some differences compared with conventional testing, including assessing preparedness for clients to receive test results in the same session and explaining the meaning of preliminary positive results. Information can be provided either face-to-face or in a pamphlet, brochure, or video [34].

Patients with reactive rapid test results must be counseled in simple terms about the meaning of a reactive test. The provider must emphasize the need for a confirmatory test and schedule a return visit for results. Providers offering rapid HIV testing should be able to collect blood or oral fluid specimens on-site for confirmatory testing. All patients with reactive tests should be counseled on risk-reduction behaviors while awaiting the results of confirmatory testing. A simple message to convey this information could be "Your preliminary test result is positive, but we won't know for sure if you are infected with HIV until we get the results from your confirmatory test. In the meantime, you should take precautions to avoid transmitting the virus" [34]. The New York State Department of Health AIDS Institute has also created guidelines for how to discuss reactive results stratifying the language based on the patient's level of risk for HIV infection. For clients at high risk, the guidelines suggest saying "Based on your risk factors, it is *highly likely* that the preliminary test result is correct and that you have HIV" (emphasis added). For those at low risk, the phrase "quite likely" is recommended, and for those with no admitted risk factors, they advise informing them "There is a chance that this result could be a false positive" [35].

Physicians and counseling staff may be apprehensive about rapid testing specifically with regards to the ability to handle preliminary positive test results at any time. Data from RESPECT-2, a large, randomized, controlled trial that compared different forms of HIV testing and risk-reduction counseling in clients at sexually transmitted disease (STD) clinics in the United States, found that after gaining experience in the field, the majority of counselors preferred rapid testing, felt that rapid HIV testing sessions resulted in enhanced counseling, and felt that it was more convenient for both clients and counseling staff [36]. Although some have expressed concern about how counselors and clients will deal with discussing and understanding reactive results [37], others have noted that providers have extensive experience managing preliminary positive test results (eg, abnormal mammograms that require biopsies and abnormal pap smears that require colposcopy) [38] and studies of rapid testing have demonstrated good client understanding of results [39].

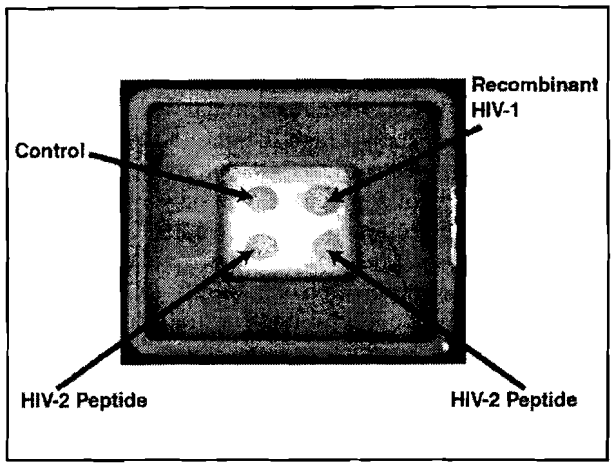


Figure 4. Multispot HIV-1/HIV-2 Rapid test.

Providing HIV counseling and testing may be challenging in some health care settings. Because the average primary care office visit in the United States is less than 18 minutes long [40], even the "brief" counseling protocol of RESPECT-2 could take up an entire office visit. In these situations, alternative procedures for HIV counseling with rapid testing should be considered, eg, providing information either in a face-to-face meeting with a counselor or in a pamphlet, brochure, or video [34].

Outcomes of Rapid HIV Testing—Receipt of Test Results

Compared with the standard two-session counseling and testing protocol, single-session, rapid HIV testing has the potential advantages of decreasing costs and increasing the number of patients who receive their results [41]. In anonymous testing and STD clinics in Dallas, the use of rapid testing with the Single Use Diagnostic System (SUDS) HIV-1 test (Murex, Norcross, GA) was associated with an increase in the number of patients learning their serostatus, lower costs, and improved patient satisfaction [39]. A randomized, controlled trial at a needle exchange and two bathhouses compared SUDS HIV-1 testing to other conventional HIV testing. This study found that more clients received their test results after rapid testing than with traditional testing: at the needle exchange, 66 (83%) of 80 versus 27 (56%) of 48 (odds ratio [OR] = 3.7; $P = 0.002$), and at the bathhouses, 102 (99%) of 103 versus 82 (74%) of 111 (OR = 36.1; $P < 0.001$) [38].

Patients failing to return for their confirmatory HIV test results remain a challenge [42]. Patients who do not return for confirmatory test results may choose to seek care at other

locations, already know their status, or seek retesting elsewhere. However, with rapid HIV testing, patients with reactive test results leave the initial testing visit with information that there is a high likelihood that they are seropositive compared with receiving no test result information at the end of a visit where a conventional HIV test specimen was collected. Because rapid HIV testing is likely to increase in the coming years, validation of an algorithm using a combination of point-of-care rapid HIV tests would enhance opportunities for individuals to get a confirmed HIV status.

Patient Satisfaction

Overwhelmingly, both patients and providers prefer rapid HIV tests to conventional EIAs [36,43–45]. Ninety percent (1038/1148) of persons seeking HIV testing at 24 clinical and nonclinical settings that offered the OraQuick® HIV test and an oral or serum EIA in New York, Utah, and Wisconsin preferred the rapid test; 13% of the clients in New York and Utah said they would not have tested that day if the rapid test had not been available [43].

Financial Considerations

The price for the FDA-approved rapid HIV test kits, as of July 2005, range from $14 to $25. Costs for multidose external control vials range from $20 to $26.25 [29]. According to the Centers for Medicare and Medicaid Services 2005 Clinical Laboratory Fee Schedule, average reimbursement for a CLIA-waived rapid HIV-1 antibody test (Current Procedural Terminology [CPT] code 86701QW) is $12.41/test and for a CLIA-waived rapid HIV-1/2 antibody test (CPT code 86703QW) is $19.17 [46–48]. Providers offering point-of-care, rapid HIV testing may be challenged by reimbursement not keeping pace with the list prices of the tests. In addition, comparable with counseling for other health issues, HIV counseling by a nonphysician is not reimbursable. Physicians performing HIV counseling may attempt to collect reimbursement for it by billing for prolonged services.

Conclusions

Rapid testing overcomes major barriers to individuals with HIV infection knowing their status: 1) HIV testing opportunities can be expanded to both medical and nonmedical settings and 2) rapid testing facilitates patients receiving their test results the same day, usually at the encounter where the test specimen was collected. Providing greater access to testing, prevention, and care services for persons living with HIV can reduce the number of new infections and lead to reductions in HIV-associated morbidity and mortality [49,50].

Disclaimer

Use of trade names and commercial sources is for identification only and does not imply endorsement by the US Department of Health and Human Services.

References to non-CDC sites on the Internet are provided as a service to this manuscript's readers and do not constitute or imply endorsement of these organizations or their programs by CDC or the US Department of Health and Human Services. CDC is not responsible for the content of these sites. URL addresses listed in this manuscript were current as of the date of publication.

References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

- • Of importance
- •• Of major importance

1. Glynn M, Rhodes P: **Estimated HIV Prevalence in the United States at the end of 2003 [abstract T1-B1101].** *Paper presented at the 2005 National HIV Prevention Conference. Atlanta, GA; June 12–15, 2005.*
2. Samet JH, Freedberg KA, Savetsky JB, et al.: **Understanding delay to medical care for HIV infection: the long-term non-presenter.** *AIDS* 2001, 15:77–85.
3. Valleroy LA, MacKellar DA, Karon JM, et al.: **HIV prevalence and associated risks in young men who have sex with men. Young Men's Survey Study Group.** *JAMA* 2000, 284:198–204.
4. Centers for Disease Control and Prevention: **HIV/AIDS Surveillance Report 2002; 14**:Table 1. http://www.cdc.gov/hiv/stats/hasrlink.htm. Accessed November 2, 2004.
5. Klein D, Hurley LB, Merrill D, Quesenberry CP Jr: **Review of medical encounters in the 5 years before a diagnosis of HIV-1 infection: Implications for early detection.** *J Acquir Immune Defic Syndr* 2003, 32:143–152.
6. Centers for Disease Control and Prevention: **Late versus early testing of HIV–6 Sites, United States, 2000–2003.** *MMWR* 2003, 52:581–586.
7. Tao G, Branson BM, Kassler WJ, Cohen RA: **Rates of receiving HIV test results: data from the U.S. National Health Interview Survey for 1994 and 1995.** *J Acquir Immune Defic Syndr* 1999, 22:395–400.
8. Centers for Disease Control and Prevention: **Advancing HIV Prevention: new strategies for a changing epidemic—United States, 2003.** *MMWR* 2003, 52:329–332.
9. Centers for Disease Control and Prevention: **Recommendations for HIV testing services for inpatients and outpatients in acute-care hospital settings.** *MMWR* 1993, 42:RR-02.
10. **OraQuick Advance Rapid HIV-1/2 Antibody Test [package insert].** Bethlehem, PA: OraSure Technologies, Inc.; 2004. Available at http//:www.orasure.com.
11. • Bulterys M, Jamieson DJ, O'Sullivan MJ, et al.: **Rapid HIV-1 testing during labor: a multicenter study.** *JAMA* 2004, 292:219–223.

The authors demonstrate that rapid HIV testing is feasible and delivers accurate and timely test results for women in hospital-based labor and delivery wards.

12. Greenwald JL, Pincus J: **Rapid HIV testing in the era of OraQuick®.** *Today's Therapeutic Trends* 2003, 21:307–344.
13. Kroc KA, GR Papiez GR, Parks DI, et al.: **Routine screening of emergency department (ED) patients using OraQuick Rapid HIV-1 Antibody Test [abstract #M3-L103].** *Paper presented at the 2003 National HIV Prevention Conference. Atlanta, GA; July 27–30, 2003.*

14. Kelen GD, Shahan JB, Quinn TC: **Emergency department-based HIV screening and counseling: experience with rapid and standard serologic testing.** *Ann Emerg Med* 1999, 33:147–155.
15. Walensky RP, Losina E, Steger-Craven KA, Freedberg KA: **Identifying undiagnosed human immunodeficiency virus: The yield of routine, voluntary inpatient testing.** *Arch Intern Med* 2002, **162**:887–892.
16. MacGowan R, McKeever J, Stodola J, et al.: **Multi-site demonstration project implementing rapid HIV testing jails [abstract #WO-C1401].** *Paper presented at the 2005 National HIV Prevention Conference. Atlanta, GA: June 12–15, 2005.*
17. Salgado CD, Flanagan HL, Haverstick DM, Farr BM: **Low rate of false-positive results with use of a rapid HIV test.** *Infect Control Hosp Epidemiol* 2002, **23**:335–337.
18. Kallenborn JC, Price TG, Carrico R, Davidson AB: **Emergency department management of occupational exposures: cost analysis of rapid HIV test.** *Infect Control Hosp Epidemiol* 2001, **22**:289–293.
19. King AM, Osterwalder JJ, Vernazza PL: **A randomised prospective study to evaluate a rapid HIV-antibody assay in the management of cases of percutaneous exposure amongst health care workers.** *Swiss Medical Weekly* 2001, **131**:10–13.
20. **Military sealift command anthrax/smallpox meeting with industry.** Monday, February 24, 2003. **Washington Navy Yard, Bldg 220.** http://www.msc.navy.mil/n10/022403meeting.htm. Accessed September 3, 2004.
21. United States Food and Drug Administration: **Premarket approval information–devices.** http://www.fda.gov/cber/pmaltr/P00002315L.htm. Accessed June 2, 2005.
22. **Reveal Rapid HIV-1 Antibody Test [package insert].** Halifax, Nova Scotia, Canada: MedMira Laboratories, Inc.; 2004. Available at http//:www.reveal-hiv.com.
23. Trinity Biotech: **Trinity Biotech receives CLIA waiver for Uni-Gold Recombigen® HIV test [press release].** June 30, 2004. http//:www.trinitybiotech.com/data/press/30-06-04.pdf. Accessed September 3, 2004.
24. Trinity Biotech: **Uni-Gold Recombigen® HIV. Summary of safety and effectiveness.** http://www.fda.gov/cber/pmasumm/P0300250S.pdf. Accessed September 3, 2004.
25. **Multispot HIV-1/HIV-2 Rapid Test [package insert].** Redmond, WA: Bio-Rad Laboratories; 2004. Available at http//:www.bio-rad.com.
26. **Genetic Systems™ HIV-1/HIV-2 Peptide EIA [package insert].** Redmond, WA: Bio-Rad Laboratories Blood Virus Division; 2000.
27. **Genetic Systems™ rLAV EIA [package insert].** Redmond, WA: Bio-Rad Laboratories Blood Virus Division; 2000.
28. **Vironostica HIV-1 Microelisa System [package insert].** Durham, NC: Organon Teknika Corporation; 1994.
29. **Health Research and Education Trust (HRET): Charts for comparing rapid HIV antibody screening tests.** http://www.hret.org/hret/programs/hivtransmrpd.html. Accessed June 28, 2005.
30. Centers for Disease Control and Prevention: **Revised guidelines for HIV counseling, testing, and referral.** *MMWR* 2001, **50(RR-19)**:1–57.
31. O'Connell RJ, Merritt TM, Malia JA, et al.: **Performance of the OraQuick rapid antibody test for diagnosis of human immunodeficiency virus type 1 infection in patients with various levels of exposure to highly active antiretroviral therapy.** *J Clin Microbiol* 2003, **41**:2153–2155.
32.• Centers for Disease Control and Prevention: **Quality assurance guidelines for testing using the OraQuick® Rapid HIV-1 antibody test.** http://www.cdc.gov/hiv/rapid_testing/materials/QA-Guide.htm. Accessed October 6, 2004.
This document provides guidance on quality assurance practices for sites using or planning to use the OraQuick® Rapid HIV-1 Antibody test to detect antibodies to HIV.
33. Centers for Disease Control and Prevention: **Notice to readers: protocols for confirmation of reactive rapid HIV tests.** *MMWR* 2004, **53**:221–222.
34. Centers for Disease Control and Prevention: **HIV counseling with rapid tests.** http://www.cdc.gov/hiv/pubs/rt-counseling.htm. Accessed August 23, 2004.
35. New York State Department of Health: **Counseling protocol for rapid HIV testing sessions.** http://www.health.state.ny.us/diseases/aids/testing/rapid/protocol.htm. Accessed July 7, 2005.
36. Padilla SM, Dillon B, Iatesta M, et al.: **Counselor perceptions and impressions of HIV counseling using a rapid HIV test: RESPECT-2.** *Paper presented at the 2001 National HIV Prevention Conference. Atlanta, GA; August 12–15, 2001.*
37. Elliott R, Jürgens R: **Rapid HIV screening at the point of care: Legal and ethical questions.** Montréal: Canadian HIV/AIDS Legal Network; 2000. http://www.aidslaw.ca/Maincontent/issues/testing/finalreports/tofc.htm. Accessed July 14, 2005.
38. Spieldberg F, Branson BM, Goldbaum GM, et al.: **Choosing HIV counseling and testing strategies for outreach settings: a randomized trial.** *J Acquir Immune Defic Syndr* 2005, 38:348–355.
39. Kassler WJ, Dillon BA, Haley C, et al.: **On-site, rapid HIV testing with same-day results and counseling.** *AIDS* 1997, 11:1045–1051.
40. Woodwell DA, Cherry DK: **National ambulatory medical care survey: 2002 summary.** *Adv Data* 2004, 346:1–44.
41. Ekwueme DE, Pinkerton SD, Holtgrave DR, Branson BM: **Cost comparison of three HIV counseling and testing technologies.** *Am J Prev Med* 2003, 25:112–121.
42. Valdiserri RO, Moore M, Gerber AR, et al.: **A study of clients returning for counseling after HIV testing: implications for improving rates of return.** *Public Health Rep* 1993, **108**:12–18.
43. Burstein GR, Newman DR, San Antonio-Gaddy ML, et al.: **Client Preferences for Quick OraQuick Test Results in 3 U.S. States [poster].** *Presented at the XV International AIDS Conference. Bangkok, Thailand; July 11–16, 2004.*
44. Birkhead GS, San Antonio-Gaddy ML, Richardson-Moore AL, et al.: **Effects of training and field experience on staff confidence and skills for rapid HIV testing in New York State (NYS) [poster].** *Presented at the XV International AIDS Conference. Bangkok, Thailand; July 11–16, 2004.*
45. Metcalf CA, Dillon BA, Malotte CK, et al.: **Client preferences regarding HIV testing and counseling in STD clinic settings: RESPECT-2.** Available at http://www.cdc.gov/hiv/projects/respect-2/slide/testprefpaper.ppt. Accessed March 18, 2005.
46. American Medical Association: **CPT code/value search.** Available at https://catalog.ama-assn.org/Catalog/cpt/cpt_home.jsp. Accessed July 6, 2005.
47. Centers for Medicare and Medicaid Services: **2005 clinical laboratory fee schedule.** http://new.cms.hhs.gov/Clinical-LabFeeSched/CFRC/list.
48. Centers for Medicare and Medicaid Services: Tests granted waived status under CLIA. http://www.cms.hhs.gov/clia/waivetbl.pdf. Accessed July 11, 2005.
49. Janssen RS, Holtgrave DR, Valdiserri RO, et al.: The serostatus approach to fighting the HIV epidemic: Prevention strategies for infected individuals. Am J Pub Health 2001, 91:1019–1024.
50. Institute of Medicine: No time to lose: getting more from HIV prevention. Washington, DC: National Academy Press; 2001.

ORIGINAL ARTICLE

Evaluation of the Centers for Disease Control and Prevention's Recommendations Regarding Routine Testing for Human Immunodeficiency Virus by an Inpatient Service: Who Are We Missing?

JEFFREY L. GREENWALD, MD; CATHERINE A. RICH, MD; SAMANTHA BESSEGA, MD; MICHAEL A. POSNER, MS; JARED LANE MAEDA, MPH; AND PAUL R. SKOLNIK, MD

OBJECTIVE: **To assess the proportion of hospitalized patients who tested positive for human immunodeficiency virus (HIV) by a routine inpatient testing service, as recommended by the Centers for Disease Control and Prevention, who might not have been identified had routine testing not been offered.**

PATIENTS AND METHODS: **In this retrospective cohort study, the medical records of patients who tested HIV positive by the inpatient testing service between 1999 and 2003 were compared with the medical records of inpatients who tested HIV negative by the inpatient testing service and the medical records of patients who tested HIV positive in ambulatory settings. We compared HIV risk factors, discharge diagnoses, CD4 cell counts, and HIV RNA concentrations.**

RESULTS: **A total of 243 patients participated in this study: 81 patients who tested HIV positive and 81 who tested HIV negative by the inpatient testing service, and 81 patients who tested HIV positive in ambulatory settings. Both HIV-positive inpatients and HIV-positive outpatients had similar frequencies of HIV risk factors (46% vs 43%; *P*=.75). Both groups differed significantly from HIV-negative inpatients (4%; *P*<.001). Comparing HIV-positive inpatients with HIV-positive outpatients, CD4 cell counts were lower (196 vs 371 cells/mm^3; *P*<.001), and HIV RNA levels were higher (4.61 vs 4.09 $\log_{10}$ HIV RNA; *P*=.001). At diagnosis, 64 HIV-positive inpatients (79%) met criteria for acquired immunodeficiency syndrome compared with 21 HIV-positive outpatients (26%) (*P*<.001).**

CONCLUSION: **Patients who tested HIV positive through inpatient testing have more advanced disease than those identified as outpatients. Half of these patients would not have been identified had testing not been routinely offered. Routine inpatient HIV testing offers an important opportunity to identify patients with HIV infection.**

***Mayo Clin Proc.* 2006;81(4):452-458**

AIDS = acquired immunodeficiency syndrome; CDC = Centers for Disease Control and Prevention; HITS = HIV Inpatient Testing Service; HIV = human immunodeficiency virus; OP positive = outpatient positive

In 1993, the Centers for Disease Control and Prevention (CDC) published guidelines that advocated human immunodeficiency virus (HIV) counseling and testing in acute care hospitals with high seroprevalences of HIV (at least 1%) or high rates of acquired immunodeficiency syndrome (AIDS) diagnoses (at least 1 per 1000 discharges).[1] The initial recommendations and updated guidelines published in 2001[2] called for these hospitals to offer routine HIV counseling and testing to all inpatients. Since the CDC recommendations were published, only a limited number of studies have evaluated the efficacy of offering routine inpatient testing.[3-5] Little is known about the characteristics, risk factors, and stage of HIV disease of the inpatients identified by routine HIV testing programs or about the numbers of additional patients with undiagnosed HIV such a program would identify.

In 1999, Boston Medical Center, an urban, academic hospital in Boston, Mass, began offering confidential HIV counseling and testing to inpatients on the medical service on a referral basis. In 2001, the testing service expanded to screen all adult patients in the medical service. The purpose of this study is to report the experience of the HIV Inpatient Testing Service (HITS), examining its impact on identifying HIV-infected inpatients who might not otherwise have been detected had testing not been offered routinely and clarifying their immunologic, virologic, and risk factor profiles.

For editorial comment, see page 449

PATIENTS AND METHODS

The study group was composed of patients who tested HIV positive through HITS between November 1999 and August 2003. The only patients excluded were those for whom hospital medical records were unavailable. Before April 2001, all testing through the program was done on a physician referral basis. After April 2001, patients without referral were approached directly by counseling and testing staff

From the Hospital Medicine Unit, Department of Medicine (J.L.G.) and General Internal Medicine (C.A.R.), Boston Medical Center, Boston, Mass; Department of Medicine, Lowell General Hospital, Lowell, Mass (S.B.); Department of Mathematical Sciences, Villanova University, Villanova, Pa (M.A.P.); University of Illinois, School of Public Health, Chicago (J.L.M.); Center for HIV/AIDS Care and Research, Boston Medical Center and Boston University School of Medicine, Boston, Mass (P.R.S.).

Dr Skolnik was supported by Massachusetts Department of Public Health Contracts 49301802043, 49221802005, and 497144BC080. These grants funded the HIV testing program from which these data were obtained.

Individual reprints of this article are not available. Address correspondence to Jeffrey L. Greenwald, MD, Hospital Medicine Unit, Boston Medical Center, 850 Harrison Ave, NIF 6W, Boston, MA 02118 (e-mail: jeffrey.greenwald@bmc.org)

in an attempt to see as many newly admitted patients as time permitted. Patients who were HIV positive were referred to the HIV intake clinic at Boston Medical Center, although some patients received primary care elsewhere and chose to have follow-up visits at other institutions.

The first comparison group was composed of patients diagnosed as being HIV positive in an outpatient setting. This group was called the outpatient positive (OP positive) group and included patients whose conditions were diagnosed in Boston Medical Center clinics, community health centers, local correctional facilities, and surrounding medical facilities and who were referred to Boston Medical Center for HIV care. The OP positive medical records were identified by reviewing Boston Medical Center's HIV intake clinic schedule from the same day as the corresponding visit made by a patient who tested HIV positive through HITS (HITS positive group). The OP positive patient selected was the next patient on the schedule who had an HIV diagnosis made in an outpatient setting. If no OP positive patient was available on the same date, the next intake clinic schedule was reviewed for appropriate candidates. If a HITS-positive patient did not present for initial HIV care at our institution, an outpatient match was selected from HIV clinic records for the date of testing rather than the date of care.

The second comparison group included patients who tested HIV negative through HITS (HITS negative group). Using the HITS testing log, for every HITS positive subject, the medical record of the next HIV-seronegative inpatient was selected for the same testing date.

Study Design

After obtaining approval from the Boston Medical Center Institutional Review Board, records were selected and information was extracted from HITS records, the HIV intake clinic records, and the hospital's record system regarding inpatient and outpatient visits. The outpatient visits of interest included the first 3 HIV-related visits after HIV diagnosis. For HITS positive subjects who did not present to Boston Medical Center for initial HIV outpatient care, no outpatient follow-up information was available. We recorded age, sex, race, educational level, HIV risk factors, date of initiation of outpatient HIV care, inpatient diagnoses (for HITS positive and HITS negative groups), initial CD4 cell count, HIV RNA (viral load), and presence of AIDS at time of diagnosis. When CD4 and viral load were recorded twice, representing additional testing, the mean of the 2 values was used in analyses. When only 1 was present, it was used by itself. Inpatient diagnoses were taken from discharge summaries. The abstractors (C.A.R., S.B.), both physicians, were instructed to select up to 3 diagnoses that were the most active during the hospitalization. If HIV or AIDS was among the diagnoses on the discharge summary, then it was included.

To evaluate risk factors that were self-reported by the patients, we used the 2002 HIV/AIDS Surveillance Report to define high-risk categories for HIV infection.[6] We also examined the outpatient literature concerning identifiable clinical conditions that should prompt consideration of HIV testing. These conditions include oral candidal infection, pneumonia, unexplained fever, herpes zoster, seborrheic dermatitis, night sweats, unexplained weight loss, and lymphadenopathy.[7-9] We extrapolated these diagnoses to the inpatient setting, adding several diagnoses that we thought were generally recognized by physicians as inpatient prompts for HIV testing (eg, complications of injection drug use, sexually transmitted diseases, complicated bacterial infections, nephrotic syndrome). We then applied this list of diagnoses to the discharge diagnoses of the HITS positive patients and noted whether the diagnoses were related, possibly related, or unrelated to HIV disease. Related diagnoses included those that were AIDS defining and those that were likely to have been caused by HIV/AIDS. Possibly related diagnoses were those that might prompt a physician to consider HIV testing. Unrelated diagnoses were those that were not thought to be causally linked to HIV infection (Table 1 and Figure 1).

Statistical Analyses

A sample size of 80 per group was chosen as our enrollment goal to attain 80% power to detect a 100 cell difference in CD4 cell count between HIV-positive and OP positive groups. We used an estimate for a common SD of 221.7 from pilot data.

We examined bivariate associations within groups using analysis of variance tests for numerical variables and χ^2 tests of independence for categorical variables. When examining 2 samples, independent sample *t* tests were used, with equal variance *t* tests when the assumption was verified. Base 10 logarithmic transformations of CD4 cell count and HIV RNA were considered because of the nonnormality of the data. Ultimately, the nontransformed CD4 cell counts and the $\log_{10}$ HIV RNA values were used. Sensitivity comparisons using the nontransformed HIV RNA levels did not differ substantially from the transformed values. Multivariate linear regressions were used to control for covariates when quantifying the differences between groups. For applicable analyses, a significance level of .05 was used, and all tests were 2-sided. All statistical analyses were performed with SAS statistical software, version 8.2 (SAS Institute Inc, Cary, NC).

RESULTS

A total of 243 patients participated in this study. During the study period, HITS identified 89 patients as HIV positive.

TABLE 1. **Discharge Diagnoses of Patients Who Tested HIV Positive by the HIV Inpatient Testing Service***

HIV related	Possibly HIV related	Unrelated to HIV
Cerebral toxoplasmosis	Aseptic meningitis	Anaphylaxis
CMV colitis	Bacteremia	Breast cancer
CMV esophagitis	Disseminated lymphadenopathy	Celiac sprue
Cryptococcal meningitis	ESRD	Cellulitis
Esophageal candidiasis	Fever of unknown origin	Chest pain
Extrapulmonary TB	Nephrotic syndrome	CHF
HIV	Neurosyphilis	Chronic pancreatitis
HIV encephalopathy	Opiate withdrawal	Coagulation disorder
PCP	Paraspinal abscess	Crohn disease
Pott disease	Pelvic inflammatory disease	Dehydration
Pulmonary TB	Peripheral neuropathy	Diarrhea
Thrush	Pneumonia	Gastroenteritis
	Psoas abscess	Hepatic encephalopathy
	Pulmonary hypertension	Hypertension
	Septic arthritis	Idiopathic pulmonary fibrosis
	Thrombocytopenia	Malaria
		Medication toxicity
		Non-Hodgkin lymphoma
		Pyelonephritis
		Reactive arthritis
		Septic thrombophlebitis
		Soft tissue infection
		Stroke
		Suicide attempt
		Viral syndrome

*CHF = congestive heart failure; CMV = cytomegalovirus; ESRD = end-stage renal disease; HIV = human immunodeficiency virus; PCP = *Pneumocystis carinii* pneumonia; TB = tuberculosis.

Eighty-one (91%) of these 89 had medical records accessible to the study staff. These 81 medical records were included in the study as the HITS positive group. These patients were compared with 81 OP positive patients and 81 HITS negative patients.

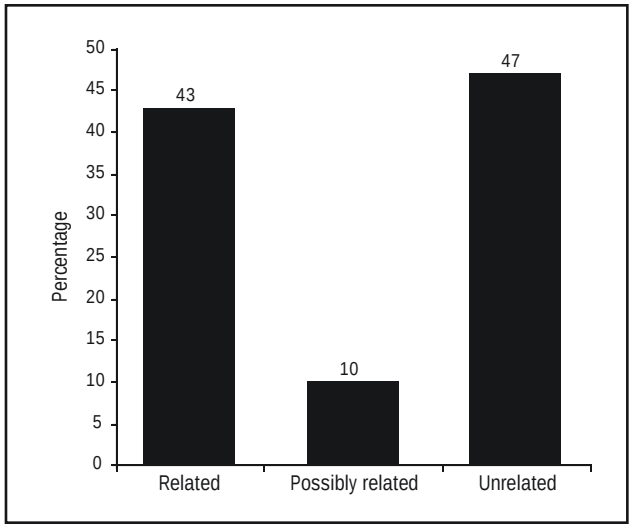


Figure 1. Percentage of the discharge diagnoses of patients who tested positive for human immunodeficiency virus (HIV) according to the HIV Inpatient Testing Service that were related, possibly related, or unrelated to HIV disease.

The demographics of the 243 study participants are presented in Table 2. No significant sex differences were apparent among the 3 groups (P=.61), but the mean age of the HITS positive group was 4.7 years higher than the OP positive group (P=.004). Differences were also apparent in racial categories and levels of education. Specifically, the HITS positive group was more likely to be Latino and less likely to be African than the OP positive group. Compared with the HITS negative group, the HITS positive group was more likely to be Haitian and less likely to have had a high school education.

Compared with the risk profiles of the OP positive patients, the HITS positive patients were more likely to be injection drug users (35% vs 6%; P<.001) but less likely to have reported sex with HIV-positive partners (9% vs 25%; P=.006). Despite their overall greater morbidity (ie, requiring hospitalization), the HITS positive group had a similar percentage of overall high-risk individuals as the OP positive group (46% vs 43%; P=.75). High-risk behaviors were more prevalent in the HITS positive and OP positive patients than in the HITS negative patients (overall P<.001).

The HITS positive and OP positive groups differed in CD4 cell count and HIV RNA. The HITS positive group had a lower mean CD4 cell count (196 vs 371 cells/mm^3) and a higher mean HIV RNA load (122,806 vs 50,302 copies/mL; 4.61 vs 4.09 $\log_{10}$ HIV RNA) than the OP

TABLE 2. **Demographics of the 243 Study Patients by Group***

Demographics	HITS positive (n=81)	OP positive (n=81)	HITS negative (n=81)	*P* value
Mean age (y)	42.3	37.7	42.5	.01
Female	30 (37)	34 (42)	28 (35)	.61
Race				<.001
African†	8 (10)	22 (27)	2 (2)	
African American	33 (41)	27 (33)	39 (48)	
Haitian‡	10 (12)	14 (17)	0 (0)	
Latino†	17 (21)	4 (5)	13 (16)	
White§	11 (14)	10 (12)	21 (26)	
Other	2 (2)	4 (5)	6 (7)	
Schooling				<.001
Elementary	13 (16)	7 (9)	9 (11)	
High school‡	32 (40)	31 (38)	50 (62)	
College	12 (15)	14 (17)	20 (25)	
NA‡	24 (30)	29 (36)	2 (2)	
Risk factors				<.001
High risk‡//	39 (46)	35 (43)	3 (4)	
MSM‡	7 (9)	12 (15)	0 (0)	
IDU†‡	28 (35)	5 (6)	1 (1)	
Sex with HIV-positive person†//	7 (9)	20 (25)	1 (1)	
Sex with CSW	4 (5)	2 (2)	1 (1)	
Sex with IDU	4 (5)	3 (4)	0 (0)	
Sex for drugs or money	3 (4)	0 (0)	0 (0)	

*Data are number (percentage) of patients unless otherwise indicated. CSW = commercial sex worker; HITS = HIV Inpatient Testing Service; HIV = human immunodeficiency virus; IDU = injection drug user; MSM = men who have sex with men; NA = not available; OP = outpatient.
†HITS positive vs OP positive, $P<.01$.
‡HITS positive vs HITS negative, $P<.01$.
§HITS positive vs HITS negative, $P<.05$.
//Totals may not equal the sum of individual risks because some patients may have noted 1 or more high-risk factors.

positive group. All differences were statistically significant ($P\leq.001$).

When corrected for sex, educational level, racial group, age, and high-risk category, the CD4 and HIV RNA values changed only minimally, and group differences remained statistically significant. The adjusted CD4 cell count was 170 cells/mm^3 lower for HITS positive relative to OP positive subjects ($P<.001$) compared with the unadjusted difference of 175 cells/mm^3 ($P<.001$). The adjusted HIV RNA was 0.48 $\log_{10}$ RNA higher for HITS positive patients compared with OP positive patients ($P=.005$); the unadjusted difference was 0.52 $\log_{10}$ RNA ($P=.001$). The untransformed viral load produced similar results.

In HITS positive subgroup analyses (sex, race, risk category, level of education, and age), only sex was associated with significant differences in CD4 counts and $\log_{10}$ HIV RNA. The mean CD4 cell count among males was 130 cells/mm^3 compared with 302 cells/mm^3 among females ($P=.007$). The CD4 cell count difference by sex was not significant in the OP positive group (361 vs 378 cells/mm^3, respectively; $P=.77$). Of note, the mean CD4 cell count in the high-risk HITS positive group was 209 cells/mm^3 compared with 184 cells/mm^3 in the low-risk group ($P=.65$); in OP positive patients, these values were 399 vs 350 cells/mm^3, respectively ($P=.39$).

Pairwise comparison of the mean CD4 cell counts of each of the 3 HITS positive discharge diagnosis categories were significantly different except when comparing the related group to the possibly related group (Figure 2). In HITS positive patients, the subgroup who had discharge diagnoses unrelated to HIV disease had a CD4 cell count of

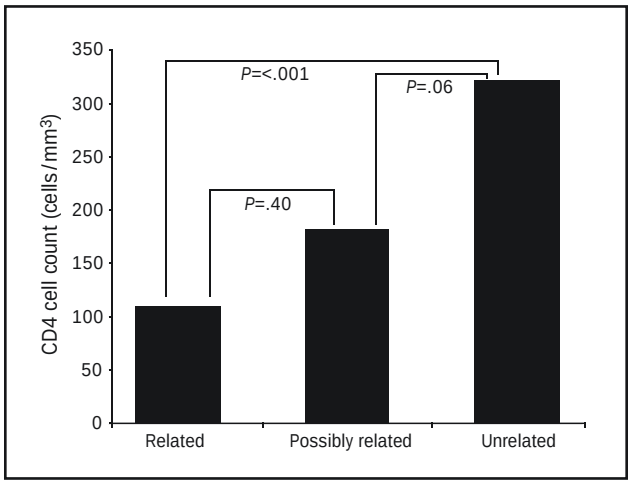


Figure 2. CD4 cell count by discharge diagnosis group in patients who tested positive for human immunodeficiency virus (HIV) according to the HIV Inpatient Testing Service.

321 cells/mm^3, which was statistically similar to the OP positive patients' mean CD4 cell count of 371 cells/mm^3 (P=.41).

Based on baseline CD4 cell count and inpatient diagnoses, 64 (79%) of the HITS positive group had an AIDS diagnosis made concurrently with their HIV diagnosis. In the OP positive group, 21 (26%) had AIDS at the time of their HIV diagnosis (P<.001).

Seven (9%) of the HITS positive patients were found to have known their HIV positive serostatus before testing by HITS, although they did not disclose this fact to the counseling staff before testing. This subgroup had a similar CD4 cell count of 208 cells/mm^3 compared with 194 cells/mm^3 in the group of patients whose HIV diagnosis was truly new (P=.88). The HIV RNA levels in these 2 groups were also similar (4.64 vs 4.34 $\log_{10}$ HIV RNA; P=.59). The percentage of patients in these subgroups considered high risk was different, but the small sample size limits the ability to discern a statistically significant difference (71% vs 43%; P=.15). Prior HIV-positive serostatus could not be included in the multivariate models because of the small sample size.

DISCUSSION

Whether it is due to logistic difficulties, financial impediments, or medical traditions, the 1993 CDC recommendation[1] to offer routine HIV counseling and testing in hospitals with HIV seroprevalences of 1% or higher or AIDS diagnoses in 1 per 1000 discharges or more has not been broadly implemented. Before this study, no data existed to clarify how many inpatients had undiagnosed HIV infection conditions despite hospitalization. Additionally, no literature exists to illuminate what types of patients routine HIV testing would identify. Relevant hypotheses include that the patients would have more advanced HIV disease than their outpatient counterparts, because their need for inpatient care, regardless of diagnosis, may reflect a higher level of immunocompromise. Alternatively, one could hypothesize that their HIV disease would be less advanced, because, in most cases, the admitting diagnoses were not opportunistic infections or other clearly HIV-related diseases. In this view, routine inpatient testing would result in earlier diagnosis than ambulatory testing, which most often occurs in response to a clinical prompt or historical "red flag."

Our data suggest that the former hypothesis is more likely to be correct since 64 (79%) of the HITS positive patients presented with AIDS at the time of HIV diagnosis compared with 21 (26%) of the OP positive group. The CD4 cell count in the HITS positive group was 175 cells/mm^3 lower (170 cells/mm^3 lower after adjusting for potential confounders) than the CD4 cell count in the OP positive group. Extrapolating from the average rate of CD4 cell count decline noted in the study by Mellors et al[10] of 60 cells/mm^3 yearly suggests that the HITS positive patients undergo testing almost 3 years later in their course of HIV disease relative to the OP positive group.

Despite the lack of widely implemented routine inpatient testing for HIV, 3 studies suggest that the inpatient setting is a common place where the diagnosis is made. In their analysis of the national HIV/AIDS Reporting System, Kates et al[11] reviewed HIV patient data between 1994 and 1999 from 25 reporting states. They found that 26.6% of the 104,780 patients in the study were diagnosed as having HIV infection while hospitalized. Of note, 67.1% of patients were diagnosed as having advanced HIV disease and developed AIDS within 1 year. Mellors et al[10] demonstrated that 33% of the HIV-positive patients had their conditions diagnosed in inpatient settings. Wortley et al[12] reported that one third of patients in their study had their conditions diagnosed as inpatients and that 58% were tested because of an illness; 51% of their patients with HIV infection developed AIDS within 1 year of diagnosis.

Although inpatients appear to have undergone testing approximately 3 years later in their disease course than their outpatient counterparts, 2 important additional points must be recognized. First, inpatients who presented with HIV-unrelated diagnoses had CD4 cell counts similar to outpatients (321 vs 371 cells/mm^3; P=.41) and therefore are not clearly later in their disease stage than outpatients. Second, more than a quarter of outpatients also had an AIDS diagnosis at presentation. Diagnosing HIV in its advanced stages among ambulatory populations has been found by other authors.[7,13,14] In the study by Neal and Fleming,[15] 43% of patients presenting with HIV were diagnosed as having AIDS within 1 year. Samet et al[16] found that 37% of patients who presented with newly diagnosed HIV infection had CD4 cell counts below 200 cells/mm^3. Liddecoat et al[8] found that 44% of HIV-positive patients who presented for initial care had CD4 cell counts less than 200 cells/mm^3 at diagnosis. The CDC reported testing patterns from 16 sites in the United States from 2000 to 2003.[17] Among patients with evaluable data, 40% of patients had less than 1 year between HIV and AIDS diagnoses. Of these late presenters, 87% tested HIV positive in acute or referral-based medical centers, and 65% were tested because of an illness.

Racial and other demographic differences in HIV testing populations have also been noted in the literature. Neal and Fleming[15] showed that among late testers the highest rates of late presentation were among patients older than 50 years, men, whites, Hispanics, men who have sex with men, and injection drug users. The CDC study[17] identified

the following risk factors for late presentation: younger age (18-29 years old), black race, Hispanic ethnicity, heterosexuality, and education level of high school or less. In our study, HITS positive patients were more likely to be older than OP positive subjects (42.3 vs 37.7 years old) and Latino, but the education levels were not statistically different.

As noted in the CDC study[17] and the study by Wortley et al,[12] most patients historically have been tested in response to an illness. Unfortunately, such a testing scheme excludes a significant number of HIV-positive patients from testing. The study by Peterman et al[18] of almost 1.3 million patients suggested that 20% to 26% of HIV-positive patients acknowledged no HIV risk factors before testing. In our study, 47% of patients had no diagnoses that were related or possibly related to HIV, and 54% were considered low risk. These data suggest that clinicians may not have obtained HIV testing had it not been offered routinely. Our data suggest that routine HIV testing on an inpatient service may double the number of patients identified with HIV infection (38-44 of the 81 HITS positive patients) by testing those who might previously have not undergone testing because of a lack of identifiable HIV risk factors or HIV-related medical problems.

Our study has a number of limitations, including the fact that before April 2001 physician referral was required to obtain HIV testing, leading to the possibility of referral bias. Subgroup analyses comparing pre-April 2001 (n=15) to post-April 2001 (n=66) patients and post-April 2001 to the entire HITS positive group showed no significant differences from the findings reported. Although the seroprevalence during the earlier period was 5.5% compared with 2.7% in the 2 years after non–referral-based testing was begun (data not shown), indicating selection bias, excluding these patients from the analysis did not significantly affect the CD4 cell count or HIV viral load data analysis or conclusions. The higher rates of HIV risk factors identified among the pre-April 2001 group also did not statistically alter the HITS positive risk factor analyses.

We have no follow-up information on 16 patients who went to other institutions for their HIV follow-up care. These missing data may also have biased our results. However, these 16 patients did not differ from the other 65 in their distribution of sex, race, schooling, or risk category, according to χ^2 tests of independence or Fisher exact tests.

Another limitation of our study is that acute illnesses can cause CD4 lymphopenia[19,20] and total lymphopenia[21] in patients not infected with HIV. This finding has led to the accepted practice of not evaluating CD4 cell count during an acute phase of illness. In our study, the median time from HIV testing to CD4 testing was 18 days, after patients retrospectively known to be HIV positive before being identified by HITS were excluded. In 43% of patients, a second CD4 measurement was obtained a week or more after the first (data not shown). Therefore, it is unlikely that the acute illness leading to hospitalization accounted for the lower CD4 cell counts observed.

The inclusion of the 7 patients previously known to be HIV positive within the analysis could have affected the results. Subgroup analyses of patients whom we were able, retrospectively, to identify that they knew about their HIV-positive status before HITS testing showed that they had CD4 cell counts and HIV RNA levels that were not statistically different from the rest of the HITS positive group. Unfortunately, some patients who know they have HIV infection do undergo retesting, and inpatient testing services, like outpatient testing services, will find these patients. Reidentifying these patients gives the testing service an opportunity to educate the patient further about HIV and to ensure that the patient is linked to care appropriately.

Also, the patient population at our medical center may not mirror the demographics at other institutions. Walensky et al[3] used the American Hospital Association database and identified 72 hospitals in the United States with demographic characteristics similar to Boston Medical Center. Therefore, there are a large number of hospitals for which these data are relevant.

An additional limitation of our study is the sample size. Although this study represents the largest examination to date of inpatient HIV testing, the limited patient numbers may have skewed our results, particularly with respect to some ethnic groups who were only minimally represented in this study. For example, although our inpatient population during the study period was 4% Haitian, there were no Haitian patients represented in the HITS negative group. Additional larger studies on inpatient testing would clearly be helpful.

Despite these limitations, the lower CD4 cell count and higher HIV RNA identified in the HITS positive group place most of these patients in a population that should strongly be considered for therapy based on current guidelines for initiation of antiretroviral therapy.[22] Most of the HITS positive patients also meet criteria for prophylactic therapy against *Pneumocystis carinii* pneumonia.[23]

CONCLUSION

In 2003, the CDC proposed a program called Advancing HIV Prevention in which HIV testing would be incorporated into routine medical care.[24] Broad implementation of routine HIV testing could have a profound effect on identifying some of the 252,000 to 312,000 of the 1,039,000 to 1,185,000 patients in the United States who are unaware of their HIV infection.[25] This recommendation follows a decade of recommendations regarding routine inpatient HIV

testing.[1,2] Although broad-based inpatient screening appears to diagnose HIV late in the course of the disease, those inpatients identified who were not hospitalized for HIV-related illnesses had CD4 cell counts similar to outpatients.

Additionally, approximately half of those inpatients whose conditions were diagnosed by HITS may not have been tested had the testing not been offered routinely. Offering HIV tests only to inpatients who present with traditional risk factors or HIV-related diagnoses clearly fails to identify a large proportion of the undiagnosed HIV disease on the inpatient service. The yield from routine screening in the inpatient setting is sufficiently high and the opportunity to integrate patients directly into care is sufficiently great to merit establishing routine inpatient HIV testing services at hospitals that meet the CDC's recommended thresholds.[1]

REFERENCES

1. Centers for Disease Control and Prevention. Recommendations for HIV testing services for inpatients and outpatients in acute-care hospital settings. *MMWR Recomm Rep*. 1993;42(RR-2):1-6.

2. Centers for Disease Control and Prevention. Revised guidelines for HIV counseling, testing, and referral. *MMWR Recomm Rep.* 2001;50(RR-19):1-57.

3. Walensky RP, Losina E, Steger-Craven KA, Freedberg KA. Identifying undiagnosed human immunodeficiency virus: the yield of routine, voluntary inpatient testing. *Arch Intern Med*. 2002;162:887-892.

4. Cohen MH, Olszewski Y, Branson B, et al. Using point-of-care testing to make rapid HIV-1 tests in labor really rapid [letter]. *AIDS*. 2003;17:2121-2124.

5. Irwin K, Olivo N, Schable CA, Weber JT, Janssen R, Ernst J, CDC-Bronx-Lebanon HIV Serosurvey Team. Performance characteristics of a rapid HIV antibody assay in a hospital with a high prevalence of HIV infection. *Ann Intern Med*. 1996;125:471-475.

6. Centers for Disease Control and Prevention. Cases of HIV infection and AIDS in the United States, 2002. *HIV/AIDS Surveill Rep*. 2002;14:1-48.

7. Klein D, Hurley LB, Merrill D, Quesenberry CP Jr, Consortium for HIV/AIDS Interregional Research. Review of medical encounters in the 5 years before a diagnosis of HIV-1 infection: implications for early detection. *J Aquir Immune Defic Syndr*. 2003;32:143-152.

8. Liddicoat RV, Horton NJ, Urban R, Maier E, Christiansen D, Samet JH. Assessing missed opportunities for HIV testing in medical settings. *J Gen Intern Med*. 2004;19:349-356.

9. Freedberg KA, Samet JH. Think HIV: why physicians should lower their threshold for HIV testing. *Arch Intern Med*. 1999;159:1994-2000.

10. Mellors JW, Munoz A, Giorgi JV, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med*. 1997; 126:946-954.

11. Kates J, Levi J, Neal J, Gallagher K. Learning more about the HIV-infected population not in care in the United States: using public health surveillance data to inform current policy challenges in enhancing access. In: Program and Abstracts of the XIV International AIDS Conference; July 7-12, 2002; Barcelona, Spain. Abstract TuPeG5690.

12. Wortley PM, Chu SY, Diaz T, et al. HIV testing patterns: where, why, and when were persons with AIDS testing for HIV? *AIDS*. 1995;9:487-492.

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

14. Dybul M, Bolan R, Condoluci D, et al. Evaluation of initial CD4+ T cell counts in individuals with newly diagnosed human immunodeficiency virus infection, by sex and race, in urban settings. *J Infect Dis*. 2002;185:1818-1821.

15. Neal JJ, Fleming PL. Frequency and predictors of late HIV diagnosis in the United States, 1994 through 1999. In: Program and Abstracts of the 9th Conference on Retroviruses and Opportunistic Infections; February 24-28, 2002; Seattle, Wash. Abstract 474-M.

16. Samet JH, Freedberg KA, Stein MD, et al. Trillion virion delay: time from testing positive for HIV to presentation for primary care. *Arch Intern Med*. 1998;158:734-740.

17. Centers for Disease Control and Prevention. Late versus early testing of HIV—16 sites, United States, 2000-2003. *MMWR Morb Mortal Wkly Rep*. 2003;52:581-586.

18. Peterman TA, Todd KA, Mupanduki I. Opportunities for targeting publicly funded human immunodeficiency virus counseling and testing. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1996;12:69-74.

19. Aldrich J, Gross R, Adler M, King K, MacGregor RR, Gluckman SJ. The effect of acute severe illness on CD4+ lymphocyte counts in nonimmunocompromised patients [letter]. *Arch Intern Med.* 2000:160:715-716.

20. Fantin B, Joly V, Elbim C, et al. Lymphocyte subset counts during the course of community-acquired pneumonia: evolution according to age, human immunodeficiency virus status, and etiologic microorganisms. *Clin Infect Dis*. 1996;22:1096-1098.

21. Collazos, J, Mayo J, Ibarra S, Martinez E. The primary deficiency resulting from acute severe illness in nonimmunocompromised patients is not the CD4+ subtype, but the total number of lymphocytes [letter]. *Arch Intern Med.* 2001;161:771-772.

22. US Department of Health and Human Services. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents—October 29, 2004. Available at: http://aidsinfo.nih.gov/Guidelines/ArchivedGuidelines.aspx?MenuItem=Guidelines&Search=Off&GuidelineID=18&ClassID=6. Accessed February 21, 2006.

23. Kaplan JE, Masur H, Holmes KK. Guidelines for preventing opportunistic infections among HIV-infected persons—2002: recommendations of the U.S. Public Health Service and the Infectious Diseases Society of America. *MMWR Recomm Rep*. 2002;51(RR-8):1-52.

24. Centers for Disease Control and Prevention. Advancing HIV prevention: new strategies for a changing epidemic—United States, 2003. *MMWR Morb Mortal Wkly Rep.* 2003;52:329-332.

25. Glynn M, Rhodes P. Estimated HIV prevalence in the United States at the end of 2003. In: Program and Abstracts of the National HIV Prevention Conference; June 12-15, 2005; Atlanta, Ga. Abstract T1-B1101.

CHEST

Special Features

Grading Strength of Recommendations and Quality of Evidence in Clinical Guidelines*

Report From an American College of Chest Physicians Task Force

Gordon Guyatt, MD, MSc, FCCP; David Gutterman, MD, FCCP; Michael H. Baumann, MD, MSc, FCCP; Doreen Addrizzo-Harris, MD, FCCP; Elaine M. Hylek, MD, MPH; Barbara Phillips, MD, FCCP; Gary Raskob, PhD; Sandra Zelman Lewis, PhD; and Holger Schünemann, MD, PhD, FCCP

While grading the strength of recommendations and the quality of underlying evidence enhances the usefulness of clinical guidelines, the profusion of guideline grading systems undermines the value of the grading exercise. An American College of Chest Physicians (ACCP) task force formulated the criteria for a grading system to be utilized in all ACCP guidelines that included simplicity and transparency, explicitness of methodology, and consistency with current methodological approaches to the grading process. The working group examined currently available systems, and ultimately modified an approach formulated by the international GRADE group. The grading scheme classifies recommendations as strong (grade 1) or weak (grade 2), according to the balance among benefits, risks, burdens, and possibly cost, and the degree of confidence in estimates of benefits, risks, and burdens. The system classifies quality of evidence as high (grade A), moderate (grade B), or low (grade C) according to factors that include the study design, the consistency of the results, and the directness of the evidence. For all future ACCP guidelines, The College has adopted a simple, transparent approach to grading recommendations that is consistent with current developments in the field. The trend toward uniformity of approaches to grading will enhance the usefulness of practice guidelines for clinicians. ***(CHEST 2006; 129:174–181)***

Key words: grading recommendations; grading system; methodology

Abbreviations: ACCP = American College of Chest Physicians; RCT = randomized controlled trial; RRR = relative risk reduction

Treatment decisions involve a tradeoff between benefits on the one hand, and risks, burdens, and, potentially, costs on the other. Guideline panels provide recommendations for the management of typical patients. To integrate these recommendations with their own clinical judgment, and with individual patient values and preferences, clinicians need to understand the basis for the recommendations that expert guidelines offer. A systematic approach to grading the strength of management recommendations can minimize bias and aid interpretation.[3] Indeed, most guideline groups have accepted the necessity for some sort of grading scheme.

While the grading of recommendations represents

*From the Departments of Medicine, and Clinical Epidemiology and Biostatistics (Dr. Guyatt), McMaster University, Hamilton, ON, Canada; the Department of Medicine (Dr. Gutterman), Medical College of Wisconsin, Milwaukee, WI; the University of Mississippi Medical Center (Dr. Baumann), Jackson, MS; New York University School of Medicine (Dr. Addrizzo-Harris), New York, NY; the Department of Medicine (Dr. Hylek), Research Unit-Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; the University of Kentucky College of Medicine (Dr. Phillips), Lexington, KY; College of Public Health (Dr. Raskob), University of Oklahoma Health Sciences Center, Oklahoma City, OK; the American College of Chest Physicians (Dr. Lewis), Northbrook, IL; and the Department of Clinical Epidemiology, Italian National Cancer Institute (Dr. Schünemann), Rome, Italy.

Manuscript received August 16, 2005; revision accepted August 21, 2005.

Correspondence to: Gordon Guyatt, MD, MSc, FCCP, Department of Clinical Epidemiology and Biostatistics, HSC-2C12, McMaster University, 1200 Main St West, Hamilton, ON, Canada L8N 3Z5; e-mail: guyatt@mcmaster.ca

a positive development for guideline development and interpretation, the proliferation of grading systems has proved to be an unfortunate consequence. Methodologists and guideline developers have given much thought and effort to considering the criteria and approaches to an optimal grading system. The American College of Chest Physicians (ACCP) convened a working group to review the issue and to agree on a grading system that would be consistent with the latest developments in the field.

For editorial comment see pages 7 and 10

The task force began by developing criteria that define an optimal grading system (Table 1), placing them in an order that approximates their relative importance. These criteria guided the decisions of the group in the choice of the grading system that follows.

Strength of Recommendation

Guideline panels should make recommendations to administer, or not administer, an intervention, on the basis of tradeoffs between benefits on the one hand, and risks, burdens, and, potentially, costs on the other. If benefits outweigh risks and burdens, experts will recommend that clinicians offer a treatment to appropriately chosen patients. The uncertainty associated with the tradeoff between the benefits and the risks and burdens will determine the strength of recommendations.

The ACCP task force chose to classify recommendations into two levels, strong and weak (Table 2). If guideline panelists are very certain that benefits do, or do not, outweigh risks and burdens, they will make a strong recommendation, grade 1. If they think that the benefits and the risks and burdens are finely balanced, or if appreciable uncertainty exists about the magnitude of the benefits and risks, they must offer a weak, grade 2 recommendation.

A two-level grading system has the merit of simplicity. Two levels also facilitate the clear interpretation of the implications of strong and weak recommendations by clinicians. We offer three ways that clinicians can interpret strong and weak recommendations. We have already presented the first way. A strong recommendation signifies that benefits clearly outweigh the risks, or the reverse; a weak recommendation signifies that benefits and risks are closely balanced, or uncertain.

Clinicians are becoming increasingly aware of the importance of patient values and preferences in clinical decision making.[4] A second way to interpret strong and weak recommendations is in relation to patient values and preferences. For decisions in which it is clear that benefits far outweigh risks, or risks far outweigh benefits, virtually all patients will make the same choice (see box 1 for an example). In such instances, guideline panels can offer a strong (grade 1) recommendation. In contrast, there are other choices in which patient values and preferences will play a crucial role and in which patients will, as a result, make different choices. See boxes 2

Box 1: Short-term aspirin reduces the relative risk of death after myocardial infarction by approximately 25%. Aspirin has minimal side effects and very low cost. Peoples' values and preferences are such that virtually all patients suffering a myocardial infarction would, if they understood the choice they were making, opt to receive aspirin. Guideline panels can thus offer a strong recommendation for aspirin administration in this setting.

Box 2: Consider a patient a 40 year-old man who has suffered an idiopathic deep venous thrombosis and has been taking adjusted dose warfarin for one year. If the patient continues on standard-intensity warfarin his risk of recurrent DVT will be reduced by approximately 10% per year.[1] The inevitable burdens of the treatment include taking a warfarin pill daily, keeping dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. Some patients who are very averse to a recurrent DVT may consider the down sides of taking warfarin well worth it. Others are likely to consider the benefit not worth the risks and inconvenience.

Table 1—*Criteria for an Optimal Grading System*

Criteria	Description
1	Separation of grades of recommendations from quality of evidence
2	Simplicity and transparency for clinician consumer
3	Sufficient (but not too many) categories
4	Explicitness of methodology for guideline developers
5	Simplicity for guideline developers
6	Consistent with general trends in grading systems
7	Explicit approach to different levels of evidence for different outcomes

Table 2—*Grading Recommendations*

Grade of Recommendation/ Description	Benefit vs Risk and Burdens	Methodological Quality of Supporting Evidence	Implications
1A/strong recommendation, high-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	RCTs without important limitations or overwhelming evidence from observational studies	Strong recommendation, can apply to most patients in most circumstances without reservation
1B/strong recommendation, moderate quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies	Strong recommendation, can apply to most patients in most circumstances without reservation
1C/strong recommendation, low-quality or very low-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	Observational studies or case series	Strong recommendation but may change when higher quality evidence becomes available
2A/weak recommendation, high-quality evidence	Benefits closely balanced with risks and burden	RCTs without important limitations or overwhelming evidence from observational studies	Weak recommendation, best action may differ depending on circumstances or patients' or societal values
2B/weak recommendation, moderate-quality evidence	Benefits closely balanced with risks and burden	RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies	Weak recommendation, best action may differ depending on circumstances or patients' or societal values
2C/weak recommendation, low-quality or very low-quality evidence	Uncertainty in the estimates of benefits, risks, and burden; benefits, risk, and burden may be closely balanced	Observational studies or case series	Very weak recommendations; other alternatives may be equally reasonable

Box 3: A systematic review of randomized trials suggests that in 1,000 patients with ST elevation myocardial infarction who are receiving thrombolytic therapy and aspirin and who are treated with heparin (versus no treatment with heparin) 5 fewer will die, 3 fewer will have reinfarction, and 1 fewer will have a pulmonary embolus, while 3 more will have major bleeds.[2] Further, these estimates are not precise, and the advantage in decreased infarctions may be lost after six months. The small, imprecise and possibly transient benefit leaves us less confident about any recommendation to use heparin in this situation. Hence, the recommendation is likely to be weak.

and 3 for examples. When, across the range of patient values, fully informed patients are liable to make different choices, guideline panels should offer weak (grade 2) recommendations.

Following closely from this reasoning, a third way for clinicians to interpret strong recommendations is, for typical patients, to just do it. On the other hand, when clinicians face weak recommendations, or when they face patients with very atypical circumstances or values, they should carefully consider the benefits, risks, and burdens in the context of the individual patient before them.

How to individualize decision making in weak recommendations remains a challenge. One strategy uses decision aids that present patients with both the benefits and downsides of therapy.[5] Because of time constraints, clinicians cannot use decision aids in all patients. For strong recommendations, using a decision aid is likely, for most patients, to constitute a poor use of time and energy. For weak recommendations, clinicians should consider the use of a decision aid or, alternatively, a detailed conversation with the patient to ensure that the ultimate decision is consistent with the patient's values.

Factors That Influence the Strength of a Recommendation

Guideline panels must consider a number of factors in grading recommendations (Table 3). One issue is their confidence in the best estimates of benefit and harm. The rating of methodological quality, which we discuss below, captures that degree of confidence.

Table 3—*Factors Panels Should Consider in Deciding on a Strong or Weak Recommendation

Issue	Example
Methodological quality of the evidence supporting estimates of likely benefit, and likely risk, inconvenience, and costs	Many high-quality randomized trials have demonstrated the benefit of therapy with inhaled steroids in patients with asthma, while only case series have examined the utility of pleurodesis in patients with pneumothorax
Importance of the outcome that treatment prevents	Preventing postphlebitic syndrome with thrombolytic therapy in DVT patients in contrast to preventing death from PE
Magnitude of treatment effect	Clopidogrel vs aspirin leads to a smaller stroke reduction in patients with TIAs (RRR,[19] 8.7%) than anticoagulation vs placebo in patients with AF (RRR, 68%)
Precision of estimate of treatment effect	ASA therapy vs placebo in AF patients has a wider confidence interval than ASA therapy for stroke prevention in patients with TIA
Risks associated with therapy	ASA and clopidogrel for anticoagulation therapy in patients with acute coronary syndromes has a higher risk for bleeding than ASA alone
Burdens of therapy	Therapy with adjusted-dose warfarin is associated with a higher burden than that with aspirin; warfarin requires monitoring the intensity of anticoagulation and a relatively constant dietary vitamin K intake
Risk of target event	Some surgical patients are at very low risk of post-operative DVT and PE while other surgical patients have considerably higher rates of DVT and PE
Costs	Clopidogrel has a much higher cost in patients with TIA than does aspirin
Varying values	Most young, healthy people will put a high value on prolonging their lives (and thus incur suffering to do so); the elderly and infirm are likely to vary in the value they place on prolonging their lives (and may vary in the suffering they are ready to experience to do so)

*DVT = deep vein thrombosis; PE = pulmonary embolism; TIA = transient ischemic attack; AF = atrial fibrillation; ASA = aspirin.

The prevention of outcomes with high patient importance[6] should, in general, lead to stronger recommendations than the prevention of outcomes of lesser patient importance. For instance, one needs to expose four patients to a respiratory rehabilitation program for one patient to gain a small but important improvement in dyspnea in daily life.[7] In low-risk patients who have experienced a myocardial infarction, one might need to treat 100 patients with agents such as aspirin, β-blockers, angiotensin-converting enzyme inhibitors, or statins, to extend the life of one patient. Despite the much higher number needed to treat, since we value prolongation of life more highly than relieving dyspnea, the latter intervention may warrant a stronger recommendation.

The choice of adjusted-dose warfarin vs aspirin for the prevention of stroke in patients with atrial fibrillation illustrates a number of the factors that will influence the strength of a recommendation. A systematic review and metaanalysis[8] found a relative risk reduction (RRR) of 46% in all strokes with warfarin vs aspirin. This large effect supports a strong recommendation for warfarin. Furthermore, the relatively narrow 95% confidence interval (RRR, 29 to 57%) suggests that warfarin provides an RRR of at least 29%, and further supports a strong recommendation. At the same time, warfarin is associated with the inevitable burdens of keeping the dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. Most patients, however, are much more stroke averse than they are bleeding averse.[8] As a result, almost all patients with high risk of stroke would choose therapy with warfarin, suggesting the appropriateness of a strong recommendation.

This last point emphasizes the importance of the patient's baseline risk of the adverse outcome that treatment is designed to avoid. Consider a 65-year-old patient with atrial fibrillation and no other risk factors for stroke. This individual's risk for stroke in the next year is approximately 2%. Therapy with dose-adjusted warfarin can, relative to aspirin, reduce the risk to approximately 1%. Some patients who are very stroke-averse may consider the downside of receiving warfarin therapy to be well worth it. Others are likely to consider the benefit not worth the risks and inconvenience. When, across the range of patient values, fully informed patients are liable to make different choices, guideline panels should offer weak (grade 2) recommendations.

As benefits and risks become more finely balanced, or more uncertain, decisions to administer an effective therapy also become more sensitive to resource use (cost) implications. When dealing with resource allocation issues, guideline panels face challenges of limited expertise, paucity of rigorous and unbiased cost-effectiveness analyses, and wide variability of costs across jurisdictions or health-care systems. Ignoring the issue of resource use (costs) is, however, becoming less and less tenable for guideline panels.[9]

When guideline developers make recommendations, they assume a particular set of values as they

weigh the possible beneficial and detrimental outcomes. When value or preference judgments are particularly salient, guideline panels should describe the key values attached to these outcomes and that influenced the direction of a recommendation or its grade. Guideline panels often do not elicit direct or indirect representation from patients in arriving at these values. Moreover, recommendations can only reflect average values. These considerations emphasize the importance of guideline panels making explicit the key values and preference judgments that drive their recommendations.

Wording of Recommendations

Given the proliferation of grading systems, and the resulting confusion, it is desirable to provide clinicians with as many indicators as possible in interpreting the strength of recommendations. ACCP panels, when they are making a strong recommendation, will use the terminology, "We recommend. . . . " When they make a weak recommendation, ACCP guideline panels will use less definitive wording, such as, "We suggest. . . . " Further, the clarity of recommendations requires that the target patient population be defined and, when appropriate, the details of how clinicians should administer the intervention.

Confidence in Estimates of Magnitude of Benefits, Risks, Burdens, and Costs

Early systems of grading methodological quality relied primarily on the basic study design (*ie*, randomized control trials [RCTs], or observational studies). The fundamental study design remains critically important in determining our confidence in estimates of beneficial and detrimental treatment effects. Because of prognostic differences between groups, and the lack of safeguards such as blinding that can avoid biased ascertainment of outcomes, evidence based on observational studies will, in general, be appreciably weaker than evidence from RCTs. The last several years have seen, however, an increased awareness of a number of other factors that influence our confidence in our estimates of risk and benefit (Table 4).

ACCP recommendations will henceforth use a three-category system of quality of evidence, as follows: high (grade A); moderate (grade B); and low quality (grade C) [Table 2]. Ideally, guideline panels will have available to them systematic reviews of the evidence regarding the benefits and risks of the alternative management strategies they are considering. Guideline panels will have the strongest evidence possible when such reviews reveal one or more well-designed and well-executed RCTs yielding consistent directly applicable results. Strong evidence can also come, under unusual circumstances, from observational studies yielding very large effects.

Table 4—*Factors Panels Should Consider in Deciding on Their Confidence in Estimates of Benefits, Risks, Burden, and Costs*

Factor Type	Factors
Factors that may decrease the quality of evidence based on RCTs	Poor quality of planning and implementation of the available RCTs suggesting high likelihood of bias
	Inconsistency of results
	Indirectness of evidence
	Sparse evidence
Factors that may increase the quality of evidence based on observational studies	Large magnitude of effect
	All plausible confounding would reduce a demonstrated effect
	Dose-response gradient

The moderate quality category is populated by randomized trials with important limitations and by exceptionally strong observational studies. Observational studies, and on occasion RCTs with multiple serious limitations, will fill the low-quality evidence category. This categorization follows the principle that all relevant clinical studies provide evidence, the quality of which varies. Following this principle, the ACCP does not use a threshold for "acceptable evidence" in the peer-reviewed published medical literature.

Factors That Modify the Quality of Evidence: Limitations in RCTs

When RCTs have addressed the impact of alternative management strategies (both benefits and harms) on all relevant outcomes, they will yield high-quality evidence unless they have one of a number of limitations. The following limitations may decrease the quality of evidence supporting a recommendation (Table 4).

1. Our confidence in recommendations decreases if the available RCTs have major deficiencies that are likely to result in a biased assessment of the treatment effect. These methodological limitations include a very large loss to follow-up, or an unblinded study with subjective outcomes that are highly susceptible to bias. How lack of blinding can influence the grading is exemplified by a recommendation to treat heparin-induced thrombocytopenia complicated by thrombosis with danaparoid sodium. The

randomized trial evidence for danaparoid use in patients with heparin-induced thrombocytopenia comes from an unblinded trial[20] in which the outcome was the clinicians' assessment of when the thromboembolism had resolved, which is a subjective judgment. As a result, an ACCP guideline panel rated the quality of the evidence as moderate rather than strong.[10]

2. When several RCTs yield widely differing estimates of treatment effect (heterogeneity or variability in results) investigators look for explanations for that heterogeneity. For instance, drugs may have larger relative effects in sicker, or in less sick, populations. When heterogeneity exists, but investigators fail to identify a plausible explanation, the strength of recommendations from even rigorous RCTs is weaker. For example, RCTs of pentoxifylline in patients with intermittent claudication have shown conflicting results that so far defy explanation. Acknowledging the unexplained heterogeneity, an ACCP guideline panel rated the quality of the evidence for pentoxifylline as moderate, rather than high.[11]
3. Investigators may have undertaken RCTs in similar populations, but not identical populations, to those of interest to a guideline panel. Panels should consider this to be indirect evidence and, to the extent they are uncertain about the applicability to their relevant population, should downgrade the quality of evidence. For instance, while graduated compression stockings have proven to be of benefit in a variety of populations at risk for venous thrombosis, they have never been tested directly in trauma patients. An ACCP guideline panel judged the available RCTs to be relevant to trauma patients in whom the administration of low-molecular-weight heparin is contraindicated, but because of concern about generalizing from other populations (that is, concern about the indirectness of the evidence), rated the quality of the evidence as moderate. Had they had no concerns about directness, they would have considered the evidence to be of high quality, whereas if there were no relevant RCTs available, and the best evidence came from observational studies, they would have rated the evidence to be of low quality.[12] Indirectness may also apply to the intervention [(*eg*, RCTs of similar but not identical interventions or different doses and formulations)] and outcomes (*eg*, RCTs measuring laboratory exercise capacity when a panel is really interested in quality-of-life improvement).
4. Investigators may have conducted RCTs, but included very few patients and observed very few events. For instance, a well-designed and rigorously conducted RCT addressed the use of nadroparin, a low-molecular-weight heparin, in patients with cerebral venous sinus thrombosis. Of 30 treated patients, 3 had a poor outcome, as did 6 of 29 patients in the control group. The investigators' analysis suggested a 38% reduction in the relative risk of a poor outcome, but the result was not statistically significant.[13] Because of the small number of patients, and the small number of events, an ACCP guideline panel judged the quality of the evidence for anticoagulation in cerebral sinus thrombosis as moderate rather than high.[12]

Factors That Modify the Quality of Evidence: Observational Studies Can Provide Moderate or Strong Evidence

While observational studies will generally yield only low-quality evidence, there may be unusual circumstances in which guideline panels will classify such evidence as of moderate quality, or even high quality.

1. On the rare occasions when they yield extremely large and consistent estimates of the magnitude of a treatment effect, we may be confident about the results of observational studies. For example, oral anticoagulation in mechanical heart valves has not been compared to placebo in an RCT. However, evidence from observational studies suggests that the probability of experiencing thromboembolic events without anticoagulation is 12.3% annually in patients with bileaflet prosthetic aortic valves and higher for those with other valve types,[14] and estimates of the RRR with oral anticoagulation are in the range of 80%. While the observational studies are likely to overestimate the true effect, the weak study design is very unlikely to explain the entire benefit. Thus, an ACCP guideline panel concluded that these data, despite the absence of randomized trials, constituted strong evidence of the effectiveness of anticoagulation in bileaflet aortic prosthetic valves.[15]
2. On equally rare occasions, all plausible biases from observational studies may be working to underestimate an apparent treatment effect. In other words, the actual treatment effect is very likely to be larger than what the data suggest. For instance, a rigorous systematic review of observational studies including a total of 38 million patients compared private for-profit vs

private not-for-profit hospital care. The meta-analysis[16] demonstrated higher death rates in the private for-profit hospitals.

The investigators postulated two likely sources of bias. The first was residual confounding with disease severity. It is likely that, if anything, patients in the not-for-profit hospitals were sicker than those in the for-profit hospitals. Thus, to the extent that residual confounding existed, it would bias results against the not-for-profit hospitals.

The second likely bias was the possibility that higher numbers of patients with excellent private insurance coverage could lead to a hospital having more resources and to a "spillover" effect that would benefit those without such coverage. Since for-profit hospitals are likely to admit a larger proportion of such well-insured patients than are not-for-profit hospitals, the bias is once again against the not-for-profit hospitals. Because the plausible biases would all diminish the demonstrated treatment effect, one might consider the evidence from these observational studies as being of moderate quality rather than of low quality.

What to Do When Quality of Evidence Differs Across Outcomes?

When RCT results are available, the quality of evidence will often differ between primary efficacy and toxicity outcomes, usually between efficacy outcomes and cost, and almost always between efficacy outcomes and rare but serious side effects. On most occasions, efficacy outcomes will be the most important, and guideline panels can base their rating of the quality of the evidence exclusively on these end points. Panels should, however, consider whether toxicity end points are also crucial to the decision regarding the optimal management strategy. If they are, panels should consider the quality of evidence regarding those end points, and should make a final rating about the quality of evidence accordingly.

For instance, consider a guideline panel addressing the use of long-term oral steroids for patients with stage 2 or 3 sarcoidosis with moderate-to-severe symptoms and radiographic changes. Randomized trials have addressed the impact of steroids on radiographic findings, symptoms, and spirometry over a period of 2 years.[17] These trials failed, however, to address steroid toxicity. If a guideline panel ignored toxicity, they might well rate the quality of evidence as high. If, however, they consider steroid toxicity as crucial in their decision, the uncertainty about the impact of treatment increases. If they look for observational studies to estimate steroid toxicity, the quality of the evidence about toxicity is likely to be low, and this may be the most appropriate rating for the overall quality of evidence. Alternatively, they may seek randomized trials of steroids in other conditions and face limitations of directness. They may then conclude that the evidence regarding steroid toxicity, and the overall quality of the evidence, is moderate.

The ACCP Grading System and Initiatives Toward Uniform Grading Across Guideline Panels

In considering alternative grading systems, we found that the structure and guides for application and interpretation suggested by the GRADE group largely met the criteria in Table 1.[18] As a result, the categories presented in Table 2 permit similar interpretation to those of the GRADE group. The important aspect in which the ACCP task force approach differs is in combining low-quality and very low-quality evidence. While we achieved the primary goal of the ACCP task force, to identify a unified grading system for all future ACCP evidence-based guidelines, this exercise went beyond that goal. This article will facilitate the adoption of uniform guidelines through a simple, straightforward presentation that any guideline panel interested in the principles underlying Table 2 will find useful.

Clinicians' understanding of systems of grading the strength of recommendations and quality of evidence will also benefit if systems map easily onto one another. The ACCP mapping onto the GRADE system is obvious, and the approach that the ACCP has adopted also maps easily onto other systems, including that of the ACC/AHA and prior ACCP guideline grading systems, further facilitating understanding and usefulness.

Summary

In the system that the ACCP has adopted, the strength of any recommendation depends on the following two factors: the tradeoff between the benefits and the risks and burdens; and the quality of the evidence regarding treatment effect. We grade the tradeoff between the benefits, and the risks and burdens into the following two categories; category 1, in which the tradeoff is clear enough that most patients, despite differences in values, would make the same choice, leading to a strong recommendation; and category 2, in which the tradeoff is less clear, and individual patient values will likely lead to different choices, leading to a weak recommendation. We grade methodological quality in terms of the following three categories: randomized trials that show consistent results, or observational studies with

very strong treatment effects; randomized trials with limitations, or observational studies with exceptional strengths; and observational studies without exceptional strengths and case series. The framework summarized in Table 2 generates recommendations from the very strong (benefit/risk tradeoff unequivocal, high-quality evidence, grade 1A) to the very weak (benefit/risk questionable, low-quality evidence, grade 2C). Whatever the grade of the recommendation, clinicians must use their judgment, considering both local and individual patient circumstances, and patient values, in making individual decisions. In general, however, they should place progressively greater weight on expert recommendations as they move from grade 2C to grade 1A.

References

1 Büller H, Agnelli G, Hull R, et al. Antithrombotic therapy for venous thromboembolic disease. Chest 2004; 126:401S–428S
2 Collins R, MacMahon S, Flather M, et al. Clinical effects of anticoagulant therapy in suspected acute myocardial infarction: systematic overview of randomised trials. BMJ 1996; 313:652–659
3 Guyatt G, Sinclair J, Cook D, et al. Grading recommendations: a qualitative approach. In: Guyatt GR, Rennie D, ed. Users' guide to the medical literature: a manual for evidence-based practice. Chicago, IL: AMA Press, 2002
4 Guyatt G, Straus S, McAlister F, et al. Incorporating patient values. In: Guyatt GR, Rennie D, ed. Users' guide to the medical literature: a manual for evidence-based practice. Chicago, IL: AMA Press, 2002
5 O'Connor A, Stacey D, Entwistle V, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev, Issue 2, (database online) 2003
6 Guyatt G, Montori V, Devereaux P, et al. Patients at the centre: in our practice, and in our use of language. ACP J Club 2004; 140:A11–A12
7 Goldstein R, Gort E, Guyatt G, et al. Economic analysis of respiratory rehabilitation. Chest 1997; 112:370–379
8 Devereaux P, Anderson D, Gardner M, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ 2001; 323:1218–1222
9 Guyatt G, Baumann M, Pauker S, et al. Addressing resource allocation issues in recommendations from clinical guideline panels. Chest 2006; 129:182–187
10 Warkentin T, Greinacher A. Heparin-induced thrombocytopenia: recognition, treatment, and prevention: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004; 126:311S–337S
11 Clagett G, Jackson M, Lip G, et al. Antithrombotic therapy in peripheral arterial occlusive disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004; 126:609S–626S
12 Geerts W, Pineo G, Heit J, et al. Prevention of venous thromboembolism: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004; 126: 338S–400S
13 de Bruijn S, Stam J. Randomized, placebo-controlled trial of anticoagulant treatment with low-molecular-weight heparin for cerebral sinus thrombosis. Stroke 1999; 30:484–488
14 Baudet E, Puel V, McBride J, et al. Long-term results of valve replacement with the St. Jude Medical prosthesis. J Thorac Cardiovasc Surg 1995; 109:858–870
15 Salem D, Stein P, Al-Ahmad A, et al. Antithrombotic therapy in valvular heart disease—native and prosthetic: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004; 126:457S–482S
16 Devereaux P, Choi P, Lacchetti C, et al. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. CMAJ 2002; 166:1399–1406
17 Paramothayan S, Jones P. Corticosteroid therapy in pulmonary sarcoidosis: a systematic review. JAMA 2002; 287:1301–1307
18 Atkins D, Best D, Briss P, et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328:1490
19 CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). Lancet 1996; 348:1329–1339
20 Chong BH, Gallus AS, Cade JF, et al. Prospective randomised open-label comparison of danaparoid with dextran 70 in the treatment of heparin-induced thrombocytopenia with thrombosis: a clinical outcome study. Thromb Haemost 2001; 86:1170–1175

Letters to the Editor

Stroke *welcomes Letters to the Editor and will publish them, if suitable, as space permits. They should not exceed 750 words (including references) and may be subject to editing or abridgment. Please submit letters in duplicate, typed double-spaced. Include a fax number for the corresponding author and a completed copyright transfer agreement form (published in every issue).*

Response to Letter by Testa et al

Response:

We appreciate the thoughtful comments of Testa et al, particularly regarding the pivotal issue of stroke prevention among elderly individuals with atrial fibrillation. In our study, patients with planned cardioversion were treated with warfarin consistent with current guidelines.[1] Antiplatelet agents would not be used in this setting. Based on risk factors, the patients in our study were predominantly classified as high risk for stroke. Aspirin in this group would not be considered a surrogate for warfarin. For each patient who was not discharged on warfarin, we ascertained the reason directly either from the discharge record or the treating physician. Among patients considered to have a contraindication to warfarin, 18 patients (9%, not 18% as cited by Testa et al) were alternatively treated with aspirin plus clopidogrel. Currently, aspirin is the only alternative for warfarin ineligible patients. Aspirin has been shown to be only modestly efficacious in preventing stroke among patients with atrial fibrillation (19% stroke reduction).[2] The efficacy of dual antiplatelet therapy versus aspirin monotherapy among patients who are warfarin ineligible has yet to be determined. This specific question is being addressed in the ongoing ACTIVE A arm of the ACTIVE trial.[3] Testa et al cite the results of the recently published ACTIVE W trial which compared clopidogrel plus aspirin to oral anticoagulation and bears no relevance to a warfarin ineligible subset. Regarding dose of aspirin, we found that hemorrhage was the only independent predictor of receiving low dose versus full-strength aspirin at hospital discharge (adjusted odds ratio, 5.4 [95% CI 2.1 to 13.7]).

Twenty-two percent of the patients in our study were discharged on no antithrombotic therapy. Table 4 in our article provides a full account of the cited warfarin contraindications for these 44 patients who were discharged on neither warfarin nor aspirin. Fifty-two percent of these patients had current or recurrent extracranial hemorrhage or prior intracranial hemorrhage. Stroke prevention is particularly challenging among patients with a propensity for bleeding which often leads to cessation of therapy. The optimal management of these patients has yet to be defined.

It is important to note that guidelines, although evidence-based, are derived from trial populations that are often trial-eligible based on a favorably low bleeding risk profile. Among the older, hospitalized patients with atrial fibrillation in our study, the prevalence of risk factors for stroke was rivaled by the increased presence of potential contraindications to anticoagulant therapy. Given the devastating consequences of an atrial fibrillation–related stroke and the associated 30-day mortality of 24%,[4] the risk to benefit analysis continues to weigh in favor of anticoagulation. In clinical care, the practical challenge lies in maintaining patients on oral anticoagulant therapy, particularly after a hemorrhagic event. To the degree that anticoagulation intensity contributes to bleeding risk, newer drugs with wider therapeutic margins and shorter half-lives should help to minimize this risk. To the extent that frequent monitoring is a barrier to the use of vitamin K antagonists, newer drugs without this requirement should help to extend therapy to more of those for whom it is indicated.

Disclosures

Dr Hylek has served on an advisory board for Bristol-Myers Squibb and has received research funding from Bristol-Myers Squibb and AstraZeneca.

Elaine M. Hylek, MD, MPH
Lori E. Henault, MPH
Department of Medicine
Research Unit-Section of General Internal Medicine
Boston University School of Medicine
Boston, Mass

Carmella Evans-Molina, MD
Department of Medicine
University of Virginia
Charlottesville, Va

Carol Shea, RN
Susan Regan, PhD
Department of Medicine
Massachusetts General Hospital
Boston, Mass

1. Writing Committee Members, Fuster V, Ryden LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heuzey J-Y, Kay GN, Lowe JE, Olsson SB, Prystowsky EN, Tamargo JL, Wann S, ACC/AHA Task Force M, Smith SC Jr, Jacobs AK, Adams CD, Anderson JL, Antman EM, Halperin JL, Hunt SA, Nishimura R, Ornato JP, Page RL, Riegel B, Esc Committee For Practice G, Priori SG, Blanc J-J, Budaj A, Camm AJ, Dean V, Deckers JW, Despres C, Dickstein K, Lekakis J, McGregor K, Metra M, Morais J, Osterspey A, Tamargo JL, Zamorano JL. ACC/AHA/ESC 2006 Guidelines for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation): Developed in Collaboration With the European Heart Rhythm Association and the Heart Rhythm Society. *Circulation*. 2006; 114:e257–e354.
2. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy To prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med*. 1999;131:492–501.
3. Active Writing Group on behalf of the ACTIVE Investigators. Clopidogrel plus aspirin versus oral anticoagulation for atrial fibrillation in the Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE W): a randomised controlled trial. *Lancet*. 2006;367: 1903–1912.
4. Hylek EM, Go AS, Chang Y, Jensvold NG, Henault LE, Selby JV, Singer DE. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med*. 2003;349:1019–1026.

(*Stroke*. **2006;37:2868.**)

Stroke **is available at http://www.strokeaha.org** **DOI: 10.1161/01.STR.0000248210.87666.c6**

Translating the Results of Randomized Trials into Clinical Practice

The Challenge of Warfarin Candidacy Among Hospitalized Elderly Patients With Atrial Fibrillation

Elaine M. Hylek, MD, MPH; James D'Antonio, MD; Carmella Evans-Molina, MD; Carol Shea, RN; Lori E. Henault, MPH; Susan Regan, PhD

Background and Purpose—Numerous studies have documented under use of warfarin particularly among elderly patients. A better understanding of the discrepancy between trials and clinical practice will help inform stroke prevention strategies in this vulnerable age group. The study objective was to prospectively assess the use of antithrombotic therapy among a contemporary cohort of patients with atrial fibrillation at the time of hospital discharge. In addition to baseline characteristics, we sought to define the physician-cited reason for not prescribing warfarin for each patient.

Methods—Patients with atrial fibrillation were prospectively identified and followed to hospital discharge. Enrolled patients were ≥65 years of age, not taking warfarin on admission, and had their longitudinal care provided at our institution. Predictors of warfarin use were determined and physician-cited contraindications were compared across age groups.

Results—Fifty-one percent (n=206) of patients were discharged on warfarin: 75% of those 65 to 69 years of age, 59% 70 to 79, 45% 80 to 89, and 24% age ≥90 years. Of the remaining 199 patients, 83% had ≥2 major risk factors for stroke, and 98% were felt to have contraindications including nearly 25% who were unable to tolerate warfarin in the past. Among patients age ≥80, falling was the most often physician-cited reason for not prescribing warfarin (41%) followed by hemorrhage (28%).

Conclusion—Our findings suggest that many elderly patients at high risk for stroke may not be optimal candidates for anticoagulant therapy. There is a pressing need for alternative stroke prevention strategies for this expanding patient population. (***Stroke.* 2006;37:1075-1080.)**

Key Words: atrial fibrillation ■ geriatrics ■ warfarin

Approximately 25% of strokes in patients age ≥80 years are attributable to atrial fibrillation (AF).[1] AF–related stroke is associated with a 30-day mortality of ≈24%.[2–5] The prevalence of AF increases with age approaching 10% for individuals age ≥80 years.[6,7] It is projected that 3.3 million adults in the United States will have AF by the year 2020 largely attributable to the aging of our population.[8]

Warfarin has been shown to be highly effective in preventing stroke in AF.[9,10] Despite its proven benefit, numerous studies have documented under use of warfarin particularly among elderly patients who would seem to benefit the most from anticoagulant therapy.[11–19] These findings have raised concerns regarding quality of care, physician adherence to guidelines, and translation of clinical trial results into real-world practice.[20,21]

An important unanswered question is whether published estimates of warfarin use represent inappropriate under-prescribing or withholding of therapy attributable to questionable patient candidacy for a class of drugs with associated toxicity. Estimates of warfarin use vary depending on study methodology. Existing data, based largely on retrospective studies or secondary analyses, may overestimate the proportion of patients eligible for anticoagulation because of incomplete identification of contraindications. Resolution of this issue is central to the care of aging patients. Published low rates of warfarin use have importantly focused attention on the need to better inform physicians and patients on the benefits and risks of anticoagulant therapy. However, if the low rate of warfarin use represents a growing and problematic group of patients in whom anticoagulant therapy is felt not to be safe, a critical problem in stroke prevention is emerging. Patients at the highest risk of stroke currently have little in the way of an effective alternative.[22,23] Given the proven benefit

Received November 11, 2005; final revision received January 20, 2006; accepted January 23, 2006.

From the Department of Medicine (E.M.H., L.E.H.), Research Unit-Section of General Internal Medicine, Boston University School of Medicine, Mass; the Department of Medicine (J.D.), Cardiology Division, University of Pittsburgh, Pa; the Department of Medicine (C.E.-M.), University of Virginia, Charlottesville, Va; and the Department of Medicine (C.S., S.R.), Massachusetts General Hospital, Boston, Mass.

Correspondence and Reprint Requests to Elaine M. Hylek, MD, MPH, Boston University School of Medicine, Research Unit-Section of General Internal Medicine, 91 East Concord Street, Suite 200, Boston, MA 02118. E-mail ehylek@bu.edu

***Stroke* is available at http://www.strokeaha.org** **DOI: 10.1161/01.STR.0000209239.71702.ce**

of anticoagulant therapy, it is important that we understand the factors that are driving the discrepancy between published recommendations and actual practice. Direct patient-level assessment is needed to better inform strategies aimed at optimizing stroke prevention in this vulnerable age group.

The objective of our study was to prospectively assess the use of antithrombotic therapy among a contemporary cohort of patients with AF at the time of hospital discharge. In addition to baseline characteristics, we also sought to determine the physician-cited reason for not prescribing warfarin for each patient.

Methods

Study Participants

Consecutive patients were prospectively identified by daily searches of electronic admission notes and ECGs of all patients admitted to the medical service of our hospital from January 2001 to June 2003. Searches intentionally did not include patients admitted to the surgical service or stroke unit. To be eligible, patients had to be ≥65 years of age, had to have AF verified by ECG, could not be taking warfarin on admission or have another indication for long-term therapy, and had to have their longitudinal care provided by a physician at our institution. The eligibility criterion related to site of care minimized referral bias of more complex procedure-oriented patients and ensured that the anticoagulation decision would not be deferred at the time of discharge to the patient's outside physician. It also ensured access to the comprehensive ambulatory and hospital medical records. In addition, the presence of an on-site anticoagulation clinic helped to minimize any logistical barriers to warfarin prescription.

Baseline Patient Characteristics

Demographic data and medical diagnoses were extracted from the medical record. Major risk factors for stroke included age ≥75 years, diabetes mellitus, heart failure, hypertension, prior stroke, and age ≥65 years with coronary artery disease.[24–26] Risk factors for hemorrhage and potential contraindications to warfarin included history of hemorrhage, falls, cognitive impairment or dementia, active alcohol abuse, liver disease, advanced malignancy, and renal dysfunction defined as creatinine concentration >133 μmol/L or 1.5 mg/dL.[27] To better discern absolute versus relative contraindication, hemorrhage was further defined as (1) intracranial, (2) recurrent, (3) related to the index hospitalization, or (4) warfarin-related. Similarly, prior falls were further characterized by the presence or absence of associated head trauma or fracture. Past history of warfarin use and reason for discontinuation were also recorded.

TABLE 1. Baseline Characteristics of Hospitalized Patients With AF, Stratified by Warfarin Use

Characteristic	All Patients, n=405	Warfarin, n=206	No Warfarin, n=199	*P* Value
Age, y, mean	80.0	78.0	81.7	<0.001
Women, n (%)	234 (58)	108 (52)	126 (63)	0.027
Coronary heart disease, n (%)	193 (48)	94 (46)	99 (50)	0.409
AF, n (%)				<0.001
First clinical episode	174 (43)	123 (60)	51 (26)	
Recurrent episode	136 (34)	73 (35)	63 (32)	
Persistent	95 (23)	10 (5)	85 (42)	
Risk factors for stroke, n (%)				
Age ≥75 years	299 (74)	135 (66)	164 (82)	<0.001
Heart failure	192 (47)	97 (47)	95 (48)	0.976
Hypertension	299 (74)	150 (73)	149 (75)	0.642
Diabetes mellitus	109 (27)	59 (29)	50 (25)	0.491
Prior stroke	49 (12)	18 (9)	31 (16)	0.049
Risk factors for stroke, n (%)	0.156			
No major	13 (3)	7 (3)	6 (3)	
1	74 (18)	46 (22)	28 (14)	
2	141 (35)	73 (36)	68 (34)	
≥3	177 (44)	80 (39)	97 (49)	
Risk factors for hemorrhage, n (%)				
Prior hemorrhage	97 (24)	34 (16)	63 (32)	<0.001
Prior fall	100 (25)	24 (12)	76 (38)	<0.001
Cognitive impairment/dementia	51 (13)	8 (4)	43 (22)	<0.001
Active alcohol abuse	9 (2)	4 (2)	5 (3)	0.697
Advanced malignancy	25 (6)	8 (4)	17 (9)	0.051
Renal dysfunction*	101 (25)	43 (21)	58 (29)	0.051
Liver disease	11 (3)	3 (1)	8 (4)	0.113
Length of hospital stay, d, median	6	5	6	0.309

*Creatinine concentration >133 μmol/L or 1.5 mg/dL.

Discharge Antithrombotic Medication and Physician-Cited Reason for Not Prescribing Warfarin

Patients were followed from hospital admission to the day of discharge. Prescription of antithrombotic therapy was ascertained from 2 sources: the discharge summary and the electronic discharge medication list. The reason for not prescribing warfarin was recorded as explicitly stated in the discharge summary or, if ambiguous, as clarified by the treating physician on inquiry by the study nurse.

The study was approved by the institutional review board at Massachusetts General Hospital (Boston, Mass). The nature of the study did not require written informed consent.

Statistical Analysis

Baseline characteristics of patients according to warfarin status were compared using χ^2 test for proportions for categorical variables and Student t test for continuous variables. Independent predictors of warfarin use were determined using multivariate logistic regression. Variables whose probability values were ≤0.10 on univariate analysis were included in the multivariate model. For all analyses, a 2-sided probability value <0.05 was considered statistically significant. For the subset of patients who were not discharged on warfarin, we also sought to define the patient features associated with prescription of either low dose aspirin, 81 mg per day, or no antithrombotic therapy compared with those patients discharged on full dose aspirin, 325 mg per day. Five patients with missing data on aspirin dose were excluded from this analysis. Analyses were performed using Stata statistical software, release 8.0 (Stata Corporation).

Results

Patient Clinical Characteristics

During the study period, 426 patients were identified with ECG–verified AF, were ≥65 years of age, were not taking warfarin on admission, and had an established physician at our institution. Twenty-one patients died before discharge. Of the 405 patients, 74% were age ≥75 years (51% 80 years or greater), 74% had hypertension, 47% heart failure, and 12% prior stroke (Table 1). Ninety-seven percent of patients had ≥1 major risk factor for stroke and 79% had ≥2. Symptoms related to an uncontrolled ventricular rate prompted admission for the majority of patients.

Warfarin Use

Of the 405 patients, 51% (n=206) were started on warfarin: 75% of those 65 to 69 years of age, 59% 70 to 79, 45% 80 to 89, and 24% age ≥90 ($P<0.001$). Seventy-five percent of patients starting warfarin had ≥2 major stroke risk factors versus 83% of those patients not taking warfarin. Patients starting warfarin were younger, more likely to be male, newly presenting with AF, and less likely to have a history of hemorrhage, falls, or cognitive impairment. Twenty-two percent of patients not started on warfarin had at least 1 documented fall that resulted in head trauma or fracture compared with 4% of those patients who were discharged on warfarin. Of the 199 patients not started on warfarin, 7% (n=14) were admitted from a nursing home and a total of 13% (n=26) were discharged to a long-term care facility. Older age, cognitive impairment, history of falling, history of hemorrhage, and advanced malignancy were identified as independent baseline predictors of not receiving warfarin at hospital discharge (Table 2).

TABLE 2. Independent Predictors of Not Receiving Warfarin Among Hospitalized Patients With AF

Variable	Adjusted Odds Ratio (95% CI)	P Value
Age, per decade	1.49 (1.08–2.06)	0.015
Female	1.56 (0.98–2.47)	0.059
Cognitive impairment	6.35 (2.74–14.74)	<0.001
History of falling	5.61 (2.50–12.57)	<0.001
History of hemorrhage	3.22 (1.87–5.56)	<0.001
History of ischemic stroke	1.19 (0.59–2.40)	0.63
Advanced malignancy	4.57 (1.82–11.46)	0.001
Renal dysfunction	1.45 (0.86–2.45)	0.16

Physician-Cited Reason for Not Prescribing Warfarin

Potential contraindications were cited for 98% (n=195) of the 199 patients who were not started on warfarin including 46 patients (23%) who were unable to tolerate warfarin in the past (6 patients had sustained an intracranial hemorrhage and 21 patients extracranial bleeding [Table 3]). Of the 199 patients, the most often physician-cited reasons were hemorrhage for 33% (n=66), falls for 32% (n=64), and patient refusal or history of nonadherence for 14% (n=27). To better validate hemorrhage, we further characterized the bleeding experienced by those patients with hemorrhage cited as the reason warfarin was not prescribed. Of the 66 patients, 31 had recurrent bleeding, 9 had a prior intracranial hemorrhage, 16 had evidence of bleeding during the index hospitalization, and 7 of the remaining 10 patients had sustained a warfarin-related hemorrhage in the past. Among patients with falls cited as the contraindication, 53% had at least 1 documented spontaneous fall that resulted in closed head trauma or an orthopedic fracture. Among patients age ≥80 years, falling

TABLE 3. Physician-Cited Reason for Not Prescribing Warfarin, Stratified by Patient Age

Reason	All, n=199*	<80 Years, n=76	≥80 Years, n=123
Hemorrhage, n (%)	66 (33)	32 (42)	34 (28)
Recurrent bleeding	31 (16)	17 (22)	14 (11)
Current bleeding	16 (8)	7 (9)	9 (7)
Past intracranial bleeding	9 (4)	3 (4)	6 (5)
Past other bleeding	10 (5)	5 (7)	5 (4)
Falls	64 (32)	14 (18)	50 (41)
Patient refused or history of nonadherence	27 (14)	13 (17)	14 (11)
Cognitive impairment	6 (3)	1 (1)	5 (4)
Active alcohol abuse	4 (2)	4 (5)	0
Advanced illness, comfort care	16 (8)	4 (5)	12 (10)
Other†	16 (8)	8 (11)	8 (7)

*Includes 46 (23%) patients who had been taken off warfarin in the past. Twenty-seven patients had sustained a warfarin-related hemorrhage (6 intracranial, 21 extracranial), and 19 patients had their warfarin discontinued because of falls (n=9), nonadherence (n=8), and other (n=2).

†Includes 2 patients with an intracranial mass, 2 labile hypertension, 2 previous hypersensitivity reaction, 5 anticipated procedures, and major surgery, 4 normal sinus rhythm after cardioversion, 1 pericarditis.

TABLE 4. Choice of Antithrombotic Therapy Stratified by Physician-Cited Warfarin Contraindication

		Antithrombotic Medication				
		None	Aspirin	Clopidogrel+Aspirin	Aspirin Dose†	
Potential Contraindication	n	n	n	n	81 mg %	325 mg %
Hemorrhage, extracranial	56	20	32	4*	58%	42%
Hemorrhage, intracranial	10	3	4	3	57	43
Prior falls	64	6	52	6*	37	63
Patient refusal	27	3	21	3	22	78
Cognitive impairment	6	2	4	0	50	50
Active alcohol abuse	4	2	2	0	100	0
Terminal illness	16	6	10	0	33	67
Other	16	2	12	2	0	100
Total	199	44	137	18	38%	62%

*Includes 1 patient taking clopidogrel without aspirin.
†Five patients are missing aspirin dose.

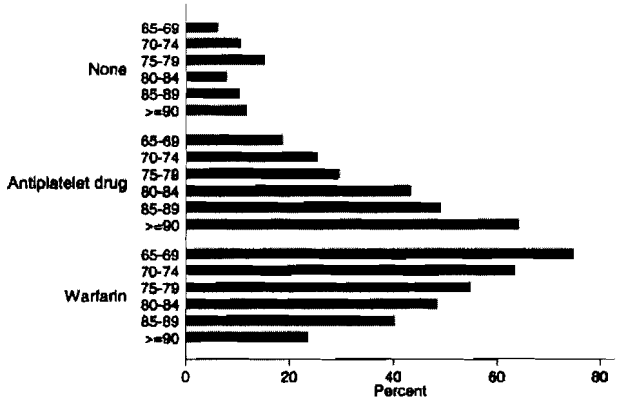
was the most often-cited reason for not prescribing warfarin (41%) followed by hemorrhage (28%).

Alternative Antithrombotic Therapy

Overall, 44 patients (11%) were not discharged on any antithrombotic medication. Thirty-eight percent (n=155) of patients were discharged on antiplatelet therapy. We stratified discharge antithrombotic therapy and prescribed dose of aspirin by the physician-cited warfarin contraindication to gain insight into physician choice of alternative medication (Table 4). Among patients prescribed aspirin, 62% were taking 325 mg per day and 38% 81 mg. Hemorrhage was the only independent predictor of receiving either no antithrombotic therapy or low-dose aspirin versus full-dose aspirin at hospital discharge (adjusted odds ratio, 5.4 [95%CI, 2.1 to 13.7]). After controlling for bleeding complications, older patients were not more likely to be prescribed low-dose aspirin or no antithrombotic medication (adjusted odds ratio, 1.1 [95% CI, 0.7 to 1.6]). The breakdown of antithrombotic therapy at hospital discharge by age is shown in the Figure.

Discussion

AF is a powerful risk factor for stroke. Stroke is a leading cause of death and disability. Warfarin has been shown to reduce the risk of stroke in AF by 68% compared with placebo, which is far superior to the relative risk reduction of 21% associated with full-dose aspirin.[22] Despite its proven efficacy, numerous studies have documented underutilization of warfarin, most notably in the elderly. In our study among 405 patients identified with AF, 51% (n=206) were discharged on warfarin. Of the remaining 199 patients, 97% had ≥1 major risk factor for stroke and 83% had ≥2. All of these patients had been considered for anticoagulant therapy, reflecting an understanding of the stroke risk imparted by AF and current evidence-based guidelines.[24,26] However, nearly all of these patients were felt to have contraindications that precluded its safe use including nearly 25% of patients who had taken warfarin in the past, but were unable to tolerate it long-term. Hemorrhage, falls, and patient refusal or history of nonadherence to therapy constituted nearly 80% of the physician-cited reasons for not prescribing warfarin. Currently, there is no effective treatment option for these patients. Despite its modest effect, aspirin is widely used in this setting.



Antithrombotic therapy at hospital discharge by patient age.

Our study is the first prospective assessment of warfarin prescription at the time of the treatment decision among consecutive patients admitted and discharged with AF. In addition to baseline patient characteristics, we directly determined the reason for not prescribing warfarin for each patient. We independently validated the physician-cited contraindications of hemorrhage and falls and qualified their severity. Retrospective studies and studies based on secondary analysis of administrative databases have likely underreported contraindications and have overestimated warfarin eligibility among hospitalized patients of age ≥80 years. Accurate assessment of warfarin candidacy using these methodologies is limited by the extent of documentation and reliance on search strategies using ICD-9-CM codes (*International Classification of Diseases, Ninth Edition, Clinical Modification*) that would not comprehensively capture patient-specific clinical details. Most notably, recurrent nonhospitalized falls, previous attempts at warfarin use, recurrent bleeding, patient preference, patient nonadherence, excessive alcohol use, and comfort care status would not be systematically ascertained.

Our study illustrates the challenges that physicians and patients face when trying to apply recommendations derived from clinical trials to older patients with AF in real-world practice. Few patients >80 years were enrolled in the early trials of oral anticoagulation versus placebo. Two meta-analyses of these trials report 20% of the study population being ≥75 years of age. This is in contrast to the patients enrolled in our study of which 51% were ≥80 years of age (74% age ≥75 years). Two more recently conducted trials that compared warfarin to ximelagatran enrolled 16% of patients age ≥80 years (38% age ≥75). However, unlike the early AF trials that compared warfarin to placebo, 74% and 84% of patients, respectively, were taking a vitamin K antagonist at entry, thereby selecting populations at already proven lower risk for hemorrhage. Six percent of patients had a prior history of bleeding compared with 24% in this study.[28–30]

A limitation of our study is that it was conducted at a single center: a large, urban, teaching hospital which may not reflect other practice settings. However, our finding that 51% of patients with AF were discharged on warfarin is consistent with a study funded by the Health Care Financing Administration assessing the quality of care delivered to Medicare beneficiaries during the period 1998 to 1999.[20] Using hospital claims data, performance in the median state was 55% and ranged from 42% to 65%. A more recent retrospective study of 21 teaching, 13 community, and 4 Veterans Administration hospitals similarly found that 55% of patients with AF were discharged on warfarin in 2002 (53% among the teaching hospitals).[31] Another potential limitation of our study is that physicians who were directly queried may have altered their subsequent prescribing behavior. However, we would expect this bias to have resulted in more warfarin use. In addition, the reason for not prescribing warfarin was explicitly stated in the discharge summary for the majority of patients.

The study institution has several features that overcome limitations of other settings. The existence of an anticoagulation clinic minimized monitoring as a potential barrier to warfarin prescription. In addition, stroke awareness is high among physician staff, many of whom participated in the Boston Area Anticoagulation Trial for Atrial Fibrillation which helped to establish the efficacy of warfarin.[32] Enrolled patients had established providers at our institution, which facilitated validation and severity assessment of contraindications in the outpatient and inpatient medical records, and ensured that the warfarin decision was not deferred to an outside provider. Because we did not include referral patients, we believe our study population is similar to that of other large, general, urban hospitals serving a Medicare population.

Our data suggest that among very elderly patients with AF the decision to prescribe warfarin is strongly influenced by contraindications. More research is needed to discern absolute from relative contraindications to help guide physicians and patients in their assessment of the overall risks and benefits of therapy.[33–35] Strategies are needed to optimize candidacy among elderly patients, particularly interventions to reduce hemorrhage and falls.[36–38] Prospective studies are needed to better define the true hazard of falls in the presence of anticoagulant therapy. In addition, our study underscores the need for continued research on the mechanisms of AF and triggers for thrombus formation, insights which may lead to alternative stroke prevention strategies without the attendant hemorrhagic risk.[39–41]

Summary

Despite the first published evidence of warfarin's efficacy 16 years ago, prophylaxis against stroke in AF remains suboptimal particularly among those at highest risk, patients of age ≥80 years. Given the aging of the population, stroke prevention in AF is a pressing health imperative.

Acknowledgments

This study was funded by The Robert Wood Johnson Foundation Generalist Physician Faculty Scholars Program (RWJF 039174). Dr Hylek has received research funding from Bristol-Myers Squibb and 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. Candelise L, Pinardi G, Morabito A. Mortality in acute stroke with atrial fibrillation. The Italian Acute Stroke study group. *Stroke*. 1991;22:169–174.
3. Hylek EM, Go AS, Chang Y, Jensvold NG, Henault LE, Selby JV, Singer DE. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med*. 2003;349:1019–1026.
4. Lin H-J, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, D'Agostino RB. Stroke severity in atrial fibrillation: The Framingham study. *Stroke*. 1996;27:1760–1764.
5. Sandercock P, Bamford J, Dennis M, Burn J, Slattery J, Jones L, Boonyakarnkul S, Warlow C. Atrial fibrillation and stroke: Prevalence in different types of stroke and influence on early and long term prognosis (Oxfordshire Community Stroke Project). *BMJ*. 1992;305:1460–1465.
6. Feinberg WM, Blackshear JL, Laupacis A, Kronmal R, Hart RG. Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications. *Arch Intern Med*. 1995;155:469–473.
7. Furberg CD, Psaty BM, Manolio TA, Gardin JM, Smith VE, Rautaharju PM. Prevalence of atrial fibrillation in elderly subjects (The Cardiovascular Health Study). *The Am J Cardiol*. 1994;74:236–241.
8. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, Singer DE. 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.
9. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med*. 1999;131:492–501.
10. 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.
11. Albers GW, Yim JM, Belew KM, Bittar N, Hattemer CR, Phillips BG, Kemp S, Hall EA, Morton DJ, Vlasses PH. Status of antithrombotic therapy for patients with atrial fibrillation in university hospitals. *Arch Intern Med*. 1996;156:2311–2316.
12. Brass LM, Krumholz HM, Scinto JM, Radford M. Warfarin use among patients with atrial fibrillation. *Stroke*. 1997;28:2382–2389.
13. Bungard TJ, Ghali WA, Teo KK, McAlister FA, Tsuyuki RT. Why do patients with atrial fibrillation not receive warfarin? *Arch Intern Med*. 2000;160:41–46.
14. Cohen N, Almoznino-Sarafian D, Alon I, Gorelik O, Koopfer M, Chachashvily S, Shteinshnaider M, Litvinjuk V, Modai D. Warfarin for stroke prevention still underused in atrial fibrillation: patterns of omission. *Stroke*. 2000;31:1217–1222.
15. Gage BF, Boechler M, Doggette AL, Fortune G, Flaker GC, Rich MW, Radford MJ. Adverse outcomes and predictors of underuse of antithrombotic therapy in medicare beneficiaries with chronic atrial fibrillation. *Stroke*. 2000;31:822–827.
16. Go AS, Hylek EM, Borowsky LH, Phillips KA, Selby JV, Singer DE. Warfarin use among ambulatory patients with nonvalvular atrial fibril-

lation: The Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. *Ann Intern Med.* 1999;131:927–934.
17. Smith NL, Psaty BM, Furberg CD, White R, Lima JAC, Newman AB, Manolio TA. Temporal trends in the use of anticoagulants among older adults with atrial fibrillation. *Arch Intern Med.* 1999;159:1574–1578.
18. Stafford RS, Singer DE. Recent national patterns of warfarin use in atrial fibrillation. *Circulation.* 1998;97:1231–1233.
19. White RH, McBurnie MA, Manolio T, Furberg CD, Gardin JM, Kittner SJ, Bovill E, Knepper L. Oral anticoagulation in patients with atrial fibrillation: adherence with guidelines in an elderly cohort. *The Am J Med.* 1999;106:165–171.
20. Jencks SF, Cuerdon T, Burwen DR, Fleming B, Houck PM, Kussmaul AE, Nilasena DS, Ordin DL, Arday DR. Quality of medical care delivered to medicare beneficiaries: a profile at state and national levels. *JAMA.* 2000;284:1670–1676.
21. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the united states. *N Engl J Med.* 2003;348:2635–2645.
22. Investigators AF. The efficacy of aspirin in patients with atrial fibrillation. Analysis of pooled data from 3 randomized trials. *Arch Intern Med.* 1997;157:1237–1240.
23. Stroke Prevention in Atrial Fibrillation I. Adjusted-dose warfarin versus low-intensity, fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation: Stroke prevention in atrial fibrillation iii randomised clinical trial. *The Lancet.* 1996;348:633–638.
24. Fuster V, Ryden LE, Asinger RW, Cannom DS, Crijns HJ, Frye RL, Halperin JL, Kay GN, Klein WW, Levy S. For the american college of cardiology/american heart association/european society of cardiology: Guidelines for the management of patients with atrial fibrillation. *J Am Coll Cardiol.* 2001;38:1231–1265.
25. 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.
26. 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.
27. Beyth RJ, Quinn LM, Landefeld CS. Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin. *The Am J Med.* 1998;105:91–99.
28. Executive Steering Committee on behalf of the SPORTIF III. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): Randomised controlled trial. *The Lancet.* 2003;362:1691–1698.
29. Halperin JL. Ximelagatran compared with warfarin for prevention of thromboembolism in patients with nonvalvular atrial fibrillation: Rationale, objectives, and design of a pair of clinical studies and baseline patient characteristics (SPORTIF III and V). *Am Heart J.* 2003;146: 431–438.
30. Investigators SV. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: A randomized trial. *JAMA.* 2005;293:690–698.
31. Waldo AL, Becker RC, Tapson VF, Colgan KJ; for the NABOR Steering Committee. Hospitalized patients with atrial fibrillation and a high risk of stroke are not being provided with adequate anticoagulation. *J Am Coll Cardiol.* 2005;46:1729–1736.
32. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. The Boston Area Anticoagulation Trial for Atrial Fibrillation investigators. *N Engl J Med.* 1990;323:1505–1511.
33. Gage BF, Birman-Deych E, Kerzner R, Radford MJ, Nilasena DS, Rich MW. Incidence of intracranial hemorrhage in patients with atrial fibrillation who are prone to fall. *The Am J Med.* 2005;118:612–617.
34. Hart RG, Halperin JL, Pearce LA, Anderson DC, Kronmal RA, McBride R, Nasco E, Sherman DG, Talbert RL, Marler JR. Lessons from the stroke prevention in atrial fibrillation trials. *Ann Intern Med.* 2003;138:831–838.
35. Mina AA, Bair HA, Howells GA, Bendick PJ. Complications of preinjury warfarin use in the trauma patient. *J Trauma-Injury Infection & Critical Care.* 2003;54:842–847.
36. Dhesi JK, Jackson SHD, Bearne LM, Moniz C, Hurley MV, Swift CG, Allain TJ. Vitamin D supplementation improves neuromuscular function in older people who fall. *Age Ageing.* 2004;33:589–595.
37. Priplata AA, Niemi JB, Harry JD, Lipsitz LA, Collins JJ. Vibrating insoles and balance control in elderly people. *The Lancet.* 2003;362: 1123–1124.
38. Tinetti ME. Preventing falls in elderly persons. *N Engl J Med.* 2003;348: 42–49.
39. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, Tracy RP, Van Wagoner DR, Psaty BM, Lauer MS, Chung MK. Inflammation as a risk factor for atrial fibrillation. *Circulation.* 2003;108:3006–3010.
40. Li D, Shinagawa K, Pang L, Leung TK, Cardin S, Wang Z, Nattel S. Effects of angiotensin-converting enzyme inhibition on the development of the atrial fibrillation substrate in dogs with ventricular tachypacing-induced congestive heart failure. *Circulation.* 2001;104:2608–2614.
41. Sievert H, Lesh MD, Trepels T, Omran H, Bartorelli A, Della Bella P, Nakai T, Reisman M, DiMario C, Block P, Kramer P, Fleschenberg D, Krumsdorf U, Scherer D. Percutaneous left atrial appendage transcatheter occlusion to prevent stroke in high-risk patients with atrial fibrillation: Early clinical experience. *Circulation.* 2002;105:1887–1889.

SPECIAL ARTICLE

The "Gender Gap" in Authorship of Academic Medical Literature — A 35-Year Perspective

Reshma Jagsi, M.D., D.Phil., Elizabeth A. Guancial, M.D., Cynthia Cooper Worobey, M.D., Lori E. Henault, M.P.H., Yuchiao Chang, Ph.D., Rebecca Starr, M.B.A., M.S.W., Nancy J. Tarbell, M.D., and Elaine M. Hylek, M.D., M.P.H.

ABSTRACT

BACKGROUND

Participation of women in the medical profession has increased during the past four decades, but issues of concern persist regarding disparities between the sexes in academic medicine. Advancement is largely driven by peer-reviewed original research, so we sought to determine the representation of female physician-investigators among the authors of selected publications during the past 35 years.

METHODS

Original articles from six prominent medical journals — the *New England Journal of Medicine* (*NEJM*), the *Journal of the American Medical Association* (*JAMA*), the *Annals of Internal Medicine* (*Ann Intern Med*), the *Annals of Surgery* (*Ann Surg*), *Obstetrics & Gynecology* (*Obstet Gynecol*), and the *Journal of Pediatrics* (*J Pediatr*) — were categorized according to the sex of both the first and the senior (last listed) author. Sex was also determined for the authors of guest editorials in *NEJM* and *JAMA*. Data were collected for the years 1970, 1980, 1990, 2000, and 2004. The analysis was restricted to authors from U.S. institutions holding M.D. degrees.

RESULTS

The sex was determined for 98.5 percent of the 7249 U.S. authors of original research with M.D. degrees. The proportion of first authors who were women increased from 5.9 percent in 1970 to 29.3 percent in 2004 (P<0.001), and the proportion of senior authors who were women increased from 3.7 percent to 19.3 percent (P<0.001) during the same period. The proportion of authors who were women increased most sharply in *Obstet Gynecol* (from 6.7 percent of first authors and 6.8 percent of senior authors in 1970 to 40.7 percent of first authors and 28.0 percent of senior authors in 2004) and *J Pediatr* (from 15.0 percent of first authors and 4.3 percent of senior authors in 1970 to 38.9 percent of first authors and 38.0 percent of senior authors in 2004) and remained low in *Ann Surg* (from 2.3 percent of first authors and 0.7 percent of senior authors in 1970 to 16.7 percent of first authors and 6.7 percent of senior authors in 2004). In 2004, 11.4 percent of the authors of guest editorials in *NEJM* and 18.8 percent of the authors of guest editorials in *JAMA* were women.

CONCLUSIONS

Over the past four decades, the proportion of women among both first and senior physician-authors of original research in the United States has significantly increased. Nevertheless, women still compose a minority of the authors of original research and guest editorials in the journals studied.

From the Departments of Radiation Oncology (R.J., N.J.T.) and Nephrology (C.C.W.), the General Medicine Division (Y.C.), and the Office of Women's Careers (R.S., N.J.T.), Massachusetts General Hospital; Harvard Medical School (R.J., E.A.G., C.C.W., Y.C., N.J.T.); and the Section of General Internal Medicine–Research Unit, Boston Medical Center, Boston University School of Medicine (L.E.H., E.M.H.) — all in Boston. Address reprint requests to Dr. Jagsi at the Office for Women's Careers, Bulfinch 370, Massachusetts General Hospital, 55 Fruit St., Boston, MA 02114, or at reshma_jagsi@post.harvard.edu.

N Engl J Med 2006;355:281-7.

During the past four decades, the participation of women in medicine has increased dramatically. Women now represent 49 percent of all medical students,[1] as compared with 6 percent in 1960.[2] Overall, 25 percent of practicing physicians in the United States are women,[3] and women now make up 32 percent of full-time medical faculty members.[4] However, there is considerable evidence that women continue to be underrepresented in the top tiers of academic medicine.[5-7] Women currently make up 10 percent of medical school deans, 11 percent of department chairs, and 14 percent of full professors among the clinical faculty in medical schools.[4] Women last composed 14 percent of all medical students in 1972.[8] In addition, only 10 percent of female clinical faculty members as compared with 28 percent of male clinical faculty members are full professors.[4] Figure 1 depicts the number of female faculty members who served as professors and role models for both male and female residents in the main medical specialties in 2004. For example, in internal medicine, the ratio of residents to female professors was 31 to 1; this ratio was 44 to 1 with the inclusion of fellows.[9,10]

Publication in medical journals is an important measure of academic productivity. It is also highly emphasized in the academic promotion process and an important means by which the academic medical community communicates. Although several survey studies have suggested that female faculty members may be less likely to publish academic papers than their male colleagues,[11,12] other studies have not found apparent differences.[13-15] Few studies have attempted to quantify the sex distribution of authors of published research, and those that have done so have focused on the fields of otolaryngology[16] and epidemiology[17] or on authors of research published in journals outside the United States.[18,19] In this study, we examined whether there was a "gender gap" in the authorship of six prestigious medical journals in the United States and we sought to quantify its magnitude. In addition, we examined the patterns of change in this gap over time and variations according to specialty area. We focused on published original research in these journals from 1970 to the present. We also assessed the sex composition of authors of guest editorials published during the same period.

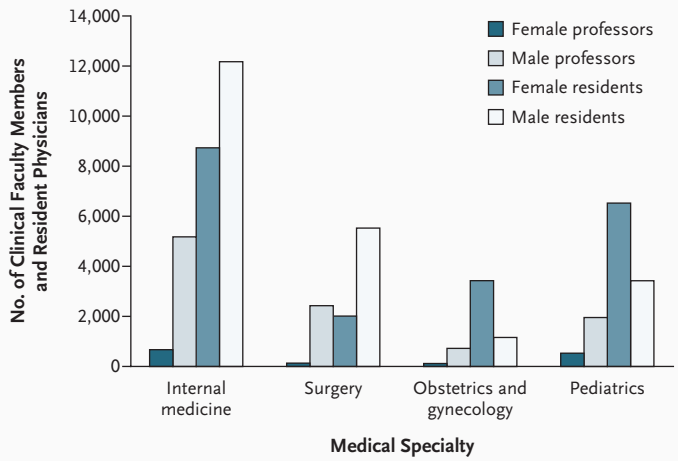


Figure 1. Sex Distribution of Clinical Faculty Members and Resident Physicians in Medical Specialties, 2004.

Data from the Association of American Medical Colleges indicate that a relatively small absolute number of female faculty members serve as professors and role models for the large number of both male and female residents in the main medical specialties.

METHODS

DATA COLLECTION

We focused on the four medical specialties that have traditionally constituted the core clerkships in the education of medical students. These specialties, which together include the largest proportion of practicing physicians, include internal medicine, surgery, pediatrics, and obstetrics and gynecology. Journals were selected on the basis of "impact factors,"[20,21] citation half-life,[20] and comments solicited from faculty members regarding the long-term prestige and importance of the various journals in their fields. Six prominent medical journals published in the United States were included in this study: the *New England Journal of Medicine* (*NEJM*), the *Journal of the American Medical Association* (*JAMA*), the *Annals of Internal Medicine* (*Ann Intern Med*), the *Annals of Surgery* (*Ann Surg*), *Obstetrics & Gynecology* (*Obstet Gynecol*), and the *Journal of Pediatrics* (*J Pediatr*).

All original articles published in 1970, 1980, 1990, 2000, and 2004 were included in the data set. For each of these articles, we determined both the first and senior (last listed) authors' sex, graduate degrees, and institutional affiliation. An author's sex was determined by initial inspection of his or her first name. For cases in which an author's sex was not certain, attempts were made to discern the sex by visiting the institutional Web

site and performing Internet searches with the use of the Google search engine.

Also included and separately identified in the study were guest editorials in the two nonspecialty journals, *NEJM* and *JAMA*. Only editorials authored by persons other than editorial-board members were considered for analysis. In the rare cases in which editorials were written by more than two authors, our analysis included just the first and last authors.

STATISTICAL ANALYSIS

Our analysis was restricted to investigators from U.S. institutions who held an M.D. degree. The tabulated data were stored in a Microsoft Access database and analyzed (with the use of SAS software, version 9.1) to determine the sex distributions of the first and senior (last listed) authors of original articles for each journal and the sex distributions of authors of guest editorials in *NEJM* and *JAMA*. The Cochran–Armitage trend test was used to test for the trend over time. Reported P values pertain to the significance of trends over time in these data.

RESULTS

AUTHORSHIP OF ORIGINAL RESEARCH

A total of 7249 authors of original articles who held M.D. degrees and were from U.S. institutions were identified in the six journals during the years studied; 3872 were first authors, and 3377 were senior authors. The sex of the author was determined for 98.5 percent. Overall, 15.9 percent of the first authors and 10.3 percent of the senior authors were women. An analysis of the data according to year demonstrated significant gains by female physician-investigators since 1970 (Fig. 2). The proportion of women serving as first authors of published original research in these journals increased from 5.9 percent to 29.3 percent, and the proportion of women serving as senior authors increased from 3.7 percent to 19.3 percent. The data also suggested that this momentum may be reaching a plateau.

Significant trends of increased female representation were evident for each of the six journals during the 35-year period (Table 1). The proportions of first and senior authors who were women increased most sharply in the specialty journals of obstetrics and pediatrics and remained low in the journal having to do with surgery. In

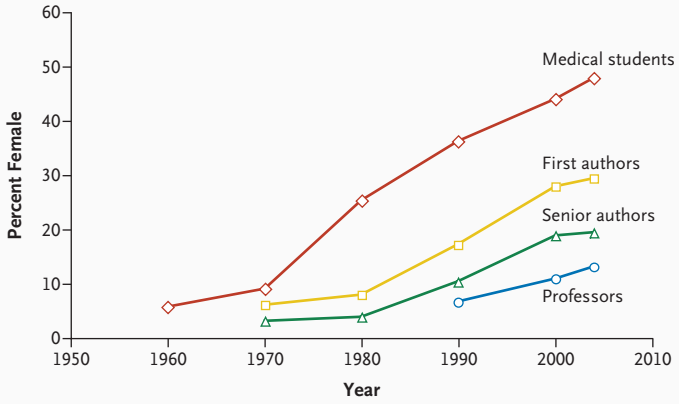


Figure 2. Female Physician-Investigators Who Were First and Senior Authors of Published Original Research.

In the six journals studied, the representation of women among first and senior authors of published original research increased during the past four decades. The cumulative trends over time are depicted by curves showing female representation among students enrolled in medical school and among professors on medical school faculties (data on faculty rank according to sex were not available from the Association of American Medical Colleges for 1980, 1970, or 1960).

2004, in the three general medical journals (*Ann Intern Med*, *NEJM*, and *JAMA*) collectively, female physicians made up 23.2 percent of the first authors and 12.7 percent of the senior authors of original research articles.

Of the U.S. authors in the years studied, 181 first authors held both M.D. and Ph.D. degrees and 236 senior authors held both M.D. and Ph.D. degrees. In this subgroup, the sex distribution over time was similar to that in the overall data set. Among the first authors holding both M.D. and Ph.D. degrees, 7.7 percent were female in 1970, 6.7 percent in 1980, 10.0 percent in 1990, 20.0 percent in 2000, and 17.9 percent in 2004 (P for trend=0.05). Of the senior authors holding both M.D and Ph.D. degrees, 3.7 percent were female in 1970, 0 percent in 1980, 12.5 percent in 1990, 22.9 percent in 2000, and 9.3 percent in 2004 (P for trend=0.02).

SEX OF AUTHORS OF GUEST EDITORIALS

We determined the sex of 99.6 percent of the 808 U.S. investigators with an M.D. degree who served as first or senior (last listed) authors of guest editorials in *NEJM* and *JAMA* during the years studied. During this period, of the 514 authors of guest editorials in *NEJM*, women made up 8.8 percent overall and 1.5 percent in 1970, 2.4 percent in 1980, 9.7 percent in 1990, 20.4 percent in 2000, and 11.4

Table 1. Representation of Female Physician-Investigators among First and Senior Authors of Published Original Research in Six U.S. Journals.*

Variable	1970	1980	1990	2000	2004	P Value
	number/total number (percent)					
Overall						
First author	58/982 (5.9)	67/810 (8.3)	137/814 (16.8)	169/614 (27.5)	178/607 (29.3)	<0.001
Senior author	29/783 (3.7)	25/692 (3.6)	69/681 (10.1)	106/578 (18.3)	112/580 (19.3)	<0.001
NEJM						
First author	8/188 (4.3)	14/117 (12.0)	23/143 (16.1)	23/110 (20.9)	13/92 (14.1)	<0.001
Senior author	6/153 (3.9)	3/108 (2.8)	11/122 (9.0)	13/106 (12.3)	11/97 (11.3)	<0.001
JAMA						
First author	13/227 (5.7)	7/151 (4.6)	25/125 (20.0)	26/121 (21.5)	30/113 (26.5)	<0.001
Senior author	5/173 (2.9)	3/128 (2.3)	13/102 (12.7)	19/115 (16.5)	16/118 (13.6)	<0.001
Ann Intern Med						
First author	5/107 (4.7)	8/126 (6.3)	13/106 (12.3)	15/44 (34.1)	17/54 (31.5)	<0.001
Senior author	5/93 (5.4)	4/115 (3.5)	4/92 (4.3)	11/43 (25.6)	7/52 (13.5)	0.009
Ann Surg						
First author	4/175 (2.3)	7/168 (4.2)	7/135 (5.2)	13/110 (11.8)	15/90 (16.7)	<0.001
Senior author	1/153 (0.7)	1/149 (0.7)	1/117 (0.9)	2/101 (2.0)	6/89 (6.7)	0.034
Obstet Gynecol						
First author	12/178 (6.7)	13/161 (8.1)	45/227 (19.8)	62/164 (37.8)	61/150 (40.7)	<0.001
Senior author	8/117 (6.8)	6/116 (5.2)	29/185 (15.7)	41/140 (29.3)	37/132 (28.0)	<0.001
J Pediatr						
First author	16/107 (15.0)	18/87 (20.7)	24/78 (30.8)	30/65 (46.2)	42/108 (38.9)	<0.001
Senior author	4/94 (4.3)	8/76 (10.5)	11/63 (17.5)	20/73 (27.4)	35/92 (38.0)	<0.001

* The analysis was restricted to authors from U.S. institutions holding an M.D. degree or equivalent for whom sex could be determined.

percent in 2004 (P for trend <0.001). Sex was determined for 291 of the 294 U.S. authors of guest editorials with M.D. degrees in *JAMA* during the years studied. Of these 291 authors, women made up 10 percent overall and 0 percent in 1970, 2.0 percent in 1980, 7.4 percent in 1990, 10.0 percent in 2000, and 18.8 percent in 2004 (P for trend <0.001).

DISCUSSION

Advancement in academic medicine is largely contingent on productivity and the measured external influence of one's scholarly work. Objective measures of the effect of one's work include the publication of original research in prominent journals and invitations by editors to provide scientific opinions on the published research of others. In this study, we focused on six medical journals chosen specifically for their prominence and high visibility to medical students, residents, and fellows. We found that from 1970 to 2004, the proportion of women among the U.S. physician-authors of original research in these journals increased from 5.9 percent to 29.3 percent of first authors and from 3.7 percent to 19.3 percent of senior authors. The magnitude of change for both groups was highest for *J Pediatr* and *Obstet Gynecol* and lowest for *Ann Surg;* these findings may have reflected, at least in part, the numbers of women entering these fields.

Despite these positive overall findings, the results also raise potential areas of concern. Although the proportion of women among authors has increased over time, the data suggest a possible lack of continued momentum among both first authors and senior authors in 2004 as compared with 2000. The data also suggest that a gender

Table 2. Academic Rank of Clinical Faculty in Main Specialties, According to Sex.*

Variable	1990	1995	2000	2005
		% *women*		
Overall				
Medical students	36	39	43	47
Instructors	**38**	44	46	—
Assistant professors	27	**32**	35	38
Associate professors	16	20	**23**	27
Professor	7	9	11	**14**
Internal Medicine				
Instructors	**33**	37	41	—
Assistant professor	23	**29**	33	36
Associate professor	11	17	**20**	24
Professor	5	6	9	**12**
Obstetrics and gynecology				
Instructor	**56**	65	66	—
Assistant professor	34	**43**	47	53
Associate professor	14	21	**27**	34
Professor	7	9	12	**16**
Pediatrics				
Instructor	**55**	60	64	—
Assistant professor	41	**48**	51	53
Associate professor	26	32	**34**	40
Professor	15	17	19	**22**
Surgery				
Instructor	**21**	29	31	—
Assistant professor	12	**15**	17	21
Associate professor	6	8	**9**	12
Professor	2	3	4	**6**

* Data are from the Association of American Medical Colleges.[4,8,22-24] Bold face values reflect a superimposed 20-year pipeline to full professor from an estimated year of graduation from medical school of 1985. Thirty percent of graduates from U.S. medical schools in 1985 were women. Data on instructor-level appointments according to sex were available in the Association of American Medical Colleges Data Books for 1990, 1995, and 2000, but not 2005.

gap in authorship remains, particularly among senior authors and editorial commentators.

Of the many possible explanations for our findings, one factor that probably explains at least some of the gender gap observed is that the pool of female faculty members who are eligible to serve as senior authors or editorial commentators remains limited. Nonnemaker examined the rates of academic advancement of men and women among different cohorts of U.S. medical school faculties from 1979 through 1997.[7] The study revealed that the numbers of women advancing to the ranks of associate and full professor were significantly lower than expected. Longitudinal data from the American Association of Medical Colleges seem to reaffirm this finding (Table 2).[4,8,22-24] In 2004, women made up only 19 percent of associate and full professors on the clinical faculties of medical schools.[9] The low overall percentage of female senior authors in 2004 — 19 percent in the six journals studied — may reflect this smaller pool of senior faculty members who are women. Similarly, the low percentage of women among authors of guest editorials may indicate that there is a limited pool of women who have achieved sufficient international recognition and expertise to

merit these invitations. Since the pool is limited, senior women may also be inundated with academic activities and may find it necessary to decline invitations more often, notwithstanding the potential for prestige and influence.

Several studies have explored the basis for the gender gap in academic medicine. In a study by Yedidia and Bickel,[25] three important barriers to the academic advancement of women were identified from interviews of department chairs — the constraints of traditional sex roles, manifestations of sexism in the medical environment, and lack of effective mentors. Carr et al. reported that female faculty members who had children published less and had less institutional support than did male colleagues who had children.[26] In addition, a study of female faculty members in the School of Science at the Massachusetts Institute of Technology found unanticipated patterns of inequity relative to the allocation of resources, space, salary, outside professional activities, and positions of influence.[27]

Some of the gap observed may also stem from career choices made by men and women. Studies have documented differences in career preferences between male and female medical students,[28,29] and women may devote more of their working time to teaching and clinical activity than to research.[30] Some have also speculated that women may have different priorities regarding the balance between work and other pursuits,[31-33] although recent studies have suggested that a balance between work and other activities is as important to men as it is to women, at least among younger physicians.[34-36] Ultimately, Nonnemaker found that fewer women were choosing academic career paths in the late 1990s.[7]

Finally, it has been suggested that the most productive period of women's careers is delayed, and this delay conflicts with traditional tenure clocks.[37] Strategies like the National Institutes of Health supplements to promote reentry into biomedical and behavioral research careers are grounded in the assumption that measures that help women to address these issues of timing may promote their more equal participation in academic medicine. On the basis of a similar logic, it may also be appropriate to consider making awards for career development independent of the number of years since medical school or since one's first faculty appointment.

Given the design of our study, we were unable to assess the contribution of productivity, career choice, or other possible factors to the gender gap in authorship in the journals we studied. Future research should explore these questions, since it is only through analysis of the underlying forces that this gap in academic medicine may be understood. Faculty diversity is valuable in promoting new insights into and approaches to medical research, so efforts to increase the representation of women in academic medicine should be grounded in rigorous, evidence-based analysis.

Our findings validate the perception that although women have made substantial strides in the past four decades, a gender gap remains among the authors of original articles in prestigious academic medical journals. Further investigation is necessary to understand more fully the causes for this gap, including the possibility that certain barriers may impede women's participation as authors early in their careers and in turn may diminish the pool of female senior faculty members who may serve in prominent authorship positions.

Presented in part at the Society for General Internal Medicine Annual Meeting, New Orleans, May 13, 2005.

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Women in U.S. academic medicine: statistics and medical school benchmarking 2004-2005. Washington, D.C.: Association of American Medical Colleges, 2005:Table 1.
2. Women in U.S. academic medicine: statistics and medical school benchmarking 2003-2004. Washington, D.C.: Association of American Medical Colleges, 2004:Table 1.
3. Physician characteristics and distribution in the U.S. Chicago: American Medical Association, 2004:Table 1.
4. Women in U.S. academic medicine: statistics and medical school benchmarking 2004-2005. Washington, D.C.: Association of American Medical Colleges, 2005: Table 3.
5. Bickel J. 1989. Women in medical education: a status report. N Engl J Med 1989; 319:1579-84.
6. Bickel J, Wara D, Atkinson BF, et al. Increasing women's leadership in academic medicine: report of the AAMC Project Implementation Committee. Acad Med 2002;77:1043-61.
7. Nonnemaker L. Women physicians in academic medicine: new insights from cohort studies. N Engl J Med 2000;342:399-405.
8. AAMC data book 2004. Washington, D.C.: Association of American Medical Colleges, 2004:Table B10.

9. Women in U.S. academic medicine: statistics and medical school benchmarking 2003-2004. Washington, D.C.: Association of American Medical Colleges, 2004: Table 3.
10. Women in U.S. academic medicine: statistics and medical school benchmarking 2004-2005. Washington, D.C.: Association of American Medical Colleges, 2005:Table 2.
11. Tesch BJ, Wood HM, Helwig AL, Nattinger AB. Promotion of women physicians in academic medicine: glass ceiling or sticky floor? JAMA 1995;273:1022-5.
12. Barnett RC, Carr P, Boisnier AD, et al. Relationships of gender and career motivation to medical faculty members' production of academic publications. Acad Med 1998;73:180-6.
13. Carr P, Friedman RH, Moskowitz MA, Kazis LE, Weed HG. Research, academic rank, and compensation of women and men faculty in academic general internal medicine. J Gen Intern Med 1992;7:418-23.
14. Carr PL, Friedman RH, Moskowitz MA, Kazis LE. Comparing the status of women and men in academic medicine. Ann Intern Med 1993;119:908-13.
15. Vydareny KH, Waldrop SM, Jackson VP, et al. Career advancement of men and women in academic radiology: is the playing field level? Acad Radiol 2000;7:493-501.
16. Bhattacharyya N, Shapiro NL. Increased female authorship in otolargyngology over the past three decades. Laryngoscope 2000; 110:358-61.
17. Dickersin K, Fredman L, Flegal KM, Scott JD, Crawley B. Is there a sex bias in choosing editors? Epidemiology journals as an example. JAMA 1998;280:260-4.
18. Schiaffino A, Garcia M, Fernandez E. Authorship and data reporting according to gender in four Spanish biomedical journals. Gac Sanit 2001;15:251-4. (In Spanish.)
19. Bastuji-Garin S, Broquet E, Revuz J. Publications of French dermatology professors from 1996 to 1998. Ann Dermatol Venereol 2002;129:1354-8. (In French.)
20. Thomson Institute for Scientific Information. Journal citation reports. Philadelphia: ISI Press, 2002.
21. Garfield E. The history and meaning of the journal impact factor. JAMA 2006; 295:90-3.
22. AAMC data book 1990. Washington, D.C.: Association of American Medical Colleges, 1990:Table C4.
23. AAMC data book 1995. Washington, D.C.: Association of American Medical Colleges, 1995:Table C5.
24. AAMC data book 2000. Washington, D.C.: Association of American Medical Colleges, 2000:Table C5.
25. Yedidia MJ, Bickel J. Why aren't there more women leaders in academic medicine? The views of clinical department chairs. Acad Med 2001;76:453-65.
26. 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;129:532-8.
27. A study on the status of women faculty in science at MIT. Cambridge, Mass.: The MIT Faculty Newsletter, March 1999. (Accessed June 23, 2006, at http://web.mit.edu/fnl/women/women.html.)
28. Cuca JM. The specialization and career preferences of women and men recently graduated from U.S. medical schools. J Am Med Womens Assoc 1979;34:425-35.
29. Bickel J, Ruffin A. Gender-associated differences in matriculating and graduating medical students. Acad Med 1995;70: 552-9.
30. Buckley LM, Sanders K, Shih M, Hampton CL. Attitudes of clinical faculty about career progress, career success and recognition, and commitment to academic medicine: results of a survey. Arch Intern Med 2000;160:2625-9.
31. 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. [Erratum, JAMA 2003;290:2666.]
32. Richtel M. Young doctors and wish lists: no weekend calls, no beepers. The New York Times. January 7, 2004:A1, A19.
33. Bland KI, Isaacs G. Contemporary trends in student selection of medical specialties: the potential impact on general surgery. Arch Surg 2002;137:259-67.
34. Dorsey ER, Jarjoura D, Rutecki GW. The influence of controllable lifestyle and sex on the specialty choices of graduating U.S. medical students, 1996-2003. Acad Med 2005;80:791-6.
35. Lambert EM, Holmboe ES. The relationship between specialty choice and gender of U.S. medical students, 1990-2003. Acad Med 2005;80:797-802.
36. Wendel TM, Godellas CV, Prinz RA. Are there gender differences in choosing a surgical career? Surgery 2003;134:591-6.
37. Grosz B, Dulac C, Dymecki S, et al. Report from the Task Force on Women in Science and Engineering. Cambridge, Mass.: Harvard University, May 2005. (Accessed June 23, 2006, at http://www.news.harvard.edu/gazette/daily/2005/05/wise.pdf.)

Need and Non-Need Factors Associated with Addiction Treatment Utilization in a Cohort of Homeless and Housed Urban Poor

Stefan G. Kertesz, MD, MSc,† Mary Jo Larson, PhD, MPA,‡ Debbie M. Cheng, ScD,§ Jalie A. Tucker, PhD, MPH,¶ Michael Winter, MPH,‖ Ashley Mullins, BS,** Richard Saitz, MD, MPH, FACP, FASAM,†††‡ and Jeffrey H. Samet, MD, MA, MPH††§§*

Background: Research on addiction treatment utilization in indigent samples mainly has been retrospective, without measures of addictive consequences, social network influences, and motivation. Prospective assessment of factors influencing utilization could inform policy and clinical care.

Objective: We sought to identify factors associated with utilization of addiction treatment and mutual help groups among substance-dependent persons with high rates of homelessness.

Research and Methods: This was a prospective cohort of patients detoxified from alcohol or drugs at baseline who were followed for 2 years in a randomized clinical trial of linkage to primary care (n = 274). Outcomes included utilization of Inpatient/Residential, Outpatient, Any Treatment, and Mutual Help Groups. Predictor variables in longitudinal regression analyses came from the literature and clinical experience, organized according to theoretical categories of Need, and non-Need (eg, Predisposing and Enabling).

Results: Many subjects used Inpatient/Residential (72%), Outpatient (62%), Any Treatment (88%) or Mutual Help Groups (93%) at least once. In multivariable analyses, addictive consequences (odds ratio [OR] 1.38, 95% confidence interval [CI] 1.12–1.71), motivation (OR 1.32, 95% CI 1.09–1.60), and female gender (OR 1.80, 95% CI 1.13–2.86) were associated with most treatment types (ORs are for Any Treatment). Homelessness was associated with Residential/Inpatient (for Chronically Homeless vs. Housed, OR 1.75, 95% CI 1.04–2.94). Living with one's children (OR 0.51, 95% CI 0.31–0.84) and substance-abusing social environment (OR 0.65, 95% CI 0.43–0.98) were negatively associated with Any Treatment.

Conclusions: In this cohort of substance-dependent persons, addictive consequences, social network variables, and motivation were associated with treatment utilization. Non-need factors, including living with one's children and gender, also were significant.

Key Words: homelessness, substance abuse, utilization, longitudinal

(*Med Care* 2006;44: 225–233)

From the *Divisions of Preventive and General Internal Medicine, University of Alabama at Birmingham School of Medicine and †Deep South Center on Effectiveness, Birmingham Veterans Affairs Medical Center, Alabama; ‡Institute for Health Services Research & Policy, New England Research Institutes, Boston, Massachusetts; §Department of Biostatistics, Boston University School of Public Health, Massachusetts; ¶Department of Health Behavior, University of Alabama at Birmingham School of Public Health; ‖Data Coordinating Center, Boston University School of Public Health, Massachusetts; **The Wright Institute, Berkeley, California; ††Section of General Internal Medicine, Boston University School of Medicine/Boston Medical Center, Massachusetts; ‡‡Department of Epidemiology, and the Youth Alcohol Prevention Center, Boston University School of Public Health, Massachusetts; and §§Department of Social and Behavioral Sciences, Boston University School of Public Health, 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 the College on Problems of Drug Dependence (San Juan, Puerto Rico 6/16/04), the Addiction Health Service Research Conference (Philadelphia, Pennsylvania, 10/8/04) and the Association for Medical Education and Research on Substance Abuse (Baltimore, Maryland, 11/13/05).

Reprints: Stefan G. Kertesz, MD, MSc, Division of Preventive Medicine, University of Alabama at Birmingham School of Medicine, 1530 3rd Ave South, MT608, Birmingham, AL 35294-4410. E-mail: skertesz@uab.edu.

ISSN: 0025-7079/06/4403-0225

Substance abuse and dependence affect 22 million Americans,[1] with effects on health and life-functioning, including unemployment, disrupted relationships, and homelessness.[2] Treatment can be effective, depending on the particular addictive problem and population.[2–5] However, in the United States, where only one-fourth of persons needing treatment obtain it,[6] a compelling challenge lies in the identification of factors affecting treatment utilization.[7] A national shortfall in treatment availability[8] likely accounts for the magnitude of the discrepancy between the population needing treatment and the population receiving it. This factor alone, however, does not explain why some persons with addictions seek and find treatment and others do not effectively seek treatment.

In this study, the housed and homeless urban poor drew our attention in part because of the devastating impact of substance abuse in inner-city populations and because of national commitments to eliminating health disparities[9] and chronic homelessness.[10]

Research on general community samples shows that interpersonal consequences of addiction and social network

factors are associated with treatment-seeking.[11,12] For instance, utilization is higher in substance abusers with greater negative consequences and social pressure to seek treatment and reduce misuse.[12] Whether such factors predict treatment utilization in poor, marginally housed, and homeless persons is less well-studied. A survey of soup kitchen clients (n = 190) found that drug problem recognition, desire for help, severity of substance use, and childcare responsibilities were positively related to motivation for treatment. That study, however, was not designed to test whether these factors affected utilization.[13]

Studies of indigent and/or homeless samples suggest that sociodemographic factors, including race and insurance, are associated with addiction treatment utilization.[14,15] These reports, however, lacked measures of addictive consequences, social network support for abstinence, or motivation for treatment (ie, factors relevant in general community samples). Additionally, although mutual help (eg, 12-step) groups may reduce substance misuse, few studies profile factors associated with participation among the urban poor.

Given these research gaps, we examined factors associated with utilization of formal addiction treatment and mutual help groups in an urban cohort of substance-dependent persons recruited from a publicly-funded detoxification unit and followed for 2 years as part of a randomized controlled trial of linkage to primary care.[16] We examined whether indicators of need and motivation predicted treatment, compared with non-need factors such as insurance or sociodemographics.

Our conceptual framework was the Behavioral Model for Vulnerable Populations, which arranges predictors of utilization in categories of predisposing, enabling/disabling, and need-related variables.[17,18] Predisposing characteristics typically include sociodemographics as well as life circumstances. Enabling/disabling characteristics have positive or negative instrumental relationships to the utilization under study, eg, insurance, transportation or social support. Need variables encompass objectively measured and subjectively assessed need for treatment. In light of a national commitment to resolve disparities in health care, we also assessed whether utilization met the Behavioral Model's proponents' criteria for *equity*, which is said to exist when need, as opposed to non-need (eg, race, insurance), characteristics account for utilization.[19]

This exploratory analysis attempted to improve upon existing research by focusing on the urban poor, including measures of addiction severity and consequences, motivation for change, social network support for abstinence, as well as essential sociodemographic variables, such as insurance, race, and homelessness.

METHODS

Data came from a clinical trial, conducted among urban alcohol and drug-dependent persons without primary care, recruited from a single urban residential (3–5 day) alcohol and drug detoxification unit. The trial intervention, a 90-minute clinical session conducted onsite before discharge, including a scheduled appointment, resulted in 69% of the intervention group obtaining a primary care visit, versus 53% of controls.[16] Subjects provided information regarding demographic, social, and health status, and were sought for follow-up interviews at 6-month intervals over the course of 2 years.

All subjects provided informed consent. This research was approved by the Institutional Review Boards of Boston Medical Center and the University of Alabama at Birmingham. Additional confidentiality was secured by issuance of a Certificate of Confidentiality from the Department of Health and Human Services to protect subjects from release of research data even under a court order or subpoena.

Study Sample

Trial eligibility criteria were: age older than 17 years and designation of alcohol, heroin, or cocaine as the substance of first or second choice. Exclusions were having a primary care provider; Mini-Mental State Examination[20] score <21; nonfluent in English or Spanish; <3 contacts to facilitate follow-up; or plans to leave the Boston area within 24 months. Of 642 eligible subjects, 470 (73%) participated.

Eligibility for the present analysis of addiction treatment utilization required completion of the baseline interview and at least 2 of 4 follow-up assessments (6, 12, 18, and/or 24 months). This requirement assured that analyzed subjects had adequate opportunity to be assessed. Moreover, it permitted categorization of housing status based on 3 longitudinal observations per subject, an approach approximating the U.S. federal definition for chronic homelessness[21] and demonstrated to capture a unique subgroup within this cohort.[22] Of 470 trial entrants,[16] 301 (64%) had sufficient follow-up. An additional 27 subjects were excluded because of missing covariates, leaving 274 (58% of trial entrants) for analysis. We compared these 274 subjects to those not included (n = 196) to assess potential for selection bias. Included and excluded subjects did not differ according to baseline housing status, age, gender, employment, substance of choice, number of chronic or episodic medical conditions, alcohol severity, suicidal ideation or intent, or social network variables (all $P > 0.1$, data not shown). Among included subjects, 52.6% had received the 90-minute trial intervention promoting linkage to primary care, compared with 46.4% of excluded subjects ($P = 0.19$). Included subjects differed from excluded subjects by race/ethnicity ($df = 3$, $P = 0.01$ in comparisons across 4 race/ethnicity groups); they were more likely to be black (53% vs. 37%) and less likely to be white (32% vs. 43%) compared with excluded subjects. Included subjects were more likely to identify cocaine (38% vs. 27%) and less likely to identify alcohol (36% vs. 44%) as substance of choice ($df = 2$, $P = 0.05$). Accordingly, included subjects had higher drug severity (mean Addiction Severity Index/drug score 0.3 vs. 0.2, $P = 0.01$).

To assess whether restriction to subjects providing baseline plus 2 follow-up observations affected analytic results, we performed a sensitivity analysis, repeating all models including all subjects with ≥1 follow-up (n = 374, 80% of sample), operationalizing housing status as a time-varying covariate (see below). Major findings were similar except where noted in the Results.

Outcome Variables

We modeled 4 outcomes (Residential/Inpatient Treatment, Outpatient Treatment, Mutual Help Group, and Any Treatment) based on self-report at baseline and 2 follow-ups. Residential/Inpatient Treatment was whether, during the past 6 months "or since we last met," subjects had spent 1 or more nights in programs including "halfway house or other residential facility," or "holding units" (eg, stabilization programs[23]), exclusive of detoxification facilities. Outpatient Treatment was contact with day treatment or outpatient programs (methadone programs not included), or addiction counseling from a substance abuse counselor, doctor, or health care worker. For both treatments, subjects were cued with lists of known programs (and could add others). Mutual Help Group was self-reported attendance at AA, NA, or "self-help groups." Any Treatment encompassed either inpatient or outpatient treatment or methadone (but not mutual help). This latter variable offered an "overall" result and permitted partial corroboration by an administrative indicator from the same period. Administrative records, from Massachusetts Bureau of Substance Abuse Services, were susceptible to incomplete reporting by treatment programs (Personal Communication, Teresa Anderson, Director of Statistics and Evaluation in the Center for Community Health, Commonwealth of Massachusetts, 6/29/05). We observed fair agreement between self-report and administrative sources (overall kappa = 0.29, 95% confidence interval [CI] 0.24–0.35).[24] Review of these data confirmed that discrepancies disproportionately entailed positive self-report without administrative confirmation (all $P < 0.0001$ by McNemar test), consistent with our expectation. These findings suggest the self-report outcomes were valid.

Predictor Variables

Potential predictors of utilization derived from the Behavioral Model for Vulnerable Populations,[18] as guided by relevant utilization literature[14,15] and our clinical observation. From the model's categories of Predisposing, Enabling, and Need, we chose 17 variables for inclusion in logistic regression analyses.

All subjects met criteria for needing treatment by virtue of the detoxification admission. Postdetoxification linkage to treatment is a quality target for behavioral health providers.[25,26] The National Institute on Drug Abuse states that detoxification itself is "only the first stage" and a potential "precursor" to effective treatment.[5] Accordingly, we modeled Treatment Need-related variables for their relation to whether treatment was used, including: Addiction Severity Index drug and alcohol composite scores (emphasizing intensity of use),[27] addictive consequences (total from the Inventory of Drug Use Consequences, a self-reported inventory of social/intrapersonal/physical consequences of substance misuse,[28]) and having ≥1 problem substance at baseline. Designation of a "problem substance" was based on use of a drug (or alcohol, to intoxication) 5 or more days in the prior 30, or ≥3 times weekly for 1 year.

Subjective, patient-assessed treatment need, although not directly measured, was partly reflected by the "taking steps" subscale of the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) instrument measuring motivation to change addictive behavior.[29] Appendix 1 provides detailed profiles for these psychobehavioral scales.

Predisposing variables included age, gender, and race (white vs. nonwhite). We devised housing status to approximate the federal definition for chronic homelessness using 3 observations per subject.[21] This definition incorporates persons with recurrent homeless episodes over a period of years, and a disabling condition, which by regulation includes addiction. At baseline and follow-up, subjects retrospectively estimated the number of nights in "an overnight shelter" and "on the street, without shelter" (both contributing to homeless nights) during the preceding 6 months or since last interview. A subject was "chronically homeless" based on reporting 1 night or more homeless at all 3 longitudinal assessments, and "transitionally homeless" if only 1 or 2 assessments included homeless nights. The qualifying "disabling condition" was addiction. A previous report demonstrates that among addicted persons, these 3 categories differ robustly with respect to mental-health related quality of life (worst for the chronically homeless), a contrast reflecting differences in physical and psychiatric illness, in social support, and in severity of addiction.[22]

Enabling/Disabling variables were health insurance, an overnight medical hospitalization (a potential treatment referral source), and social network factors favoring substance use (≥50% of "people you spend time with" use drugs or are heavy/problem drinkers), or abstinence (≥50% of people "you spend time with support your sobriety or abstinence"). Because we had observed clinically that cognitive and health factors sometimes resulted in patients being redirected within (or out of) the addiction treatment system, cognitive status,[20] and mental and physical component summaries of the Short Form-36 (SF-36)[30] were treated as "enabling/disabling." We included "lived with children" as a plausible proxy for childcare or familial responsibilities, a barrier to treatment in some reports.[31]

Analysis

For descriptive purposes, we computed the percentage of subjects reporting utilization of each treatment type at any 1 of the 3 study interviews. We constructed 4 separate logistic regression models, 1 for each treatment outcome: Residential/Inpatient, Outpatient, Mutual Help Group, and Any Treatment. Each subject contributed 3 observations over time for models shown (including baseline). To account for correlation due to repeated observations on each subject, a generalized estimating equation (GEE)[32] approach with an independence working correlation matrix was used (SAS PROC GENMOD). The empirical standard errors from the GEE models are reported. Each regression model included all 17 potential predictors and time of observation, a continuous variable ranging from 0 (baseline) to 24 months. On the basis of the number of observed treatment utilization events, there was sufficient data to include all independent variables in each model and maintain model stability. We permitted all predictors to vary over time except those that were unchangeable (eg, gender) and housing status, which was based on 3 observations per subject in these models (subsequently tested as a time-varying covariate in sensitivity analyses). We tested correlations between pairs of independent variables, and no

pair of variables included in the same model was highly correlated ($r > 0.40$). Because data came from a randomized controlled trial of facilitated primary care appointments,[16] models were repeated including randomization status; findings were not materially altered (data not shown). Odds ratios (ORs) were generated for each predictor based on contrast tests.[33] To facilitate comparisons, we z-transformed most continuous variables (except age), producing ORs for a 1-standard deviation increase in each continuous predictor.

Sensitivity Analyses

The main analyses are for self-reported utilization endpoints (reflecting the prior 6 months or time since last interview) in relation to predictors that included (1) static characteristics (eg, gender), (2) characteristics determined over various time frames (eg, housing status), and (3) behavior/attitude measures that may or may not reflect prior behavior/attitude (eg, addiction severity). An assumption of this approach is that behavior/attitude characteristics, measured at the same time as the self-reported treatment utilization, could function as independent variables in the prediction of treatment utilization (ie, within the last 6 months) even when, technically, the behavior/attitude instrument (eg, addictive consequences) was administered at the end of the observation interval, potentially *after* the treatment episode being predicted. This approach maximized power and assured temporal proximity of predictors to utilization. As a sensitivity analysis, we repeated the final models for our 274-subject cohort, lagging outcomes relative to predictors (ie, predictors could have been 6, 12 or 18 months prior to utilization outcomes, with 50% ≤200 days), excluding baseline utilization outcomes from analysis.

RESULTS

Subjects (n = 274) were followed a median 15.0 months, interquartile range 12.2–19.6. They were mainly middle-aged (median 35.0 years, interquartile range 31.0–41.0), male (76%), nonwhite (67.5%, with 53% black) and single (92%) (Table 1). Substances of choice were alcohol (36%), cocaine (38%), and heroin (26%), but 88% of subjects qualified as having 1 or more "problem substances" based on excess use. Of the 274 subjects, 60 (22%) and 108 (39%) were chronically and transitionally homeless, respectively. Chronically homeless subjects had a median 6.5 months homeless during 5 years before baseline compared with 1.0 month for transitionally homeless subjects.

Treatment Use

The percentage of subjects reporting each treatment at least once was 72% (Residential/Inpatient), 62% (Outpatient), 93% (Mutual Help), and 88% (Any). In multivariable analyses of treatment or mutual help group utilization, significant predictors emerged from the Need, Predisposing and Enabling/Disabling categories.

As Table 2 shows, among treatment need variables, the addictive consequences (InDUC) total score[28] was positively associated with formal treatment (Residential/Inpatient, Outpatient, Any Treatment), and Mutual Help Groups, whereas the drug and alcohol composite scores of the Addiction Severity Index were not. Higher motivation was positively associated with all utilization endpoints but associations were significant only for Residential/Inpatient, Any Treatment and Mutual Help Groups.

Among predisposing variables, Chronically and Transitionally Homeless status (compared with Housed) were both associated with Residential/Inpatient Treatment, but housing status was not associated with the other endpoints. Female gender was associated with Outpatient Treatment and with Any Treatment. We also found a strongly positive association between white (vs. nonwhite, 79% of whom were black) race and Mutual Help Groups (OR 2.67, 95% CI 1.58–4.50).

Among several significant enabling/disabling variables, the variable most consistently and positively associated with treatment utilization was health insurance, with the ORs for insured compared with noninsured subjects ranging from 1.64 to 2.81 across the 5 treatment outcomes (all $P < 0.01$), a finding that did not hold up in sensitivity analyses (see below). We also noted that having lived with one's children in the last 6 months was associated with reduced odds for all formal treatment utilization endpoints (OR's ranging from 0.43 to 0.53) except for Outpatient (OR 1.04, 95% CI 0.64–1.71); it was also associated with a reduced odds of utilizing Mutual Help groups (OR 0.53, 95% CI 0.31–0.90). Tests of interaction between Living with Children and gender, and Living with Children and housing status, were nonsignificant when applied to the Any Treatment outcome ($P = 0.52$ and $P = 0.14$, respectively).

Also, among enabling/disabling variables, the estimated associations between social network factors and utilization were consistent across all outcomes in that all ORs were greater than 1 for abstinence-supporting and less than 1 for substance-abusing social network, but 95% confidence intervals included 1.0 in several instances (Table 2). An overnight medical hospitalization was positively associated with Outpatient Treatment and Any Treatment. The inclusion of a variable representing trial arm assignment did not change these findings materially.

In sensitivity analyses, when models were broadened to include all subjects with at least 1 follow-up (n = 374, or 80% of the inception cohort), we treated homelessness as a time-varying binary variable (ie, the same subject could contribute homeless and housed observations). With this approach, major findings did not change substantially, except that homelessness was no longer associated with Residential/Inpatient treatment (OR 1.17, 95% CI 0.87–1.57). Lastly, a repetition of all models, lagging the utilization endpoints in relation to predictors (and dropping utilization endpoints at baseline) did not substantially alter the direction of most estimated relationships between predictors and outcomes, with the exception of insurance, where the estimated positive associations became nonsignificant and reversed direction (eg, OR for Any Treatment 0.78, 95% CI 0.51–1.20).

DISCUSSION

Summary

This analysis demonstrates that contrary to equity-based systems of care, some non-need related factors such as

TABLE 1. Baseline Characteristics of 274 Subjects Recruited at Detoxification and Followed for at Least 3 Observations Over the Course of 2 Years*

Characteristic	Total Sample (n = 274)
Demographics	
Age, years	35.0 (31.0–41.0)
Male, %	76.3%
Race/ethnicity	
Nonwhite	67.5%
White	32.5%
Married	8.0%
Employed, full-time	41.2%
Randomization group	
Control	47.4%
Clinic	52.6%
Housing status	
Chronically homeless	21.9%
Transitionally homeless	38.7%
Housed	39.4%
Months homeless, previous 5 years	0 (0.0–5.0)
Nights homeless, previous 6 months	0 (0.0–17.0)
Medical status	
No. chronic medical conditions	0 (0.0–1.0)
No. episodic medical conditions, previous 6 months	0 (0.0–1.0)
No. emergency department visits, previous 6 months	0 (0.0–1.0)
Overnight hospitalization, previous 6 months	12.4%
Medical severity score†	0.4 (0.0–0.7)
Physical Component Summary (PCS) score‡	48.8 (39.2–56.4)
Substance abuse	
Substance of choice§	
Alcohol	36.1%
Cocaine	37.6%
Heroin	26.3%
Drug severity†	0.3 (0.2–0.4)
Alcohol severity†	0.5 (0.1–0.8)
Tobacco use	86.9%
Drug and alcohol consequences, total¶	38.0 (33.0–41.0)
>1 problem substance	87.6%
Criminal justice problem related to substance abuse¶	75.6%
Job problem related to substance abuse¶	66.8%
Relationship problem related to substance abuse¶	69.3%
Psychiatric	
Suicide ideation/attempt, ever	30.3%
Psychiatric medication, ever	22.6%
Depressive symptoms at baseline‖	34.5 (26.0–41.0)
Physical or sexual abuse, ever**	71.4%
Mini-Mental Status Examination Score††	27.0 (25.0–29.0)
SF-36 Mental Component Summary (MCS) score‡	28.0 (21.6–40.1)
Social environment	
Substance-abusing social environment‡‡	67.9%
Abstinence-supporting social environment‡‡	38.7%
Other	
Health insurance	35.8%
Children	67.0%
Had lived with children in past 6 months	13.9%
Motivation (SOCRATES taking steps scale score)§§	36.0 (33.0–39.0)

*Continuous values are medians, interquartile ranges are given in parentheses.

†Medical severity, alcohol severity, and drug severity are from the medical, alcohol and drug composite scores of the Addiction Severity Index, on a scale of 0 to 1, with 1 indicating greatest severity.[27]

‡Physical component score (PCS) and mental component score (MCS) are computed from Short Form-36,[30] with 0 indicating lowest and 100 the highest health-related quality of life, standardized to a United States mean of 50 with standard deviation of 10.

§Substance of choice was defined on basis of subject's self-report at time of initial screening (see Methods section).

¶Total drug and alcohol consequences reflects the total score from the Inventory of Drug and Alcohol Consequences, with range from 0 to 45 (higher being worse), and separate binary indicators of job, criminal and relationship problem are also derived from this instrument.[28]

‖Depressive symptoms, using the Center for Epidemiologic Studies Depression Scale with range from 0 (none) to 60 (maximum depressive symptoms), and 16 or greater considered a significant depressive symptom burden.[48]

**Past history of abuse was based on the subject's response to questions seeking a lifetime history of exposure to physical or sexual abuse.

††The Mini-Mental Status score reflects an examiner-administered test of cognitive function and ranges from 0 (poor) to 30 (good).[20]

‡‡A substance abusing social environment was designated by subject self-assessment that ≥50% of "people you spend time with" use drugs or are heavy/problem drinkers. An abstinence-supporting social environment was designated by subject self-assessment that ≥50% of people "you spend time with support your sobriety or abstinence."

§§Raw score on the 'Taking Steps' scale from the Stages of Change Readiness and Treatment Eagerness Scale instrument, ranging from 8 (low motivation) to 40 (high motivation).[29]

having lived with one's child and nonwhite race were associated with reduced likelihood, whereas female gender was associated with increased likelihood of utilization of some addiction treatment types. Our findings were inconclusive with regard to insurance in a state (Massachusetts), in which treatment generally is more available,[34] and public treatment programs facilitated Medicaid applications. Conversely, we found meaningful indicators of treatment need (ie, addictive consequences and motivation) were positively associated with treatment. Additionally, this study extended previously-documented[12] associations between social network support for abstinence and treatment utilization to an indigent context.

Non-need Factors: Children, Gender, Insurance

"Living with children" was associated with robustly reduced odds of both inpatient/residential treatment and mutual help group use. Speculatively, competing childcare responsibilities may help explain these findings. This conjecture is buttressed by the observation that Massachusetts provided onsite childcare in outpatient (but not other) treatments during this study, and odds of outpatient treatment were not reduced by "living with children." Importantly, we found no significant interactions between the "children" variable and gender or housing status (power was limited for the latter), ie, the association between having lived with one's children and reduced utilization applied similarly across gender and housing groups. We infer that having lived with one's children served as a proxy for family responsibilities that

TABLE 2. Predictors of Residential, Outpatient, Any Treatment, and Mutual Help Group Utilization Among 274 Subjects Recruited at Detoxification and Followed for 3 Observations Over the Course of 2 Years*

	Self-Reported Residential Treatment	Self-Reported Outpatient Treatment (Non-Methadone)	Self-Reported Any Treatment	Self-Reported Mutual Help Group Treatment
	Odds Ratios (95% Confidence Intervals)			
Need variables				
Substance abuse				
Drug and alcohol consequences (+1 SD)‡	1.38 (1.12–1.71)†	1.44 (1.18–1.76)†	1.74 (1.39–2.18)†	1.54 (1.24–1.91)†
Motivation (SOCRATES, +1 SD)§	1.32 (1.09–1.60)†	1.18 (0.98–1.41)	1.36 (1.13–1.64)†	1.34 (1.10–1.62)†
>1 problem substance	1.06 (0.59–1.91)	1.40 (0.77–2.53)	1.62 (0.86–3.05)	1.80 (0.92–3.49)
Drug severity (+1 SD)¶	0.94 (0.78–1.12)	0.84 (0.69–1.03)	0.87 (0.71–1.06)	0.92 (0.72–1.17)
Alcohol severity (+1 SD)¶	0.84 (0.70–1.01)	0.96 (0.80–1.15)	0.90 (0.74–1.08)	0.85 (0.70–1.03)
Predisposing variables				
Housing status‖				
Chronically homeless vs. housed	1.75 (1.04–2.94)†	0.85 (0.50–1.46)	1.02 (0.60–1.72)	1.21 (0.66–2.21)
Transitionally homeless vs. housed	1.79 (1.23–2.62)†	0.66 (0.43–1.01)†	0.97 (0.64–1.45)	0.99 (0.65–1.53)
Chronically homeless vs. transitionally homeless	0.97 (0.59–1.60)	1.30 (0.78–2.14)	1.05 (0.65–1.70)	1.21 (0.66–2.25)
Demographics				
Female vs. male	1.10 (0.72–1.67)	1.72 (1.13–2.61)†	1.80 (1.13–2.86)†	1.07 (0.62–1.83)
Age (10 yr)	0.76 (0.60–0.96)†	1.15 (0.90–1.46)	0.83 (0.66–1.05)	0.80 (0.62–1.05)
White vs. nonwhite	1.43 (0.99–2.08)	1.17 (0.77–1.78)	1.41 (0.94–2.11)	2.67 (1.58–4.50)†
Medical**				
Physical health (+1 SD)	1.22 (1.02–1.47)†	0.96 (0.79–1.16)	1.15 (0.95–1.38)	1.07 (0.88–1.31)
Mental health (+1 SD)	1.09 (0.91–1.31)	0.89 (0.73–1.09)	0.99 (0.82–1.19)	1.10 (0.89–1.34)
Enabling variables				
Living with children	0.43 (0.26–0.69)†	1.04 (0.64–1.71)	0.51 (0.31–0.84)†	0.53 (0.31–0.90)†
Health Insurance	2.26 (1.61–3.18)†	2.04 (1.42–2.94)†	2.81 (2.00–3.95)†	2.02 (1.35–3.02)†
Substance-abusing social environment††	0.84 (0.57–1.24)	0.64 (0.44–0.93)†	0.65 (0.43–0.98)†	0.59 (0.39–0.91)†
Abstinence-supporting social environment††	1.37 (0.97–1.92)	1.16 (0.83–1.62)	1.38 (0.98–1.93)	1.61 (1.09–2.39)†
Overnight medical hospitalization	1.07 (0.69–1.66)	1.70 (1.11–2.60)†	1.68 (1.01–2.80)†	1.32 (0.75–2.34)
Mini-Mental State Examination Score (+1 SD)‡‡	0.96 (0.81–1.15)	0.98 (0.81–1.18)	0.92 (0.76–1.11)	1.07 (0.87–1.30)

*Five longitudinal regression models are shown, with adjustment for all variables shown in table.

†Results are statistically significant at the $P = 0.05$ level.

‡Total Drug and Alcohol consequences reflects the total score from the Inventory of Drug and Alcohol Consequences, with range from 0 to 45 (higher being worse), and separate binary indicators of job, criminal and relationship problem are also derived from this instrument.[28]

§Motivation = Raw score on the 'Taking Steps' scale from the Stages of Change Readiness and Treatment Eagerness Scale instrument, ranging from 8 (low motivation) to 40 (high motivation).[29]

¶Alcohol severity and Drug severity are from the medical, alcohol and drug composite scores of the Addiction Severity Index, on a scale of 0 to 1, with 1 indicating greatest severity.[27]

‖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 1 or 2 of 3 assessments were designated "transitionally homeless," while subjects reporting no homelessness were considered "housed" (see Methods).

**Physical component score (PCS) and mental component score (MCS) are computed from Short Form-36, with 0 indicating lowest and100 the highest health-related quality of life, standardized to a United States mean of 50 with standard deviation of 10.[30]

††A substance abusing social environment was designated by subject self-assessment that ≥50% of "people you spend time with" use drugs or are heavy/problem drinkers. An abstinence-supporting social environment was designated by subject self-assessment that ≥50% of "people you spend time with support your sobriety or abstinence."

‡‡The Mini-Mental Status score reflects an examiner-administered test of cognitive function and ranges from 0 (poor) to 30 (good).[20]

could vary in specifics from person to person. Even for some homeless persons, parental/familial responsibilities may compete with treatment.

The independent association of female gender with increased odds of outpatient treatment was unexpected, given prior reports of female gender-specific barriers to addiction treatment entry.[35] Although psychosocial issues, stigma and resource issues could impede women's access to treatment in some contexts,[35] childcare availability may have mitigated these issues for some clients. Finally, utilization studies may not be directly comparable across localities and treatment systems.

With the exception of Mutual Help Groups, the lack of significant race effects in this sample contrasts with prior reports from residentially unstable populations. White race predicted treatment in Los Angeles,[15] whereas black race was associated with treatment in Houston[14] and New York samples.[36] Race effects likely vary by location. This study's positive relationship between white race and Mutual Help Groups contrasts with a California sample in which blacks

had a higher degree of 12-step affiliation than whites.[37] Blacks in our cohort may have felt less comfortable attending 12-step groups in a city where only 25% of residents are black.

An apparent association between treatment utilization and insurance, noted in cross-sectional samples of homeless persons,[14,15,36] and among persons hospitalized for detoxification[38] did not hold up when insurance was used to predict utilization 6–18 months later. Since Massachusetts expanded Medicaid eligibility at the time of this cohort, and enrolled patients upon entry to the publicly funded treatment system, we believe this study's insurance-treatment association is consistent with reverse causation, with treatment leading to insurance.

Need-Related Factors: Consequences, Motivation, Problem Severity

Consistent with reports from less marginal populations,[12,39] this study shows that personal consequences of substance misuse and social network support for abstinence are associated with treatment. Consequences were positively associated with all treatment types while a substance-abusing social environment was negatively associated with outpatient treatment, any treatment (self-reported), and mutual help groups. Although motivation has been measured in indigent samples,[13] to our knowledge, this study is the first to show it predicts treatment utilization in this context. By contrast, the Addiction Severity Index drug and alcohol composite scores, heavily influenced by quantity/frequency of use (rather than consequences),[27] did not predict utilization. This pattern of results agrees with Tucker's formulation that subjectively-perceived and socially-mediated consequences of addiction, rather than use itself, largely account for help-seeking and behavioral change.[40]

Homelessness

The characteristics of chronic or transitional homelessness (previously shown to portend poor health prognoses[22] and high overall service utilization[41]) were associated with residential/inpatient treatment in our primary analysis. This association may reflect both appropriate triage, and homeless clients' response to limited housing options. Observational data, trials and reviews of the literature suggest that effective addiction treatment of homeless persons must address the need for shelter to be effective.[23,42–44] Our clinical observation has been that some housing programs require addiction treatment as a condition of entry, reinforcing residential treatment as the "housing program of first resort" for many homeless.

The association between homelessness and residential treatment diminished when modeling homelessness as a time-varying binary "state," (permitting the subject to count as housed in 1 interval and homeless in another, affecting at least 65% of subjects at least once) as opposed to a "trait" adduced from 3 observations over time. An interpretation for this contrast is that the "trait" approach generates a more extreme contrast between housing status groups, eg, the "housed" group consists of persons who were never homeless at any time during the study period. This group proves much less likely to use residential treatment. Conversely, treating homelessness as a time-varying "state" cannot differentiate a subgroup of stably housed individuals. Future research should incorporate the insight that homelessness has plausible "state" (eg, one is or one is not homeless on any given night) and "trait" attributes (eg, falling out of housed status, even once, identifies a distinctly vulnerable sociological class[22]).

Strengths and Limitations

This study's strengths include its use of longitudinal data for both treatment and housing status over time. We studied residential and outpatient treatments, and mutual help group participation, distinctions not addressed in prior studies of indigent samples. Our data included detailed, specific measures for addiction severity,[27] consequences,[28] social network substance use and motivation for treatment.[29] Inclusion of these covariates identified consistency in drivers of treatment-seeking between this indigent sample and prior reports from middle-class samples.[11,12]

Limitations include reliance on self-report. However, self-report may be more valid for assessing whether any treatment was obtained during an interval[45] as opposed to treatment quantity. Moreover, administrative data was partially corroborative.

Implications

These findings have implications for policymakers and treatment professionals. Non-need factors associated with treatment utilization are troubling and may amount to disparities in care. The reduced odds of treatment associated with "living with children" is notable. Most addiction treatment programs do not accommodate a subject's parental responsibilities; some may assume that for men, or homeless clients, that such considerations are irrelevant. Our findings suggest that parent-child relationships remain relevant, and further research may be necessary to clarify whether adjustments to the treatment system, including onsite childcare, could help.

Increased utilization of residential/inpatient treatment among chronically homeless and transitionally homeless (vs. consistently housed) persons is consistent with recommendations for this population.[42–44] Further research should assess whether residential treatment leads to long-term rehabilitative success, or merely serves as a temporary housing substitute.

Clinicians seeking to spur substance-abusing clients into treatment may wish to focus attention on substance abuse consequences, motivation for treatment, social network factors and family relationships. Our report suggests that even in a context of socioeconomic marginality, treatment-seeking behavior is guided in good part by how individuals assess substance abuse consequences in their current social context. Given these findings, motivational interviewing[46] and cultivation of social network support for abstinence are likely to enhance treatment-seeking.[47]

ACKNOWLEDGMENTS

We thank the staff of the HELP study for their help in the conduct of this research.

REFERENCES

1. Substance Abuse and Mental Health Services Administration. *Overview of Findings from the 2003 National Survey on Drug Use and Health.*

Rockville, MD: Office of Applied Studies, NSDUH Series H-24, DHHS Publication No. SMA 04-3963; 2004.
2. Graham AW, Schultz TK, Mayo-Smith MF, et al, eds. *Principles of Addiction Medicine*, 3rd ed. Chevy Chase, MD: American Society of Addiction Medicine; 2003.
3. McLellan AT, Lewis DC, O'Brien CP, et al. Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. *JAMA*. 2000;284:1689–1695.
4. Simpson DD, Joe GW, Fletcher BW, et al. A national evaluation of treatment outcomes for cocaine dependence. *Arch Gen Psychiatry*. 1999;56:507–514.
5. National Institute on Drug Abuse. *Principles of Drug Addiction Treatment: A Research-Based Guide*. Bethesda, MD: National Institutes of Health; 1999. NIH Publication No. 99-4180.
6. Schneider Institute for Health Policy. *Substance Abuse: The Nation's Number One Health Problem*. Princeton, NJ: The Robert Wood Johnson Foundation; 2001.
7. Tucker JA. Changing addictive behavior: historical and contemporary perspectives. In: Tucker JA, Donovan DM, Marlatt GA, eds. *Changing Addictive Behavior: Bridging Clinical and Public Health Strategies*. New York: Guilford Publications; 1999.
8. Amaro H. An expensive policy: the impact of inadequate funding for substance abuse treatment. *Am J Public Health*. 1999;89:657–659.
9. Department of Health and Human Services. *Healthy People 2010: Understanding and Improving Health*. 2nd ed. Washington, DC: Government Printing Office; 2000.
10. The Secretary's Work Group on Ending Chronic Homelessness. *Ending Chronic Homelessness: Strategies for Action*. Washington, DC: Department of Health and Human Services; 2003.
11. Finney JW, Moos RH. Entering treatment for alcohol abuse: a stress and coping model. *Addiction*. 1995;90:1223–1240.
12. George AA, Tucker JA. Help-seeking for alcohol-related problems: social contexts surrounding entry into alcoholism treatment or Alcoholics Anonymous. *J Stud Alcohol*. 1996;57:449–457.
13. Nwakeze PC, Magura S, Rosenblum A. Drug problem recognition, desire for help, and treatment readiness in a soup kitchen population. *Subst Use Misuse*. 2002;37:291–312.
14. Wenzel SL, Audrey Burnam M, Koegel P, et al. Access to inpatient or residential substance abuse treatment among homeless adults with alcohol or other drug use disorders. *Med Care*. 2001;39:1158–1169.
15. Koegel P, Sullivan G, Burnam A, et al. Utilization of mental health and substance abuse services among homeless adults in Los Angeles. *Med Care*. 1999;37:306–317.
16. 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.
17. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behavior*. 1995;36:1–10.
18. Gelberg L, Andersen RM, Leake BD. The Behavioral Model for Vulnerable Populations: application to medical care use and outcomes for homeless people. *Health Serv Res*. 2000;34:1273–1302.
19. Aday L, Andersen R, Fleming GV. *Health Care in the U.S.: Equitable for Whom*? Beverly Hills, CA: Sage Publications; 1980.
20. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189–198.
21. 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 3/11/2004.
22. Kertesz SG, Larson MJ, Horton NJ, et al. Homeless chronicity and health-related quality of life trajectories among adults with addictions. *Med Care*. 2005;43:574–585.
23. Kertesz SG, Horton NJ, Friedmann PD, et al. Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. *J Subst Abuse Treat*. 2003;24:197–207.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
25. Washington Circle Policy Group. Specification of Performance Measures for the Identification, Initiation and Engagement of Alcohol and Other Drug Services; 2004.
26. Mark TL, Dilonardo JD, Chalk M, et al. *Substance Abuse Detoxification: Improvements Needed in Linkage to Treatment*. Rockville, MD: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration; 2002.
27. McLellan AT, Kushner H, Metzger D, et al. The Fifth Edition of the Addiction Severity Index. *J Subst Abuse Treat*. 1992;9:199–213.
28. 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.
29. Miller WR, Tonigan JS. Assessing drinkers' motivation for change: the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychol Addict Behav*. 1996;10:81–89.
30. 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.
31. Scott-Lennox J, Rose R, Bohlig A, et al. The impact of women's family status on completion of substance abuse treatment. *J Behav Health Serv Res*. 2000;27:366–379.
32. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986;42:121–130.
33. SAS Institute Inc. SAS/STAT Software: Changes and Enhancements, Release 8.2. Cary, NC: SAS Institute Inc.; 2001.
34. McAuliffe WE, Dunn R. Substance abuse treatment needs and access in the USA: interstate variations. *Addiction*. 2004;99:999–1014.
35. Schober R, Annis HM. Barriers to help-seeking for change in drinking: a gender-focused review of the literature. *Addict Behav*. 1996;21:81–92.
36. Padgett D, Struening EL, Andrews H. Factors affecting the use of medical, mental health, alcohol, and drug treatment services by homeless adults. *Med Care*. 1990;28:805–821.
37. Kaskutas LA, Weisner C, Lee M, Humphreys K. Alcoholics anonymous affiliation at treatment intake among white and black Americans. *J Stud Alcohol*. 1999;60:810–816.
38. Mark TL, Dilonardo JD, Chalk M, Coffey RM. Trends in inpatient detoxification services, 1992–1997. *J Subst Abuse Treat*. 2002;23:253–260.
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. Tucker JA, King MP. Resolving alcohol and drug problems: influences on addictive behavior change and help-seeking processes. In: Tucker JA, Donovan DM, Marlatt GA, eds. *Changing Addictive Behavior: Bridging Clinical and Public Health Strategies*. New York: Guilford Press; 1999:97–126.
41. 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.
42. Zerger S. *Substance Abuse Treatment: What Works for Homeless People? A Review of the Literature*. Nashville, TN: National Health Care for the Homeless Council; 2002.
43. Orwin RG, Goldman HH, Sonnefeld LJ, et al. *Community Demonstration Grant Projects for Alcohol and Drug Abuse Treatment of Homeless Individuals: Final Evaluation Report*. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism; 1993. NIH Pub. No. 93-3541.
44. Milby JB, Schumacher JE, Wallace D, et al. To house or not to house? does it make a difference? *Am J Public Health*. 2005;95:1259–1265.
45. Goldberg RW, Seybolt DC, Lehman A. Reliable self-report of health service use by individuals with serious mental illness. *Psychiatr Serv*. 2002;53:879–881.
46. Miller WR, Rollnick S. *Motivational Interviewing: Preparing People for Change*, 2nd ed. New York: The Guilford Press; 2004.
47. Longabaugh R, Wirtz PW, Zweben A, Stout RL. Network support for drinking, Alcoholics Anonymous and long-term matching effects. *Addiction*. 1998;93:1313–1333.
48. Zanis DA, McLellan AT, Cnaan RA, et al. Reliability and validity of the addiction severity index with a homeless sample. *J Subst Abuse Treat*. 1994;11:541–548.
49. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psych Meas*. 1977;1:385–401.

APPENDIX 1

Descriptions of Psychologic Instruments

1. Addiction Severity Index Alcohol Composite Score

Description: An interviewer-administered questionnaire assessing intensity of alcohol use, and limited measures for consequences (eg, delirium tremens), resulting in a mathematically calculated, unanchored score (19 items)

Example Item: How many times in the last 6 months have you had Alcohol DTs?

Range: 0–1, higher scores indicating greater severity

Available Reliability in Reference Literature: 3-day test-retest Spearman-Brown coefficient among homeless subjects: 0.89

Sources: McLellan et al,[27] Zanis et al[48]

Internal Reliability in the present study: Cronbach's alpha = 0.92 (Cronbach's Standardized Alpha was computed from our data. Reported Alpha's in the literature do not consistently indicate whether the value was standardized or not.)

2. Addiction Severity Index Drug Composite Score

Description: An interviewer-administered questionnaire assessing intensity of alcohol use, and limited measures for consequences, resulting in a mathematically calculated, unanchored score (13 items)

Example Item: How many days in the past 30 have you used heroin?

Range: 0–1, higher scores indicating greater severity

Available Reliability in Reference Literature: 3-day test-retest Spearman-Brown coefficient among homeless subjects: 0.93

Sources: McLellan et al[27] Zanis et al[48]

Internal Reliability in the present study: Cronbach's alpha = 0.70

3. CES-D Scale[49]

Description: A 20-item self-report depression scale (self- or interviewer-administered) designed to produce a quantitative measure of depressive symptoms.

Example Item: Please tell me how often you have felt this way in the last week: "I was bothered by things that usually don't bother me" with response options: Rarely or none of the time, Some or a little of the time, Occasionally or a moderate amount of time, Most or all of the time.

Range: 0–60, with ≥16 considered significant for depressive symptoms.

Available Reliability in Reference Literature: 2-week test-retest correlation r = 0.51

Source: Radloff et al[49]

Internal Reliability in the present study: Cronbach's alpha = 0.89

4. Inventory of Drug Use Consequences (InDUC) Total Score

Description: Interviewer-administered 50-item questionnaire to assess adverse consequences of drug use

Example Item: During the past 6 months, about how often has this happened to you? *"I have failed to do what is expected of me because of my drinking or drug use"* with response options: Never/Once or a Few Times/Once or twice a week/Daily or almost daily

Range: 0 (no consequences) to 45 (maximum)

Available Reliability in Reference Literature: test–retest correlation r = 0.75 (personal communication with J. Scott Tonigan, 6/16/05)

Source: Tonigan[28]

Internal Reliability in the present study: Cronbach's alpha = 0.91

5. Stages of Change Readiness and Treatment Eagerness (SOCRATES) Taking Steps Scale

Description: 19 items, with 8 items used to calculate Taking Steps score

Example Item:"I am actively doing things now to cut down or stop drinking" Response options: Strongly Agree to Strongly Disagree

Range: 8 (lowest interest in taking steps) to 40 (highest interest in taking steps).

Reliability in Reference Literature: Cronbach alpha = 0.95, test-retest r = 0.93

Source: Miller[29]

Internal Reliability in the present study: Cronbach's alpha = 0.83

6. Mini-Mental State Examination

Description: A simplified, scored cognitive status examination administered face-to-face (11 items)

Example Item: "Read and obey the following 'CLOSE YOUR EYES'" (1 point)

Range: 0 (poor) to 30 (good)

Available Reliability in Reference Literature: 24-hour test-retest (same examiner) Pearson coefficient 0.89. 24-hour test-retest (2 examiners) 0.83

Source: Folstein et al[20]

Internal Reliability in the present study: not available (eg, instrument administered at screening and only final score is available)

Available online at www.sciencedirect.com
ScienceDirect
Drug and Alcohol Dependence 85 (2006) 258–262
DRUG and ALCOHOL DEPENDENCE
www.elsevier.com/locate/drugalcdep

Short communication

Low bone density in patients receiving methadone maintenance treatment

Theresa W. Kim *, Daniel P. Alford, Alan Malabanan, Michael F. Holick, Jeffrey H. Samet

Boston University School of Medicine, Department of Medicine, 91 E. Concord Street, Ste. 200, Boston, MA 02131, USA

Received 3 February 2006; received in revised form 22 May 2006; accepted 31 May 2006

Abstract

Aim: To examine the frequency and severity of low bone mineral density (BMD) among patients enrolled in a methadone maintenance treatment (MMT) program and to ascertain risk factors for low BMD in this population.
Design: Cross-sectional.
Measurements: Data derived from standardized survey, medical record review, and dual energy X-ray densitometry (DXA).
Results: DXA results were below normal in 83% (76/92) of the study sample with *T*-scores <−2.5 (osteoporosis range) in 35% [32/92] and between −1.0 and −2.5 (osteopenia range) in 48% [44/92]. Risk factors for low BMD were common: tobacco use, 91%; heavy alcohol use, 52%; and HIV infection, 28%. Only 17% (16/92) were on medications that lower the risk of osteoporosis: estrogen ($n=5$), testosterone ($n=4$), calcium ($n=4$), and Vitamin D ($n=2$). None of the participants reported a known diagnosis of osteoporosis. In bivariate analyses, significant predictors of low BMD were: male gender ($p<0.001$), lower weight ($p=0.009$), and heavy alcohol use ($p=0.02$).
Conclusion: More than three quarters of this sample of patients in a MMT program had low BMD. Treatable conditions associated with low BMD were commonplace. Efforts to increase awareness of low BMD in MMT patients should be considered so that effective treatment may be employed to lower future fracture risk.

Keywords: Methadone maintenance; Epidemiology; Comorbidity; Bone metabolism

1. Background

Osteoporosis is a systemic disease, characterized by low bone mineral density (BMD) and micro-architectural deterioration, predisposing to fracture after minimal trauma or fall. Osteoporosis-related fractures are associated with physical functioning decline (Fink et al., 2003), impaired ambulation, and premature mortality (Johnell et al., 2004). Although effective treatment exists to reduce the risk of future fracture (Wilson, 2004), osteoporosis is underecognized and undertreated (Neuner et al., 2003; Port et al., 2003; Wong et al., 2003). This study examines whether osteoporosis occurs in patients with opioid dependence in methadone maintenance treatment (MMT) to an extent that suggests that efforts should be directed towards increasing its recognition and treatment.

Patients in MMT may be at higher risk for low BMD due to several reasons. First, direct opioid effects on bone metabolism may occur through inhibition of osteoblast functioning (Perez-Castrillon et al., 2000; Rosen et al., 1998), the cells responsible for new bone formation. Moreover, hypogonadism, a potential side effect of opioids (Daniell, 2004; Woody et al., 1988), is an important secondary cause of osteoporosis (Gennari et al., 2003; Mikhail, 2003). In addition to opioid-related effects, patients in MMT may have comorbid conditions associated with osteoporosis including HIV infection (Amiel et al., 2004; Mondy et al., 2003), tobacco dependence (Izumotani et al., 2003), and alcohol use disorders (Sampson, 2002). The objectives of this study were to evaluate BMD and risk factors for bone loss in patients receiving methadone for opioid dependence.

2. Methods

2.1. Study design and sampling

This was a cross-sectional study of participants recruited from the Boston Public Health Commission's Methadone Maintenance Treatment Program. Patients were excluded from the study if they were (1) pregnant due to radiation exposure during BMD measurement or (2) over 300 lb. due to mechanical limitations of the dual energy X-ray densitometry (DXA) table.

* Corresponding author.
E-mail address: theresa.kim@bmc.org (T.W. Kim).

0376-8716/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2006.05.027

2.2. Data collection

Research associates administered standardized interviews assessing the following: demographics; Vitamin D and calcium supplementation; tobacco, alcohol, and heroin use; and MMT enrollment duration. Medication information was obtained through medical record review.

BMD was assessed using DXA (QDR®4500W series; Hologic, Inc., Bedford, Massachusetts) of the lumbar spine (L1–L4), hip (excluding Ward's region) and forearm (distal one-third radius of non-dominant arm). DXA results are reported as *T*-scores, which are calculated by comparing an individual's BMD to a gender and race/ethnicity-matched, young adult reference population. *Z*-scores are calculated in a similar fashion to *T*-scores except that the reference population is also age-matched.

Study participants received a US$ 40 gift certificate to a local supermarket. The Boston University Institutional Review Board approved the study protocol. An NIH Certificate of Confidentiality was obtained for added privacy protection.

2.3. Statistical analysis

Differences between BMD groups were tested using bivariate analyses of variance (ANOVA), Chi-square or Fisher's exact test as appropriate. *p*-Values, two-tailed, <0.05 were considered statistically significant. All statistical analyses were performed using SAS Version 8.2 (SAS Institute Inc., 2001).

3. Results

3.1. Sample characteristics

Characteristics of the study sample ($n=92$) are listed in Table 1. No statistically significant differences were found between the study sample and the entire MMT clinic population ($n=350$) in terms of age, gender, or race/ethnicity.

Median opioid use (i.e., heroin or prescription opioids) prior to enrollment in MMT was 14 years and duration of participation in MMT varied widely. One-third of women reported cessation of menses for more than one year; this subgroup had a median age of 42 years (range 31–59). Since women with opioid dependence may have menstrual cycle disruptions, including persistent amenorrhea unrelated to menopause (Schmittner et al., 2005), it is unclear whether these women were postmenopausal. Many participants had risk factors for osteoporosis. Tobacco use was almost universal and prolonged. Half the sample reported past regular use of "heavy" amounts of alcohol for years. About a quarter of the study sample reported HIV infection. Five participants had been prescribed oral steroid medications, drugs that are known to cause osteoporosis with chronic use.

Despite risk factors for osteoporosis, only 16% (15/92) of the sample was on medications that may help to decrease the rate of bone loss: estrogen ($n=5$), testosterone ($n=4$), calcium ($n=4$), Vitamin D ($n=2$), and bisphosphonate ($n=1$). No participant was aware of a history of osteoporosis.

3.2. DXA results

We used World Health Organization criteria to classify BMD as osteoporosis (*T*-score ≤ -2.5), osteopenia (*T*-score between -2.5 and -1.0) or normal (*T*-score ≥ -1) (World Health Organization, 2003). More than 3/4 of the study sample (83%, 76/92) met *T*-score criteria for either osteoporosis (35%, 32/92) or osteopenia (48%, 44/92) (Table 2). There were significant gender differences. Low *T*-scores were almost universal among the men (97%, 32/33) with the predominant abnormality, osteoporosis (61%, 20/33). In contrast, 75% (44/59) of the women had low *T*-scores with the predominant abnormality, osteopenia (54%, 32/59) rather than osteoporosis (20%, 12/59).

To assess risk factors for low BMD in this population, *Z*-scores were used to decrease the confounding effect of age

Table 1
Characteristics of study participants recruited from a methadone maintenance treatment program ($n=92$)

Categorical variables	*N*	%
Race/ethnicity		
Black	44	48
Hispanic	12	13
White	36	39
BMI[a]		
<20	1	1
20–24	35	38
25–29	20	22
≥30	36	39
Female	59	64
Persistent amenorrhea[b]	19	21
Menses within past year	40	43
Tobacco		
Lifetime	87	95
Current	84	91
Former	3	3
Never	5	5
Heavy alcohol[c]		
Lifetime	47	52
Current	15	16
Former	32	36
Never	45	49
HIV infection	26	28

Continuous variables	Median	Range
Age, years	42	20–66
Lifetime heroin use, years	14	1–38
Methadone maintenance treatment, years	3	0–25
Current methadone dose, mg	77	3–140
Lifetime tobacco use, years	27	1–47
Lifetime heavy alcohol use, [c]years	7	1–35

[a] Weight (lb.) × 703/height (in.).
[b] Assessed with the questions, "Aside from pregnancy and birth control medication, have you stopped having your period?" and "How old were you when you had your last period?".
[c] Defined as >3 drinks/occasion, >3 occasions/week for at least 1 year.

Table 2
Bone mineral density results for the total study sample and stratified by gender

T-score	Women ($n=59$)	Men ($n=33$)	Total ($n=92$)
≥−1.0	15 (25%)	1 (3%)	16 (17%)
<−1.0 and >−2.5	32 (54%)	12 (36%)	44 (48%)
≤−2.5	12 (20%)	20 (61%)	32 (35%)

All sites measured by DXA were considered except Ward's area to classify BMD.

Table 3
Bivariate analysis of variables associated with low bone mineral density[a]

	Z-score[a]			p-Value
	≥ −1.0 (n = 33)	< −1 to −2.5 (n = 44)	≤ −2.5 (n = 15)	
Gender				
Male	3 (9%)	15 (45%)	15 (45%)	<0.001
Women				
Persistent amenorrhea	6 (32%)	12 (63%)	1 (6%)	
Menses with past year	18 (45%)	20 (50%)	2 (6%)	
Race/ethnicity				
Black	14 (32%)	20 (45%)	10 (23%)	0.48
White	9 (24%)	23 (62%)	5 (14%)	
Hispanic	4 (36%)	4 (36%)	3 (27%)	
HIV infection				
Yes	9 (%)	12 (46%)	6 (23%)	0.81
No	24 (%)	35 (53%)	12 (18%)	
Recent heroin use				
Yes	4 (21%)	12 (63%)	3 (16%)	0.49
No	23 (32%)	35 (48%)	15 (21%)	
Age, mean (S.D.)	41.2 (6.9)	41.9 (10.3)	45.0 (9.2)	0.26
Weight (kg), mean (S.D.)	84.4 (17.1)	77.8 (15.2)	69.8 (10.9)	0.009
Tobacco, lifetime years, mean (S.D.)	26.0 (7.2)	26.2 (12.2)	29.3 (8.9)	0.51
Heavy drinking, lifetime years, mean (S.D)[b]	3.3 (2.1)	9.5 (8.4)	11.6 (6.8)	0.02
Heroin, lifetime years, mean (S.D)	11.7 (8.2)	13.8 (10.9)	17.5 (9.4)	0.16
MMT, lifetime years, mean (S.D)	4.7 (4.2)	5.6 (5.3)	5.6 (6.6)	0.77
Methadone dosage (mg), mean (S.D)	83.3 (37.6)	75.9 (31.2)	76.0 (33.9)	0.54

[a] Z-score is the number of standard deviations that an individual's measured BMD is compared to an age-matched reference population.
[b] >3 drinks on >3 occasions per week for at least 1 year.

on BMD. Low BMD (Z-score <−1.0) was significantly associated with the following (Table 3): male gender ($p < 0.001$), lower weight ($p = 0.009$), and more years of heavy alcohol use ($p = 0.02$). There was a positive but non-significant relationship between longer duration of heroin use among those in the groups with lower Z-scores. Current heroin use, methadone dosage, and duration of MMT were not associated with lower BMD.

4. Discussion

We found that more than three quarters of patients recruited from one MMT clinic had abnormally low BMD. These findings indicate that patients in MMT programs may be at higher risk for fracture than the general population. Increased fracture risk in this population has particular significance given the high rates of injuries (Rees et al., 2002) and worse physical functioning (De Alba et al., 2004; Friedmann et al., 2003) in individuals with addictions.

We also found that a high percentage had osteoporosis risk factors including almost universal tobacco use. Low BMD is yet another reason for continued tobacco cessation efforts in this population. Although excess body weight is considered protective of BMD, low BMD was found despite the overweight or obese character of the sample. The association between heavy alcohol use and low BMD is consistent with studies in other populations (Sampson, 2002; Turner, 2000).

BMD in patients with opioid dependence has received limited attention. Pedrazzoni et al., 1993 found lower lumbar BMD in "recent" heroin users (1–2 days after last use) compared to "former" heroin users (4–24 months since last use) in a cross-sectional study of 22 male heroin users. None of the subjects were receiving methadone. Arnsten et al., 2006 found that MMT was associated with low BMD in middle-aged women either with or at risk for HIV infection. The current study builds upon this work by examining men and premenopausal women in MMT. These studies' findings and ours suggest that low BMD may be a common comorbidity in chronic opioid users.

An unexpectedly high proportion of the male sample had abnormal BMD. Other studies have found high rates of osteoporosis in men among patients with depression (Mussolino et al., 2004) and schizophrenia (Hummer et al., 2005). Reasons for this are unknown but may reflect the high prevalence of secondary causes of osteoporosis (Licata, 2003). These findings of exceptionally high proportions of patients in MMT with osteopenia or osteoporosis merit further examination in other cohorts since effective treatment is available (Wilson, 2004).

These results should be interpreted recognizing several limitations. Despite similar demographic characteristics between the MMT clinic population and study participants, patients with a family history of osteoporosis may have been more likely to participate, potentially overestimating of the prevalence of low BMD. Second, while this study included risk factors for low BMD not accounted for in previous studies of opioid-dependent

patients, other issues such as genetic influences (Videman et al., 2002) and level of physical activity were not assessed. Third, we included multiple sites of BMD measurements in the results to better assess fracture risk, although peripheral DXA sites are generally not used for diagnostic classification. Finally, generalizability of the study's findings is limited by recruitment from one MMT clinic.

Despite these limitations, these findings are relevant to recent efforts to increase effective linkage between addiction treatment and medical care for patients with substance use disorders (Institute of Medicine, 2006). These efforts are, in part, a response to the growing recognition of the frequency of co-occurring general health and substance use conditions (De Alba et al., 2004; Mertens et al., 2003) and unmet medical needs of patients engaged in addictions treatment (Saitz et al., 2004). Coordinated medical and addictions care, whether by referral or co-located, has the potential to engage patients in risk reduction for osteoporosis and to reduce the incidence of osteoporotic fractures.

In summary, the majority of individuals examined in this methadone maintenance treatment program had low BMD. Treatable conditions associated with low BMD were commonplace including heavy alcohol use and smoking. Whether addressing osteoporosis risk factors or using bone-preserving medications within MMT settings or via referral to medical care would reduce the incidence of fractures is an area that merits further study.

Acknowledgements

The authors appreciate the invaluable contributions of Suzette Levenson, Ph.D. of the Boston University School of Public Health Data Coordinating Center for data management. Support for Theresa Kim came from the National Institute Drug Abuse (R25-DA13582). Support for this study came from the following grants: the Boston University Department of Medicine Pilot Project Grant and the Clinical Research Feasibility Fund. This research was conducted in part at the General Clinical Research Center at Boston University School of Medicine, USPHS Grant MO1 RR00533.

References

Amiel, C., Ostertag, A., Slama, L., Baudoin, C., N'Guyen, T., Lajeunie, E., Neit-Ngeilh, L., Rozenbaum, W., De Vernejoul, M.C., 2004. BMD is reduced in HIV-infected men irrespective of treatment. J. Bone Miner. Res. 19, 402–409.

Arnsten, J.H., Freeman, R., Howard, A.A., Floris-Moore, M., Santoro, N., Schoenbaum, E., 2006. HIV infection and bone mineral density in middle-aged women. Clin. Infect. Dis. 42, 1014–1020.

Daniell, H.W., 2004. Opioid osteoporosis. Arch. Intern. Med. 164, 338.

De Alba, I., Samet, J.H., Saitz, R., 2004. Burden of medical illness in drug- and alcohol-dependent persons without primary care. Am. J. Addict. 13, 33–45.

Fink, H.A., Ensrud, K.E., Nelson, D.B., Kerani, R.P., Schreiner, P.J., Zhao, Y., Cummings, S.R., Nevitt, M.C., 2003. Disability after clinical fracture in postmenopausal women with low bone density: the fracture intervention trial (FIT). Osteoporos. Int. 14, 69–76.

Friedmann, P.D., Lemon, S.C., Anderson, B.J., Stein, M.D., 2003. Predictors of follow-up health status in the Drug Abuse Treatment Outcome Study (DATOS). Drug Alcohol Depend. 69, 243–251.

Gennari, L., Merlotti, D., Martini, G., Gonnelli, S., Franci, B., Campagna, S., Lucani, B., Dal, C.N., Valenti, R., Gennari, C., Nuti, R., 2003. Longitudinal association between sex hormone levels, bone loss, and bone turnover in elderly men. J. Clin. Endocrinol. Metab. 88, 5327–5333.

Hummer, M., Malik, P., Gasser, R.W., Hofer, A., Kemmler, G., Moncayo Naveda, R.C., Rettenbacher, M.A., Fleischhacker, W., 2005. Osteoporosis in patients with schizophrenia. Am. J. Psychiatry 162, 162–167.

Institute of Medicine, 2006. Improving the Quality of Health Care for the Mental and Substance-Use Conditions, The National Academies Press, Washington, D.C.

Izumotani, K., Hagiwara, S., Izumotani, T., Miki, T., Morii, H., Nishizawa, Y., 2003. Risk factors for osteoporosis in men. J. Bone Miner. Metab. 21, 86–90.

Johnell, O., Kanis, J.A., Oden, A., Sernbo, I., Redlund-Johnell, I., Petterson, C., De Laet, C., Jonsson, B., 2004. Mortality after osteoporotic fractures. Osteoporos. Int. 15, 38–42.

Licata, A., 2003. Osteoporosis in men: suspect secondary disease first. Cleve. Clin. J. Med. 70, 247–254.

Mertens, J.R., Lu, Y.W., Parthasarathy, S., Moore, C., Weisner, C.M., 2003. Medical and psychiatric conditions of alcohol and drug treatment patients in an HMO: comparison with matched controls. Arch. Intern. Med. 163, 2511–2517.

Mikhail, N., 2003. Hypogonadism and osteoporosis in men. Arch. Intern. Med. 163, 1237–1238.

Mondy, K., Yarasheski, K., Powderly, W.G., Whyte, M., Claxton, S., DeMarco, D., Hoffmann, M., Tebas, P., 2003. Longitudinal evolution of bone mineral density and bone markers in human immunodeficiency virus-infected individuals. Clin. Infect. Dis. 36, 482–490.

Mussolino, M.E., Jonas, B.S., Looker, A.C., 2004. Depression and bone mineral density in young adults: results from NHANES III. Psychosom. Med. 66, 533–537.

Neuner, J.M., Zimmer, J.K., Hamel, M.B., 2003. Diagnosis and treatment of osteoporosis in patients with vertebral compression fractures. J. Am. Geriatr. Soc. 51, 483–491.

Pedrazzoni, M., Vescovi, P.P., Maninetti, L., Michelini, M., Zaniboni, G., Pioli, G., Costi, D., Alfano, F.S., Passeri, M., 1993. Effects of chronic heroin abuse on bone and mineral metabolism. Acta Endocrinol (Copenh) 129, 42–45.

Perez-Castrillon, J.L., Olmos, J.M., Gomez, J.J., Barrallo, A., Riancho, J.A., Perera, L., Valero, C., Amado, J.A., Gonzalez-Macias, J., 2000. Expression of opioid receptors in osteoblast-like MG-63 cells, and effects of different opioid agonists on alkaline phosphatase and osteocalcin secretion by these cells. Neuroendocrinology 72, 187–194.

Port, L., Center, J., Briffa, N.K., Nguyen, T., Cumming, R., Eisman, J., 2003. Osteoporotic fracture: missed opportunity for intervention. Osteoporos. Int. 14, 780–784.

Rees, V.W., Horton, N.J., Hingson, R.W., Saitz, R., Samet, J.H., 2002. Injury among detoxification patients: alcohol users' greater risk. Alcohol. Clin. Exp. Res. 26, 212–217.

Rosen, H., Krichevsky, A., Bar-Shavit, Z., 1998. The enkephalinergic osteoblast. J. Bone. Miner. Res. 13, 1515–1520.

Saitz, R., Larson, M.J., Horton, N.J., Winter, M., Samet, J.H., 2004. Linkage with primary medical care in a prospective cohort of adults with addictions in inpatient detoxification: room for improvement. Health Serv. Res. 39, 587–606.

Sampson, H.W., 2002. Alcohol and other factors affecting osteoporosis risk in women. Alcohol Res. Health 26, 292–298.

SAS Institute Inc., 2001. Version 8.2 of the SAS System for Windows. Cary, NC, USA.

Schmittner, J., Schroeder, J.R., Epstein, D.H., Preston, K.L., 2005. Menstrual cycle length during methadone maintenance. Addiction 100, 829–836.

Turner, R.T., 2000. Skeletal response to alcohol. Alcohol. Clin. Exp. Res. 24, 1693–1701.

Videman, T., Batti, M.C., Gibbons, L.E., Vanninen, E., Kaprio, J., Koskenvuo, M., 2002. The roles of adulthood behavioural factors and familial influences in bone density among men. Ann. Med. 34, 434–443.

Wilson, J.F., 2004. New treatments for growing scourge of brittle bones. Ann. Intern. Med. 140, 153–156.

Wong, P.K., Spencer, D.G., McElduff, P., Manolios, N., Larcos, G., Howe, G.B., 2003. Secondary screening for osteoporosis in patients admitted with minimal-trauma fracture to a major teaching hospital. Intern. Med. J. 33, 505–510.

Woody, G., McLellan, A.T., O'Brien, C., Persky, H., Stevens, G., Arndt, I., Carroff, S., 1988. Hormone secretion in methadone-dependent and abstinent patients. NIDA Res. Monogr. 81, 216–223.

World Health Organization Scientific Group, 2003. Prevention and management of osteoporosis. Technical Report Series, No. 921, Geneva: World Health Organization.

BMC Health Services Research

BioMed Central

Research article

Open Access

Episodic homelessness and health care utilization in a prospective cohort of HIV-infected persons with alcohol problems

Theresa W Kim*[1], Stefan G Kertesz[2], Nicholas J Horton[3], Nicole Tibbetts[4] and Jeffrey H Samet[1,5]

Address: [1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA, [2]Division of Preventive Medicine, Department of Medicine, University of Alabama at Birmingham School of Medicine and the Deep South Center on Effectiveness and the Birmingham Veterans' Affairs Medical Center, Birmingham, AL, USA, [3]Department of Mathematics, Smith College, Northampton, MA, USA, [4]DM-STAT Inc., Medford, MA, USA and [5]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA

Email: Theresa W Kim* - theresa.kim@bmc.org; Stefan G Kertesz - skertesz@mail.dopm.uab.edu; Nicholas J Horton - nhorton@felix.smith.edu; Nicole Tibbetts - nicole.campagnoni@dmstat.com; Jeffrey H Samet - jsamet@bu.edu

* Corresponding author

Published: 27 February 2006

*BMC Health Services Research*2006, 6:19 doi:10.1186/1472-6963-6-19

Received: 14 July 2005

Accepted: 27 February 2006

This article is available from: http://www.biomedcentral.com/1472-6963/6/19

Abstract

Background: Because individuals with HIV/AIDS often have complex medical and social needs, the impact of housing status on medical service utilization is difficult to isolate from the impact of conditions that may worsen during periods of homelessness such as depression and substance abuse. We examine whether episodes of homelessness are independently associated with suboptimal medical utilization even when accounting for concurrent addiction severity and depression.

Methods: We used data from a 30-month cohort of patients with HIV/AIDS and alcohol problems. Housing status, utilization (ambulatory visits, emergency department (ED) visits, and hospitalizations) and other features were assessed with standardized research interviews at 6-month intervals. Multivariable longitudinal regression models calculated incidence rate ratios (IRR) comparing utilization rates during 6-month intervals (homeless versus housed). Additional models assessed whether addiction severity and depressive symptoms could account for utilization differences.

Results: Of the 349 subjects, 139 (39%) reported homelessness at least once during the study period; among these subjects, the median number of nights homeless per 6-month interview period was 30. Homelessness was associated with higher ED utilization (IRR = 2.17; 95% CI = 1.72–2.74) and hospitalizations (IRR = 2.30; 1.70–3.12), despite no difference in ambulatory care utilization (IRR = 1.09; 0.89–1.33). These associations were attenuated but remained significant when adjusting for addiction severity and depressive symptoms.

Conclusion: In patients with HIV/AIDS and alcohol problems, efforts to improve housing stability may help to mitigate intensive medical utilization patterns.

Background

Indigent populations are disproportionately affected by HIV infection. The prevalence of HIV infection is 5 to 9 times higher in urban poor populations than the general population [1,2]. Approximately one half of individuals with HIV/AIDS who receive services funded by the Ryan White Comprehensive AIDS Resources Emergency Act live below the Federal Poverty Level [3]. In the HIV Cost and Utilization Study (HCSUS), 46% reported an annual income less than $10,000 [4].

In part due to the overlap between poverty and unstable housing, many individuals with HIV infection experience homelessness. About one third of HCSUS participants reported a need for housing services with 39% of those needing housing services unable to access these services [5]. Among HIV-infected veterans, 32% have been homeless at some point in their lives [6] compared to 7.4% of the general population [7]. Prevalence estimates of homelessness in persons with HIV range from 6% to 27% among a New York State Medicaid population [8], clients at federally funded HIV clinics [3], and adults with substance abuse disorders [9]. In comparison, 3.1% of the general population sampled by random digit dialing has been homeless in the past 5 years [7].

Homelessness among HIV-infected persons is associated with a lower likelihood of receiving prophylaxis for opportunistic infections [10] and highly active antiretroviral therapy [11] as well as higher mortality rates [12]. Reduced access to effective therapies occurs in the context of higher rates of emergency department (ED) utilization [9,13] and hospital admissions [8,9]. Because indigent patients with HIV/AIDS often have complex medical and social needs, the impact of housing status is difficult to isolate from other conditions that increase the risk of homelessness and affect access to care such as addictions [9,14-16] and depression [17-19]. Previous studies on the impact of housing status on utilization among individuals with HIV infection have modeled homelessness as a permanent condition despite evidence that most individuals who are homeless have intermittent periods of stable housing [20]. In this study of the relationship between homelessness and health care utilization, we elected to treat homelessness as a state that could vary over time.

To the extent that homelessness might be associated with distinct patterns of health service utilization among HIV-infected individuals, relevant policy decisions ultimately must pivot upon clarification of which particular problems need targeting. That is, high cost utilization patterns could reflect worse HIV progression, greater addiction severity, impaired access to a usual source of ambulatory care, or potentially the impact of homelessness itself. Each of these associations might arguably call for somewhat different policy responses. Hence, a secondary interest of this study is to disentangle, in an exploratory way, contributing factors to utilization patterns.

Among HIV-infected individuals, alcohol misuse demarcates a group at special risk for adverse outcomes [6] including housing instability. We therefore used data from a cohort study of HIV-infected individuals with alcohol problems to examine the hypothesis that HIV-infected persons would utilize less ambulatory care and more emergency department and inpatient care during periods with an episode of homelessness, compared to periods without homelessness. Furthermore we set out to explore the independent contributions of relevant variables (addiction severity and depressive symptoms) with the expectation that these variables would substantially attenuate and potentially explain any apparent association between homelessness and service utilization patterns in this sample of HIV-infected persons with a history of alcohol problems.

Methods

Study population

We analyzed data from the HIV-Alcohol Longitudinal Cohort (HIV-ALC) study. The primary purpose of the HIV-ALC study was to prospectively examine the impact of alcohol use on HIV disease progression. Subjects in the HIV-ALC cohort receiving antiretroviral therapy were eligible to participate in a randomized controlled trial of an antiretroviral adherence intervention. A description of the patients in the HIV-ALC cohort [21] as well as the antiretroviral adherence intervention has been previously published [22]. Briefly, eligibility for the HIV-ALC study included endorsement of two or more positive responses to the CAGE questionnaire [23] or a physician co-investigator clinical diagnosis of lifetime alcohol abuse or dependence. 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. Other entry criteria included the following: fluency in English or Spanish, Mini-Mental State Examination score greater or equal to 21 [24], and no plans to move from the Boston area in the two years following the baseline assessment. All subjects had HIV infection confirmed either as part of clinical care or as part of the study. Subjects were recruited from the following sites: 56% from the Boston Medical Center HIV Diagnostic Evaluation Unit [25]; 17% from posted flyers; 13% from Boston Medical Center Primary Care Clinic; 5% from a medical respite facility for homeless persons; 4% from a methadone clinic; 4% from subject referrals; and 2% from Beth Israel Deaconess Medical Center.

Of 444 eligible subjects screened, 349 (79%) provided informed consent to participate in the study. Because

study subjects were recruited over a four-year period (1997 to 2001), and all follow-ups ceased in August 2001, time of recruitment was the major factor affecting the number of follow-up observations in this study ($P < .0001$) [26].

Those subjects receiving antiretroviral therapy at the time of recruitment participated in a randomized controlled trial to enhance adherence to antiretroviral therapy. The intervention consisted of three nurse visits over a 3-month period to problem-solve with the patient about ways to decrease missed doses. The intervention was not significantly associated with higher adherence [22].

The Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center approved this study.

Data collection

Subjects were interviewed up to 7 times over the study period, approximately 6 months apart, from 1997 to 2001. At each scheduled interview, trained research associates, using a standardized instrument in either English or Spanish, obtained information about housing status, medical service utilization, HIV risk behaviors, antiretroviral medication use, substance use, addiction severity, and depressive symptoms. The Spanish interview instrument used the standardized Spanish versions of scales when available; the remainder of the questionnaire was translated from English into Spanish, back translated to check for accuracy, and then corrected. CD4 cell counts and HIV RNA levels were collected, using existing laboratory tests if performed as part of clinical care within six months of the interview. When clinical samples were unavailable, the Boston Medical Center Clinical Laboratory evaluated blood samples collected for study purposes.

Outcome variables

Our three outcomes of interest were the number of self-reported ambulatory visits, the number of emergency department (ED) visits, and the number of hospitalizations in the 6 months prior to the research interview.

Main predictor variable

We used subjects' report of any night spent on the street or in a shelter in the past 6 months to indicate an episode of homelessness [27]. This was assessed with the survey question, "In the last 6 months, how many nights have you spent in an overnight shelter, or on the street, without shelter?" Sleeping in environments intended for temporary shelter, or in places not meant for sleeping, corresponds to the federal McKinney Act's definition of homelessness and approximates "literal homelessness"[28].

Other explanatory variables

We used the Behavioral Model for Vulnerable Populations [29] as a conceptual framework to help guide our choice of covariates for inclusion in the multivariable regression models explaining health service use. Predisposing factors (age, gender, race/ethnicity, housing status, substance abuse severity, and depressive symptoms) and relevant indicators of need for medical health service use (CD4 cell count, HIV RNA viral load, receipt of any antiretroviral therapy) were included in the models. Health insurance status, an enabling/disabling factor, was measured but not included in the models since 99% of all subjects had access to private, Medicaid or a special publicly-funded health insurance for medications, ambulatory and ED visits, and hospitalizations. Substance abuse severity was assessed with the alcohol and drug composite scores from the Addiction Severity Index (ASI-alc and ASI-drug, respectively), an assessment instrument with documented reliability and validity, each scored 0–1, with higher scores indicating increased severity [30]. Depressive symptoms were measured with the 20-item Center for Epidemiologic Studies Depression Scale (CES-D); scores ≥ 16 are considered to reflect significant depressive symptoms [31]. Participation in the intervention trial and study time point were also included as potential explanatory variables.

Analysis

Descriptive statistics (proportions, means, standard deviations) and univariate analyses were used to compare subjects by housing status at baseline. Categorical variables were compared using chi-square test and continuous variables with the two-sample t-test. We calculated the proportion of subjects experiencing homelessness over time using a Kaplan-Meier survival estimator. Since we asked about homelessness in the 6 months prior to each interview, time 0 was considered to be 6 months prior to the first interview, therefore the survival estimator calculated the proportion of subjects experiencing homelessness over 36 months.

To examine the association of homelessness and medical service utilization, we constructed separate multivariate longitudinal regression models for each outcome: ambulatory visits, ED visits, and hospitalizations. The unit of analysis for regression models was by observation (e.g. interview). Longitudinal regression models calculated incidence rate ratios (IRR) for each available 6-month observation period (homeless versus housed). Since serial measures on the same individuals were collected, generalized estimating equation (GEE) regression models[32] were used to adjust for the correlation between these measures over time. We used an empirical working variance estimator in these models and log link function (Poisson regression).

Table 1: Demographic and Clinical Characteristics of Study Cohort (N = 349)Stratified by Housing Status at Baseline *

Categorical measures, %	Homeless N = 101	Not homeless N = 248	p-value
Female	13	24	0.02
Race/ethnicity†			0.29
Black	38	47	
Hispanic	26	21	
White	37	32	
High school graduate	49	65	0.005
Health insurance‡	98	99.6	0.15
Prescribed antiretroviral medications§	47	64	0.001
Physical injury§	17	9	0.04
Physical or sexual abuse			
Lifetime	78	82	0.40
Recent §	14	7	0.05
Jail §	36	26	0.05
Substance use‖			
Alcohol	45	42	0.60
Cocaine	21	25	0.36
Heroin	16	8	0.04
Continuous measures, N [std]			
Age	40.8 [6.8]	40.5 [7.5]	0.70
CD4 cell count ¶	405 [291]	399 [273]	0.90
HIV RNA (log_{10}) ¶	47K [95K]	26K [68K]	0.05
Depressive symptoms **	27.2 [14.13]	20.2 [12.08]	< 0.0001
Addiction Severity Index: ††			
Alcohol composite score	0.25 [0.25]	0.18 [0.20]	0.009
Drug composite score	0.13 [0.12]	0.11 [0.10]	0.05

* The majority (56%)of study participants were recruited through from the Boston Medical Center HIV Diagnostic Evaluation Unit. The remaining from posted flyers 17%; Boston Medical Center Primary Care Clinic 13%; respite facility for homeless persons 5%; methadone clinic 4%; subject referrals 4%; and Beth Israel Deaconess Medical Center 2%.

† Categories are not mutually exclusive, so total number will >100%

‡ Private, Medicare, Medicaid, or special publicly funded health insurance for individuals with HIV-infection

§ Previous 6 months

‖Previous 30 days

¶ Laboratory tests collected as part of clinical care were used if performed within six months of the interview. When clinical samples were unavailable, the Boston Medical Center Clinical Laboratory evaluated blood samples collected for study purposes.

** Depressive symptoms were measured with the 20-item the Center for Epidemiologic Studies Depression Scale

†† Range of possible scores 0–1 with a higher score indicating worse addiction severity.

In order to focus on the statistical significance of the homelessness variable (any versus none), we used the Behavioral Model for Vulnerable Populations to build multivariate models with covariates, including predictor variables that proved to not be significant. The following variables were included in all models: age, gender, race/ethnicity (2 df), study time point (6 df), CD4 cell count, HIV RNA log_{10} viral load, receipt of antiretroviral therapy (any or none) and participation in the antiretroviral adherence intervention (adherence intervention group, control group, and not on antiretroviral medication; 2 df). To explore whether the statistical significance of homelessness was affected by inclusion of addiction severity and depressive symptoms in the equations, models were also constructed with variables for alcohol abuse severity (ASI-alc score), drug abuse severity (ASI-drug score), and any depressive symptoms (CES-D ≥ 16). All the predictor variables were allowed to vary with time except for age, gender, race/ethnicity, and intervention trial assignment group.

We performed a secondary analysis to examine whether there was a "dose-response" relationship between the number of nights homeless and utilization differences. We analyzed homelessness based upon the cumulative number of nights homeless in a 6-month observation period. The median number of nights homeless (i.e., 30) and interquartile range (7,90) were used to define the 5-level categorical homelessness variable.

All analyses were run using SAS statistical software version 8.2 [33].

Table 2: Multivariable Longitudinal Regression Results of Predictors of Medical Service Utilization Among Adults with HIV Infection and Alcohol Problems (n= 1045 observations)

	IRR† (95% CI)		
	Ambulatory	Emergency Room	Hospitalization
Age	1.01 (1.00, 1.03)	1.00 (0.98, 1.01)	1.02 (1.00, 1.04)
Female	1.44 (1.11, 1.87)*	1.50 (1.11, 2.03)**	1.64 (1.11, 2.42)*
Race/ethnicity			
Black	0.86 (0.68, 1.08)	0.99 (0.74, 1.31)	1.30 (0.87, 1.92)
Hispanic	0.63 (0.50, 0.77)***	0.81 (0.56, 1.18)	0.73 (0.44, 1.23)
White	1	1	1
Lower CD4 cell count ‡	1.05 (1.01–1.10)*	1.07 (0.98, 1.20)	1.10 (1.03, 1.16)**
HIV RNA viral load $\log_{10}$	1.00 (0.94, 1.05)	1.00 (0.94, 1.07)	1.01 (0.92, 1.10)
Antiretroviral medication use	1.06 (0.86, 1.31)	0.91 (0.72, 1.15)	1.00 (0.72, 1.39)
Adherence intervention randomization assignment			
Intervention group	1.28 (1.01, 1.62)*	1.11 (0.82, 1.51)	1.46 (0.99, 2.16)
Control group	1.32 (1.04, 1.68)*	0.88 (0.65, 1.19)	1.07 (0.73, 1.57)
Not a participant §	1	1	1
Alcohol addiction severity (ASI-alcohol)‖	1.92 (1.25, 2.94)**	1.13 (0.67, 1.93)	1.54 (0.69, 3.44)
Drug addiction severity (ASI-drug) ‖	1.47 (0.59, 3.64)	2.29 (0.87, 6.06)	4.38 (1.18, 16.33)*
Depressive symptoms (CES-D)	1.01 (1.00, 1.01)	1.02 (1.01, 1.03) **	1.02 (1.01, 1.03)***
Homelessness (any versus none)	1.10 (0.91, 1.32)	1.95 (1.55, 2.45)***	1.90 (1.41, 2.57)***

* P < 0.05
** P < 0.05
*** P < 0.005
† IRR: incidence rate ratio of medical service utilization for 6-month observation period in which homelessness is reported compared to intervals without homelessness. Separate multivariate regression models were constructed for each utilization outcome. Later study time point was associated with lower ambulatory visit utilization rates (df 6, P = 0.003) and lower hospitalization utilization rates (6 df, P = .05). Study time point was not significantly associated with emergency room utilization rate differences.
‡ Per 100 reduction in CD4 cell count
§ Subjects not receiving antiretroviral medications did not participate in the adherence intervention study.
‖ Per one point higher score

Results

Subject characteristics

Descriptive characteristics of the cohort (n = 349) stratified by housing status reported at the first (baseline) interview are presented in Table 1. Compared to housed subjects, more of the subjects reporting homelessness at baseline (101/349, 29%) were male, had not graduated from high school, and had recently been incarcerated, injured, or abused (either physically or sexually). In addition, homeless subjects endorsed significantly more depressive symptoms on the CES-D (27.2 vs. 20.5, P < .0001). While the mean CD4 cell count was not significantly different between homeless and housed subjects (405 vs. 399 cells/μl, respectively, P = .81), a lower proportion of the homeless subjects reported taking any antiretroviral medications in the previous 6 months compared to the housed (47% vs. 64%, P = .001).

No difference was found in the proportion of subjects who drank any alcohol. However, among those who drank any alcohol, the homeless reported higher alcohol consumption and alcohol abuse severity as reflected by drinks per day (5.5 vs. 1.6, P = .03) and ASI-alcohol composite score (P = .009). While no difference was found in cocaine use (21% vs. 25%, P = .36), a higher proportion of the homeless reported any heroin use (16% vs. 8%, P = .04), and higher drug abuse severity as measured by the ASI-drug composite score (P = .05).

Forty-two percent (148/349) of the cohort had a history of homelessness in the 5 years before entering the study (median duration 6 months, interquartile range 3 to 18 months). As mentioned previously, 29% of the cohort reported homelessness at the baseline interview. The median number of nights homeless in the 6 months before the baseline interview was 30 nights with interquartile range of 7 and 120 (possible range 0–180). By the end of 36 months of observation, 39% (136/349) of the study cohort reported homelessness at least once in the preceding 6 months.

Using all observations (n = 1045), medical service utilization during a 6-month period is summarized as follows (median, 75% quartile, range): ambulatory visits (4, 7, 0–180); ED visits (0, 1, 0–15); and hospitalizations (0, 0, range 0–10).

Table 3: Multivariable Longitudinal Regression Results of the Cumulative Number Nights Homeless and Medical Service Utilization (n= 1045 observations).

Number nights homeless	IRR‡ (95% CI) Ambulatory	Emergency Room	Hospitalization
0	1	1	1
1–7	1.24 (0.90, 1.71)	1.31 (0.80, 2.15)	1.07 (0.64, 1.80)
8–30	1.15 (0.84, 1.57)	1.49 (1.02, 2.19)*	1.83 (1.01, 3.32)*
31–120	1.27 (0.97, 1.67)	2.17 (1.54, 3.07) *	1.85 (1.17, 2.94)**
121–180	0.79 (0.61, 1.03)	2.65 (1.94, 3.61) ***	2.88 (1.95, 4.25) ***

† IRR: incidence rate ratio of medical service utilization for 6-month observations periods Separate multivariate regression models were used for each utilization outcome. All models include age, gender, race/ethnicity (2 df), CD4 cell, HIV RNA viral load $\log_{10}$, antiretroviral medication use, adherence intervention participation, presence of depressive symptoms (Center for Epidemiologic Studies Depression Scale), and addiction severity (Addiction Severity Index, alcohol and drug composite scores).

‡ Number nights homeless was defined by the survey question, "In the last six months, how many nights have you spent in an overnight shelter, on the street, without shelter?" Homelessness was categorized based upon the median nights homeless for the sample (i.e., 30) and interquartile range (i.e., 7, 120).

Medical service utilization and homelessness

Ambulatory visits

No significant difference was found in ambulatory visit utilization between homeless and housed periods in the multivariate longitudinal regression model (IRR 1.09; 95% CI 0.89–1.33). Adjusting for alcohol, drug abuse and depressive symptoms did not markedly change these findings (IRR 1.10; CI 0.91–1.32). Other factors associated with higher ambulatory visit utilization included: female gender (IRR 1.44; CI 1.11–1.87), less severe alcohol abuse (IRR 1.92; CI 1.25–2.94 per one point reduction in ASI-alcohol composite score), and lower CD4 cell count (IRR 1.05; CI 1.01–1.10 per 100 reduction in cell count/μl), earlier study time point (df 6, P = 0.003), and participation in the antiretroviral adherence intervention (df 2, P 0.04). Identifying as Hispanic, however, was associated with lower ambulatory visit utilization (IRR 0.63; CI 0.50, 0.77).

Emergency department

Homelessness was significantly associated with greater use of the ED (IRR 2.17; CI 1.72–2.74). This finding was slightly attenuated but remained significant when adjusting for alcohol, drug abuse and depressive symptoms (IRR 1.95; CI 1.55–2.45). As presented in Table 3, one to seven homeless nights were not associated greater use of the ED (IRR 1.31; 0.85–2.15). However, there were significant associations between a higher number of nights of homelessness (8–30, 31–120, and 121–180 nights) and ED utilization rates (IRR 1.49, 2.17, and 2.65, respectively). Other significant predictors of higher ED utilization were female gender (IRR 1.50; CI 1.11–2.03) and more depressive symptoms (IRR 1.02; CI 1.01–1.03 per one point increase in CES-D score).

Hospitalization

Homelessness was also significantly associated with inpatient hospitalizations (IRR 2.30; CI 1.70–3.12). This finding was attenuated but remained significant after adjusting for alcohol, drug abuse severity and depressive symptoms (IRR 1.90; CI 1.41–2.56). Similar to the analyses of ED utilization rates, one to seven days of homelessness were not associated with hospitalization rate differences. However, hospitalization rate differences were found with 8 to 30 nights homeless (IRR 1.85; 1.17–2.94), 31–120 nights (IRR 1.85; 1.17–2.94) and 121 to 180 nights (IRR 2.88; 1.95–4.25). Other factors significantly associated with higher hospital utilization rates included: lower CD4 cell count (IRR 1.10; CI 1.03–1.16 per 100 reduction in cell count/μl); worse drug abuse severity (IRR 4.38; CI 1.18–16.33 per 1 point increase in ASI-drug composite score), more depressive symptoms (IRR 1.02; CI 1.01–1.03 per one point increase in CES-D score), and earlier study time point (6 df, P = .05).

Discussion

In this prospective cohort study of individuals with HIV infection and alcohol problems, utilization of ED and hospital inpatient care was significantly higher during periods in which homelessness was experienced. In addition, greater ED and hospital inpatient utilization differences were found with more nights homeless. These utilization findings were not fully attributable to addiction severity or depressive symptoms.

This study's findings are consistent with prior findings that homelessness is associated with higher ED visits [9,13] and hospitalizations [8,34] in individuals with HIV infection. Prior work, however, was unable to disentangle the simultaneous effects of addiction disorders and depressive symptoms, both of which are known to predict physical functioning in homeless HIV-infected individuals [40] and hypothetically could contribute to higher ED and inpatient utilization. Increased use of ED and hospital inpatient services during homeless periods may have occurred for exposure-related conditions [35] or injuries

due to victimization [36]. Even though no difference was found in ambulatory utilization, it remains plausible that the daily struggle to meet basic subsistence needs may have been a barrier to accessing outpatient care in a timely manner (e.g. earlier in the course of an acute illness) [4,37] resulting in ED visits or hospitalizations.

It is important to note that this study modeled homelessness as a time varying "state" rather than a "trait". A subject could contribute to utilization incidence rates for the homeless cohort and subsequently to the housed comparison group if that subject did not experience any homelessness in another interview period. This approach suggests that the state of homelessness contributes to higher hospitalizations and ED utilization, rather than unique features of the "homeless" person. Although causality cannot be proven [38], the finding that a higher number of nights homeless was associated with greater ED and hospitalization differences suggests that homelessness contributes to these utilization differences.

A substantial minority of 29% reported homelessness during the 6 months prior to the baseline interview and 39% at any time during the study period. The proportion of homeless subjects in this study likely reflects selective inclusion of HIV-positive persons with a history of alcohol problems. This cohort's higher incidence of homelessness compared to published reports (6 to 11%) [3,8,13] results, in part, from this study's longitudinal study design, which was more likely to capture both the long-term homeless and persons homeless for short periods of time (i.e. the transiently and episodically homeless)[20]. Since a pattern of intermittent access to conventional housing is relatively common among homeless persons [39], cross-sectional studies tend to over-sample chronically homeless persons [7]. Perhaps more importantly, since many of the patients in this study were at-risk for homelessness for a variety of reasons (e.g., low income, recent incarceration, depressive symptoms, and alcohol or drug addiction) residential instability could have resulted from the worsening of just one of these factors for individuals without much of a safety net.

Although previous studies have documented less ambulatory care utilization among homeless persons [9,13], we did not find such a difference between homeless and housed periods. We postulate that this may have been due to a recruitment strategy that drew predominantly from an HIV intake clinic that facilitated primary care linkage [25] as well as access to health insurance in a state that aggressively expanded Medicaid during the study period.

There are several important implications for our findings. First, to the extent that an investment in housing and other services to prevent homelessness among HIV-infected persons would require a major allocation of public resources, the cost of not adopting such a strategy needs to be clarified [40]. Studies of service utilization by severely mentally ill homeless persons suggest that housing costs can be offset by savings realized from hospitals and jails, [41] however similar evaluations of housing HIV-infected persons are lacking.

Second, since this study occurred in a city with relatively generous ambulatory care services for the homeless, an even greater use of ED and inpatient hospital services might be expected in settings where ambulatory care for homeless persons is less accessible [42]. Furthermore, since homeless periods were not associated with lower ambulatory utilization in our sample (perhaps reflecting local supply of these services), it seems unlikely that further expansion of homeless ambulatory care programs would fully address excess ED and hospital utilization. To the extent that statistical adjustments for substance abuse severity minimally altered the effects of homelessness, expanded access to addiction treatment services alone may not be sufficient to mitigate intensive medical utilization patterns.

Our study has several limitations. While it would have been ideal to have a night-by-night account of when homelessness and medical service utilization occurred, our data collection precluded determination of whether homeless nights were concurrent with dates of medical service utilization. Additionally, our objective indicators of medical need (CD4 cell count, HIV viral load, and antiretroviral medication use) may not have fully encompassed physical health, since co-morbid conditions have been increasingly recognized for their impact on HIV-infected individuals [43] and may have contributed to utilization differences. Also, medical service utilization was determined by self-report. However, studies have found self-reported health care use to be a valid measure among HIV-infected individuals [44], homeless persons [45], and drug users [46]. Homeless persons are a heterogeneous population, yet we did not differentiate among subgroups that have varying utilization patterns such as the unsheltered or the chronically homeless [47,48]. Also, we examined a subsample of individuals with HIV/AIDS and unstable housing, namely those with a history of alcohol problems. Our findings may not apply to homeless, HIV-infected individuals without a history of alcohol problems. The data was taken from a randomized trial of an antiretroviral adherence intervention. However, participation in the intervention was not associated with ED or hospitalization differences. Finally, while we did not find differential utilization of ambulatory care services according to housing status, the number of ambulatory visits does not capture important information such as longitudinal provider continuity across visits and the provision of

integrated case management services. Both have been associated with lower utilization of ED and hospitalizations in other studies [4,49].

The strengths of this study include its careful assessment of alcohol, drug abuse, and depressive symptoms, information frequently missing in other research on utilization in the homeless HIV-infected. Moreover, its longitudinal nature allowed examination of the episodically homeless persons, a group underrepresented in cross-sectional homeless studies [50].

In summary, in HIV-infected persons with alcohol problems, homelessness was associated with higher utilization of ED and inpatient hospitalizations, despite no difference in ambulatory visit utilization. These utilization differences were not fully attributable to alcohol, drug abuse or depressive symptoms. Even a transient episode of homelessness has potentially costly implications for health care utilization among HIV-infected persons with a history of alcohol abuse.

List of abbreviations

ED: emergency department

IRR: incidence rate ratio

CI: confidence interval

HIV Cost and Utilization Study: HCSUS

HIV-Alcohol Longitudinal Cohort: HIV-ALC

ASI: Addiction Severity Index

ASI-alc: Addiction Severity Index, alcohol composite score

ASI-drug: Addiction Severity Index, drug composite score

CES-D: Centers for Epidemiologic Studies Depression Scale

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

Theresa W. Kim (TWK)

Stefan G. Kertesz, (SGK)

Nicholas J. Horton (NJH)

Nicole Tibbetts (NT)

Jeffrey H. Samet (JHS)

Conceptualization of study: TK, SGK, NJH, JHS

NJH designed the statistical analysis.

NT performed the statistical analysis.

TWK drafted the original manuscript.

All authors participated in data interpretation, critical revisions of the manuscript, and gave final approval of the submitted manuscript.

Acknowledgements

The authors wish to thank Seville Meli, MPH for assistance in all aspects of the study. The authors appreciate the contributions of the staff at DM-STAT for data management and statistical support.

Support for this study came from the following grants: 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); and R25-DA 13582 (Clinical Addiction Research and Education (CARE) Program). This research was conducted in part at the General Clinical Research Center, Boston University School of Medicine, USPHS grant M01-RR00533. Dr. Kertesz was supported by Career Development Award from National Institute of Drug Abuse (NIDA) K23-DA15487.

The funding agencies had no part in data analysis or interpretation, the decision to submit the manuscript for publication, or writing of the manuscipt.

Presented In part at the 25th annual meeting of the Society of General Internal Medicine, Vancouver, Canada, May 2003.

References

1. Robertson MJ, Clark RA, Charlebois ED, Tulsky J, Long HL, Bangsberg , Moss AR: **HIV seroprevalence among homeless and marginally housed adults in San Francisco.** *Am J Public Health* 2004, **94**:1207-17.
2. Culhane DP, Gollub E, Kuhn R, Shapner M: **The co-occurrence of AIDS and homelessness: results from the integration of administrative databases for AIDS surveillance and public shelter utilization in Philadelphia.** *J Epidemiol Community Health* 2001, **55**:515-20.
3. **Health Resources Services Administration's HIV/AIDS Bureau. The AIDS Epidemic and the Ryan White Care Act, Past Successes and Future Challenges, 2004–2005** 2004 [http://www.hab.hrsa.gov].
4. Katz MH, Cunningham WE, Fleishman JA, Andersen RM, Kellogg T, Bozzette SA, Shapiro MF: **Effect of case management on unmet needs and utilization of medical care and medications among HIV-infected persons.** *Ann Intern Med* 2001, **135**:557-65.
5. Katz MH, Cunningham WE, Mor V, Andersen RM, Kellogg T, Zierler S, Crystal SC, Stein MD, Cylar K, Bozzette SA, Shapiro MF: **Prevalence and predictors of unmet need for supportive services among HIV-infected persons: impact of case management.** *Med Care* 2000, **38**:58-69.
6. Conigliaro J, Madenwald T, Bryant K, Braithwaite S, Gordon A, Fultz SL, Maisto S, Samet J, Kraemer K, Cook R, Day N, Roach D, Richey S, Justice A: **The Veterans Aging Cohort Study: observational studies of alcohol use, abuse, and outcomes among human immunodeficiency virus-infected veterans.** *Alcohol Clin Exp Res* 2004, **28**:313-21.

7. Link BG, Susser E, Stueve A, Phelan J, Moore RE, Struening E: **Lifetime and five-year prevalence of homelessness in the United States.** *Am J Public Health* 1994, **84:**1907-12.
8. Smith MY, Rapkin BD, Winkel G, Springer C, Chhabra R, Feldman IS: **Housing status and health care service utilization among low-income persons with HIV/AIDS.** *J Gen Intern Med* 2000, **15:**731-38.
9. Masson CL, Sorensen JL, Phibbs CS, Okin RL: **Predictors of medical service utilization among individuals with co-occurring HIV infection and substance abuse disorders.** *AIDS Care* 2004, **16:**744-55.
10. Bangsberg D, Bamberger JD, Hecht FM, Robertson M, McKenzie M, Jankowski J, Moss AR: **Protease inhibitor access, use and self-reported adherence in HIV-infected homeless adults.** *J Gen Intern Med* 1997.
11. Moss AR, Hahn JA, Perry S, Charlebois ED, Guzman D, Clark RA, Bangsberg DR: **Adherence to highly active antiretroviral therapy in the homeless population in San Francisco: a prospective study.** *Clin Infect Dis* 2004, **39:**1190-1198.
12. Lieb S, Brooks RG, Hopkins RS, Thompson D, Crockett LK, Liberti T, Jani AA, Nadler JP, Virkud VM, West KC, McLaughlin G: **Predicting death from HIV/AIDS: a case-control study from Florida public HIV/AIDS clinics.** *J Acquir Immune Defic Syndr* 2002, **30:**351-58.
13. Arno PS, Bonuck KA, Green J, Fleishman J, Bennett CL, Fahs MC, Maffeo C, Drucker E: **The impact of housing status on health care utilization among persons with HIV disease.** *J Health Care Poor Underserved* 1996, **7:**36-49.
14. Palepu A, Horton NJ, Tibbetts N, Dukes K, Meli S, Samet JH: **Substance abuse treatment and emergency department utilization among a cohort of HIV-infected persons with alcohol problems.** *J Subst Abuse Treat* 2003, **25:**37-42.
15. Latkin CA, Forman-Hoffman VL, D'Souza G, Knowlton AR: **Associations between medical service use and HIV risk among HIV-positive drug users in Baltimore, MD.** *AIDS Care* 2004, **16:**901-8.
16. Pulvirenti JJ, Glowacki R, Muppiddi U, Surapaneni N, Gail C, Kohl B, Jezisek T: **Hospitalized HIV-infected patients in the HAART era: a view from the inner city.** *AIDS Patient Care STDS* 2003, **17:**565-73.
17. Belkin GS, Fleishman JA, Stein MD, Piette J, Mor V: **Physical symptoms and depressive symptoms among individuals with HIV infection.** *Psychosomatics* 1992, **33:**416-27.
18. Sambamoorthi U, Walkup J, Olfson M, Crystal S: **Antidepressant treatment and health services utilization among HIV-infected medicaid patients diagnosed with depression.** *J Gen Intern Med* 2000, **15:**311-20.
19. Padgett DK, Struening EL: **Influence of substance abuse and mental disorders on emergency room use by homeless adults.** *Hosp Community Psychiatry* 1991, **42:**834-8.
20. Phelan JC, Link BG: **Who are "the homeless"? Reconsidering the stability and composition of the homeless population.** *Am J Public Health* 1999, **89:**1334-38.
21. Samet JH, Horton NJ, Traphagen ET, Lyon SM, Freedberg KA: **Alcohol consumption and HIV disease progression: are they related?** *Alcohol Clin Exp Res* 2003, **27:**862-67.
22. Samet JH, Horton NJ, Meli S, Dukes K, Tripps 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.
23. Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ: **Screening for alcohol abuse using CAGE scores and likelihood ratios.** *Ann Intern Med* 1991, **115:**774-7.
24. Folstein MF, Robins LN, Helzer JE: **The Mini-Mental State Examination.** *Arch Gen Psychiatry* 1983, **40:**812.
25. 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.** *Arch Intern Med* 1995, **155:**1629-33.
26. Ehrenstein V, Horton NJ, Samet JH: **Inconsistent condom use among HIV-infected patients with alcohol problems.** *Drug Alcohol Depend* 2004, **73:**159-66.
27. Kertesz SG, Horton NJ, Friedmann PD, Saitz R, Samet JH: **Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification.** *J Subst Abuse Treat* 2003, **24:**197-207.
28. Rossi PH, Wright JD, Fisher GA, Willis G: **The urban homeless: estimating composition and size.** *Science* 1987, **235:**1336-41.
29. Gelberg L, Andersen RM, Leake BD: **The Behavioral Model for Vulnerable Populations: application to medical care use and outcomes for homeless people.** *Health Serv Res* 2000, **34:**1273-302.
30. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, Pettinati H, Argeriou M: **The Fifth Edition of the Addiction Severity Index.** *J Subst Abuse Treat* 1992, **9:**199-213.
31. Radloff L: **The CES-D scale: A self-report depression scale for research in the general population.** *Applied Psychological Measurement* 1977, **1:**385-401.
32. Liang K, Zeger SL: **Longitudinal data analysis using generalized linear models.** *Biometrika* 1986, **73:**13-22.
33. SAS/STAT: **Version 8.2 of the SAS System for Windows. (8.2). 2001.** Cary, NC, USA, SAS Institute Inc; 1999. Ref Type: Computer Program
34. Weissman JS, Cleary PD, Seage GR3, Gatsonis C, Haas JS, Chasan-Taber S, Epstein AM: **The influence of health-related quality of life and social characteristics on hospital use by patients with AIDS in the Boston Health Study.** *Med Care* 1996, **34:**1037-56.
35. Hwang SW, Lebow JM, Bierer MF, O'Connell JJ, Orav EJ, Brennan TA: **Risk factors for death in homeless adults in Boston.** *Arch Intern Med* 1998, **158:**1454-60.
36. Kushel MB, Evans JL, Perry S, Robertson MJ, Moss AR: **No door to lock: victimization among homeless and marginally housed persons.** *Arch Intern Med* 2003, **163:**2492-99.
37. Gelberg L, Gallagher TC, Andersen RM, Koegel P: **Competing priorities as a barrier to medical care among homeless adults in Los Angeles.** *Am J Public Health* 1997, **87:**217-20.
38. Hwang SW: **Is homelessness hazardous to your health? Obstacles to the demonstration of a causal relationship.** *Can J Public Health* 2002, **93:**407-10.
39. Goering P, Tolomiczenko G, Sheldon T, Boydell K, Wasylenki D: **Characteristics of persons who are homeless for the first time.** *Psychiatr Serv* 2002, **53:**1472-74.
40. Starr P: **The homeless and the public household.** *N Engl J Med* 1998, **338:**1761-63.
41. 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, **12:**107-62.
42. Laine C, Markson LE, Fanning TR, Turner BJ: **Relationship between ambulatory care accessibility and hospitalization for persons with advanced HIV disease.** *J Health Care Poor Underserved* 1999, **10:**313-27.
43. Kilbourne AM, Justice A, Weissman S, Rabeneck L, Rodriguez-Barradas M: **Prevalence of general medical comorbidity among HIV+ veterans: Is there a role for general medicine in HIV care?** *J Gen Intern Med* 2003, **16:**146.
44. Weissman JS, Levin K, Chasan-Taber S, Massagli MP, Seage GR3, Scampini L: **The validity of self-reported health-care utilization by AIDS patients.** *AIDS* 1996, **10:**775-83.
45. Gelberg L, Siecke N: **Accuracy of homeless adults' self-reports.** *Med Care* 1997, **35:**287-90.
46. Palepu A, Tyndall MW, Leon H, Muller J, O'Shaughnessy MV, Schechter MT, Anis AH: **Hospital utilization and costs in a cohort of injection drug users.** *CMAJ* 2001, **165:**415-20.
47. Desai MM, Rosenheck RA, Kasprow WJ: **Determinants of receipt of ambulatory medical care in a national sample of mentally ill homeless veterans.** *Med Care* 2003, **41:**275-87.
48. O'Toole TP, Gibbon JL, Hanusa BH, Fine MJ: **Utilization of health care services among subgroups of urban homeless and housed poor.** *J Health Polit Policy Law* 1999, **24:**91-114.
49. Laine C, Hauck WW, Gourevitch MN, Rothman J, Cohen A, Turner BJ: **Regular outpatient medical and drug abuse care and subsequent hospitalization of persons who use illicit drugs.** *JAMA* 2001, **285:**2355-62.
50. 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:**207-32.

Pre-publication history

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

http://www.biomedcentral.com/1472-6963/6/19/prepub

INTRODUCTION

Advancing the Field of Health Literacy

Sunil Kripalani, MD, MSc,[1] Michael K. Paasche-Orlow, MD, MA, MPH,[2] Ruth M. Parker, MD, FACP[1], Somnath Saha, MD, MPH[3,4]

[1]Emory University School of Medicine, Atlanta, GA, USA; [2]Boston University School of Medicine, Boston, MA, USA; [3]Portland Veterans Affairs Medical Center, Portland, OR, USA; [4]Oregon Health & Science University, Portland, OR, USA.

DOI: 10.1111/j.1525-1497.2006.00568.x
J GEN INTERN MED 2006; 21:804–805.

In 1992, the National Adult Literacy Survey (NALS) showed that 40% to 45% of adult Americans struggled with functional literacy tasks.[1] In early 2006, the results of the National Adult Assessment of Literacy indicated that the status of literacy in the United States remains much the same.[2] Over 90 million adults lack the literacy skills needed to effectively function in the present health care environment.[3]

Low health literacy is associated with less medical knowledge, infrequent receipt of preventive services, increased hospitalization and use of emergency care, and worse control of chronic diseases.[4] In 2004, the Institute of Medicine (IOM) called for additional research into the associations and consequences of low health literacy. Among the items highlighted by the IOM committee were the need to develop new measures of health literacy, approaches to improve health communication in diverse populations, interventions to reduce the negative health effects of low literacy, and ways to address health literacy in graduate medical education and the health system.[3]

This special issue of *JGIM* responds to the priority areas outlined in the IOM report by bringing together state-of-the-art research related to the role of literacy in health care. The research articles and commentaries in this issue will provide insight for clinicians, educators, researchers, administrators, and policy makers on addressing literacy in various health care settings.

Among the many excellent original manuscripts featured in this special issue are several important health outcome studies. Sudore and colleagues lead off the issue with their longitudinal analysis of the Health, Aging, and Body Composition Study, demonstrating for the first time that low health literacy is an independent risk factor for mortality. In a 2-year study of patients with asthma, Mancuso and Rincon show that literacy is associated with asthma outcomes and that this relationship appears to be mediated by other variables, particularly asthma knowledge. More work like this is needed, to clarify not only the extent to which literacy is associated with disease control, but also to provide a better understanding of the potential mechanisms behind such relationships.

In another longitudinal study, Lincoln and colleagues describe the association between literacy and the severity of depression among adults with alcohol and drug dependence. A particularly innovative report in this issue is a clinical trial by Weiss and colleagues, who randomized patients with depression and low literacy to receive adult literacy education as an adjunctive therapy. The groundbreaking findings of this study should serve as a call for greater collaboration between the medical and adult education communities.

Given the importance of literacy in chronic disease management, it is surprising how few previously published articles had investigated the relationship between literacy and medication use. Four research papers in this issue shed some light on this area. Studies by Kripalani, Fang, and Davis indicate that literacy is associated with patients' ability to correctly identify their medicines, understand medication dosing instructions, and understand prescription bottle warning labels, respectively. However, the relationship of literacy with medication adherence and clinical parameters is less clear. Fang found no association between literacy and adherence to warfarin therapy or anticoagulation levels. By contrast, Paasche-Orlow actually found *better* adherence and viral load suppression among patients with low literacy and HIV, contradicting findings of prior studies.[5,6] It is evident that more research is needed to clarify how literacy may impact various aspects of medication use and adherence.

Two articles address the important question of whether differences in literacy contribute to racial and socioeconomic disparities in health and health care. Using data from the NALS, Sentell found that differences in literacy helped explain disparities in disability and chronic illness by race and educational attainment. Using data from a large cohort study, Howard similarly found that different literacy levels contributed to disparities in health status but not in vaccination rates. These findings raise as many questions as they answer.

Other original research papers in this issue concern comprehension of informed consent, follow-up after an abnormal Pap smear, and the performance of health literacy screening questions. Baker offers a perspective on the meaning of health literacy and the challenges inherent in its measurement. Three commentaries have been included to highlight federal and other national activities, strategies for teaching about health literacy and clear communication, and guidance for health systems to adapt to the needs of low-literacy patients. Many other resources are catalogued online by the SGIM Health Literacy Interest Group.[7]

It has been a privilege, a pleasure, and a great learning experience to review the latest work in health literacy and compile this special issue of *JGIM*. We wish to thank all of

Address correspondence and requests for reprints to Dr. Kripalani: Division of General Medicine, Emory University School of Medicine, 49 Jesse Hill Jr Dr SE, Atlanta, GA 30303 (e-mail: skripal@emory.edu).

the authors who submitted their manuscripts for consideration, as well as Pfizer Global Pharmaceuticals for funding the publication of this issue through an unrestricted grant. We are indebted to the staff of the journal as well as the wonderful members of the Expert Review Panel who performed many of the manuscript reviews, anchoring the peer review process. (See front matter for a complete listing.) Finally, we thank Vice Admiral Richard H. Carmona, MD, MPH, the current United States Surgeon General, for his support of this issue and for making health literacy an integral part of his public health agenda. We hope you all enjoy the issue and include efforts to improve health literacy in your own clinical practice, research, and educational activities.

Dr. Kripalani is supported by a K23 Mentored Patient-Oriented Research Career Development Award (K23 HL077597). Dr. Saha was supported by awards from the Department of Veterans Affairs Health Services Research & Development Service Advanced Research Career Development Program and from the Robert Wood Johnson Foundation Generalist Physician Faculty Scholars Program.

REFERENCES

1. **Kirsh 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.
2. **Kutner M, Greenberg E, Baer J.** National Assessment of Adult Literacy (NAAL). A first look at the literacy of America's adults in the 21st century. Available at: http://nces.ed.gov/naal. Accessed May 2, 2006.
3. **Institute of Medicine.** Health Literacy. A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
4. **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;19:1129–39.
5. **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.
6. **Kalichman S, Ramachandran B, Catz S.** Adherence to combination antiretroviral therapies in HIV patients of low health literacy. J Gen Intern Med. 1999;14:267–73.
7. **Society of General Internal Medicine Health Literacy Interest Group.** Available at: www.sgim.org

Research Report

European Addiction Research

Eur Addict Res 2006;12:12–19
DOI: 10.1159/000088578

Co-Morbidity of Infectious and Addictive Diseases in St. Petersburg and the Leningrad Region, Russia

E.M. Krupitsky[a] E.E. Zvartau[a] D.A. Lioznov[a] M.V. Tsoy[a] V.Y. Egorova[a]
T.V. Belyaeva[a] T.V. Antonova[a] N.A. Brazhenko[a] Z.M. Zagdyn[a]
E.V. Verbitskaya[a] Y. Zorina[a] G.F. Karandashova[a] T.Y. Slavina[a]
A.Y. Grinenko[a] J.H. Samet[b] G.E. Woody[c]

[a]Scientific Research Center of Addiction and Psychopharmacology, St. Petersburg State Pavlov Medical University, St. Petersburg, Russia; [b]Department of Medicine, School of Medicine, Boston University, Boston, Mass., and [c]Department of Psychiatry, School of Medicine, University of Pennsylvania, Philadelphia, Pa., USA

Key Words
Alcoholism · Co-morbid diseases · Heroin addiction

Abstract
The Russian health care system is organized around specific diseases, with relatively little focus on integration across specialties to address co-morbidities. This organizational structure presents new challenges in the context of the recent epidemics of injection drug use (IDU) and HIV. This paper uses existing and new data to examine the prevalence of reported new cases of drug dependence (heroin) and HIV over time as well as associations between drug dependence and alcoholism, hepatitis B and C, and tuberculosis in the City of St. Petersburg and the Leningrad region. We found a sharp rise in reported cases of IDU beginning in 1991 and continuing until 2002/2003, followed by a sharp rise in newly reported cases of HIV. These rises were followed by a drop in new cases of HIV and drug addiction in 2002/2003 and a drop in the proportion of HIV-positve individuals with IDU as a risk factor. Infection with hepatitis B and C were common, especially among injection drug users (38 and 85%, respectively), but also in alcoholics (7 and 14%). Tuberculosis was more common in alcoholics (53%) than in persons with alcoholism and drug dependence (10%), or with drug dependence alone (4%). Though these data have many limitations, they clearly demonstrate that drug dependence and/or alcoholism, HIV, hepatitis, and tuberculosis frequently co-occur in St. Petersburg and the Leningrad Region. Prevention and treatment services across medical specialties should be integrated to address the wide range of issues that are associated with these co-morbidities.

Introduction

The Russian health care system is organized around specific diseases. Hence, alcoholism, injection drug use (IDU) and the infectious diseases commonly associated with these disorders are serious problems that are usually treated independently. There are separate hospitals to treat addictive diseases, separate hospitals for infectious diseases and separate hospitals for tuberculosis (TB), with relatively little focus on treating multiple disorders simultaneously – even though they may occur in the same pa-

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
www.karger.com/ear

George E. Woody, MD
Department of Psychiatry, University of Pennsylvania School of Medicine
Treatment Research Institute, 600 Public Ledger Building
150 S. Independence Mall West, Philadelphia, PA 19106 (USA)
Tel. +1 215 399 0980/ext. 112, Fax +1 215 399 0987, E-Mail woody@trescarch.org

tient. Issues concerning the treatment of drug dependence and its co-morbidities are important in Russia and elsewhere since IDU is a major risk factor for HIV in many parts of the world, particularly in the Russian Federation, where it is synonymous with heroin dependence [1–4]. Though there is a clear association between drug and alcohol dependence, hepatitis B and C [5], and TB [6], the degree to which one or more of these diseases co-occur has not been systematically evaluated in Russia. This project provides data on this issue by studying the rise in reported cases of drug dependence and HIV in the city of St. Petersburg (SP) and the Leningrad region (LR) during the initial stage of the HIV epidemic in the Russian Federation [7]. The study also attempts to evaluate the degree to which drug dependence and alcoholism co-occur with HIV, hepatitis, and TB. The approach was to cover a broad range of data in order to provide a reasonable approximation of the general situation in the city of SP and the LR and the implications for the progress of the epidemic as well as for prevention and treatment services. The institutional review boards at St. Petersburg State Pavlov Medical University and the University of Pennsylvania approved this study.

Methods

To determine these relationships we examined existing and new data, focusing on three issues: (1) The prevalence of reported cases of drug dependence and HIV in SP and the LR over time, and the proportion of persons within each population who have HIV and are injection drug users; (2) the prevalence of HBsAg, HBcAb, and HCV in a cross-section of persons with alcoholism and drug dependence, and (3) the prevalence of alcoholism and drug dependence in a cross-section of patients with HIV, hepatitis and TB. SP and the LR (a rural area around the city almost the size of Belgium) are two different administrative regions (municipalities) of Russia, with populations of 4.6 million and 1.5 million, respectively.

For issue 1 (prevalence of drug dependence and HIV over time) we examined three existing databases. First, the official statistics on the overall number of registered drug-dependent individuals and the number of first-time registered drug-dependent individuals as recorded in the registry of the Leningrad Regional Center of Addictions (LRCA) for the years 1991–2003. The procedure is as follows: when an individual is diagnosed by local addiction psychiatrists with an ICD-10 diagnosis of alcohol and/or drug dependence and referred to an addiction treatment center (or receives free outpatient care in the local addiction psychiatric office), he/she is registered at the LRCA and appears in the registry until at least 5 years of abstinence have been confirmed by regular visits to the local addiction psychiatrist. The addiction psychiatrist at the local outpatient office, or at the addiction center, can establish the diagnosis of alcohol and/or drug dependence. Anonymous care (mostly outpatient) for alcoholism and addiction (i.e. care that does not result in registration) is available in SP as well as in the LR. However, only a limited number of people can afford it.

Second, we looked at individuals who tested positive for HIV as recorded at the HIV/AIDS centers of SP and the LR in the years 1997–2003. The AIDS centers collect information about all newly reported cases from every hospital and outpatient facility in their area. Records from the SP HIV/AIDS center for 1999 were not available for technical reasons. Though HIV testing in Russia is voluntary, most individuals hospitalized in maternity clinics, surgical departments of general hospitals, infectious disease hospitals, sexually transmitted disease centers and addiction hospitals are routinely tested for HIV. There are also outpatient facilities in SP and the LR where anyone can get tested for HIV anonymously; the same service is offered to all in-patients in general hospitals. Those who test positive are registered in a central database, provided they have had a positive ELISA confirmed by Western blot. Confirmation by Western blot can be done only in the HIV/AIDS centers of each administrative region, where the central registries are kept. Information about the route of HIV transmission (e.g. injection drug use or sexual transmission) is available for most registered cases in both HIV/AIDS centers.

Third, we analyzed HIV test results in the hospital records of all patients treated at the addiction center of SP (1997–2001) and the corresponding addiction treatment center for the LR (1997–2003). The HIV data from the addiction center in SP for 2002 and 2003 were not available for technical reasons. Both centers are specialized hospitals that treat alcohol- and drug-dependent persons living in SP and the LR (600 and 280 beds, respectively). All patients treated at these centers during 1997–2003 underwent HIV testing as part of the standard intake procedure; this intake procedure did not include testing for hepatitis B and C for most patients. All patients were admitted to the centers voluntarily, usually on referral by addiction psychiatrists from outpatient offices located in the districts of SP or the LR, where the patients were living. Typically, patients came to the local outpatient addiction psychiatry office and followed the psychiatrist's referral for free care at the addiction center serving that area. The standard treatment program in both addiction centers consisted of three phases: inpatient detoxification (7–10 days) followed by inpatient rehabilitation (3–4 weeks) with follow-up care at an outpatient psychiatric office from which the initial referral was obtained. A diagnosis of alcohol or drug dependence was established prior to referral to the addiction center using a clinical interview routinely administered by trained staff psychiatrists working at both centers.

For issue 2 (prevalence of hepatitis B and C infections) we tested 503 alcohol-dependent patients and 206 patients with heroin dependence for HBsAg, HBcAb, and HCV. These patients were referred to the hospital of the LR addiction center for treatment of alcohol or drug dependence in 2002 and all were tested as part of their routine medical evaluation upon admission. Patients were selected at random and testing was done with ImmunoComb-II kits using a solid-phase enzyme immunoassay (Orgenis, Israel) [8]. LRCA staff psychiatrists confirmed the diagnosis of alcohol and/or drug dependence upon admission using ICD-10 criteria.

For issue 3 (prevalence of alcoholism and drug dependence in infected patients) we determined the prevalence of alcohol and drug abuse or dependence among patients with HIV, hepatitis and TB by assessing 201 patients with HIV and 201 with hepatitis B and/or C who were being treated in the Botkin infectious diseases hospital of SP in the years 2000 and 2001. We also assessed 160 patients with

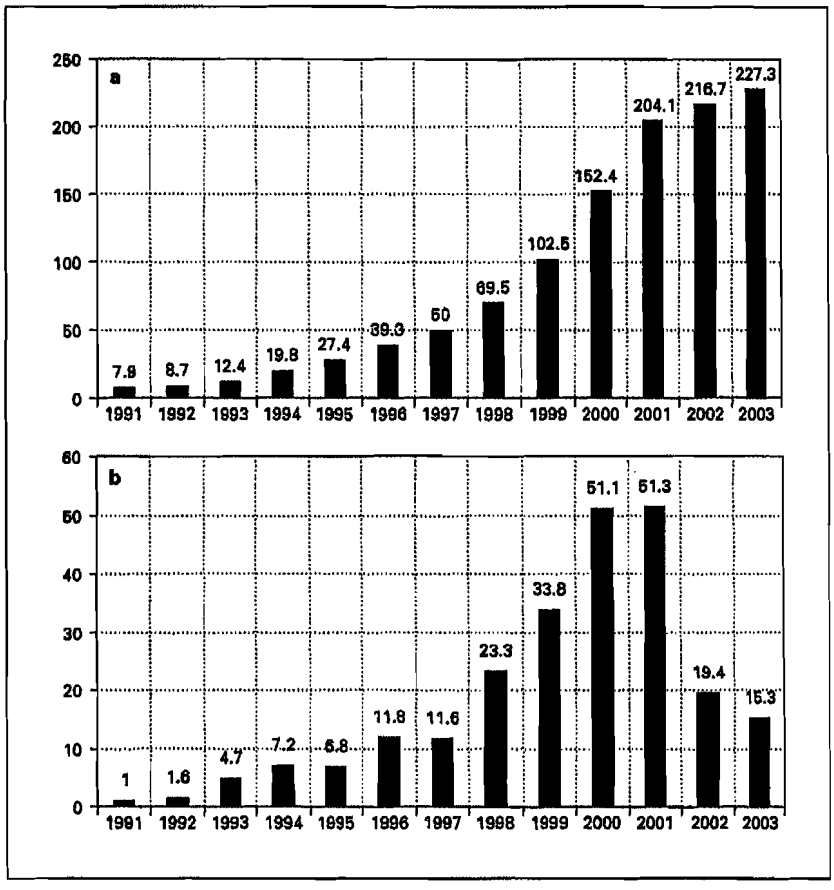


Fig. 1. Prevalence (**a**) and incidence (**b**) of drug dependence in the LR. **a** Total number of registered drug addicts per 100,000 of general population. **b** Number of first-time registered drug addicts per 100,000 of general population.

TB who were being treated in the city TB hospital in the years 2000 and 2001. All patients were selected using a random selection procedure and most of those who were asked agreed to participate. In particular, all 201 HIV and all 201 hepatitis patients agreed to participate. Of the TB patients, 160 subjects agreed and 6 refused to participate. Assessments were done by interview with a psychiatrist trained in addiction medicine who applied DSM-IV criteria for alcohol and drug abuse and dependence after obtaining informed consent from the patient. Data from the Michigan Alcoholism Screening Test [9], Addiction Severity Index [10] and Time Line Follow Back [11] supplemented the assessment for alcohol and drugs.

Results

Prevalence and Co-Occurrence of Drug Dependence and HIV over Time

Prevalence of Drug Dependence in the LR as Seen in Data from the LRCA. Between 1991 and 2003, reported cases of drug dependence increased more than 20-fold in the LR (fig. 1a). According to the LRCA registry, more than 90% of registered drug-dependent individuals were injection heroin users. Reported new cases (fig. 1b) increased 50-fold; however, there was a decrease in 2002 and 2003 compared to 2001.

Prevalence of HIV as Seen in Data from the HIV/AIDS Centers in SP and the LR. The number of registered HIV-infected persons in each area increased dramatically between 1997 and 2001 (fig. 2), though the number of first-time registered HIV-infected persons decreased in 2002 and 2003 compared to 2001.

IDU in Persons with HIV as Reflected in Data from the HIV/AIDS Centers. The proportion of injection drug users among first-time registered HIV patients in the LR increased from 67% in 1997 to 94% in 1999, then it decreased to 60% in 2003. Data were not available from the SP center for 1999; however, 86% of newly registered cases were IDUs in 2000, decreasing to 54% in 2003 (fig. 3).

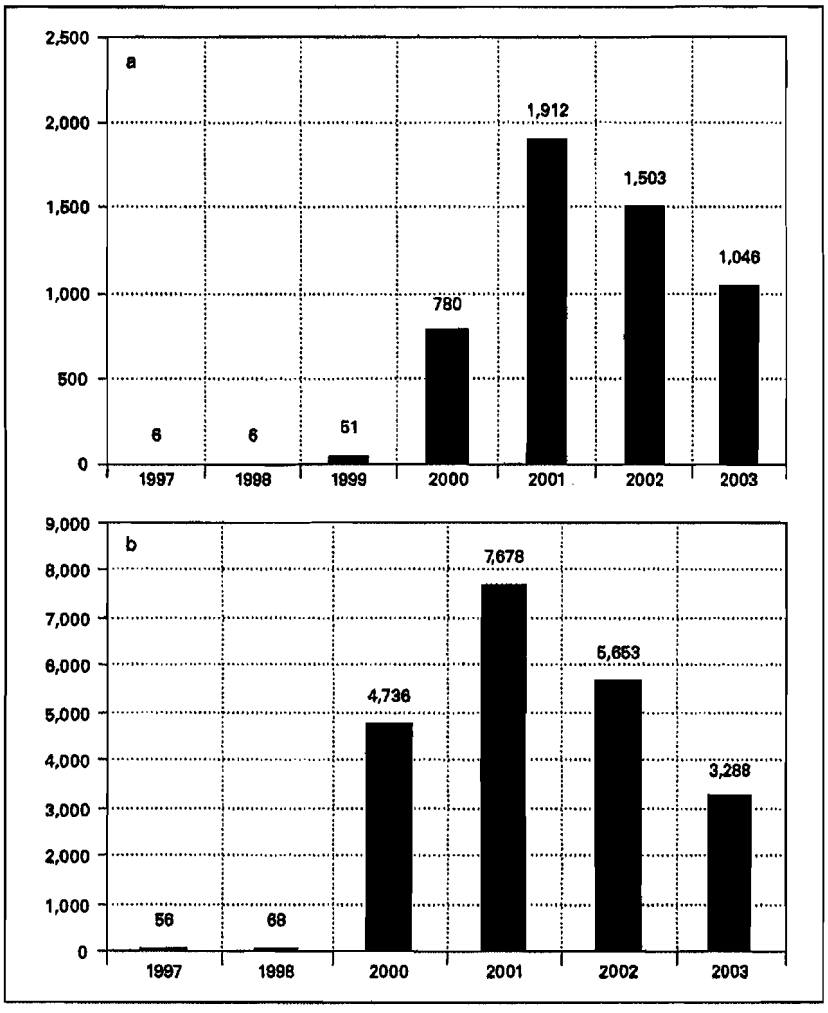


Fig. 2. Number of new HIV-positive individuals registered within a year in the LR (**a**) and SP (**b**).

Prevalence of HIV in Persons with Drug Dependence as Seen in Data from the LRCA and SP Addiction Centers. The LRCA treated 3,162 patients from 1997 to 2003. Their mean age was 23.3 (SE ± 1.0) years; 2,711 were male and 451 female. The percent with HIV increased from zero in 1997 to 40.0% in 2003. The SP addiction center treated 10,742 patients from 1997 to 2001. The mean age was 24.0 (SE ± 1.7) years; 8,357 were male and 2,385 female. The percent with HIV increased from 0% in 1997 to 18.7% by 2001 (fig. 4). The increase in HIV prevalence in 2002 and 2003 in the LRCA was lower than in the previous years.

Prevalence of HBsAg, HBcAb, and HCV in Persons with Alcoholism and Drug Dependence in the LRCA

HBsAg, evidence of active HBV infection, was found in 9% of those with drug dependence (mean age 24.5 ± 0.5 years; 172 males, 34 females) and 1% of those with alcohol dependence (mean age 42.3 ± 0.5 years; 442 males, 61 females). HBcAb was found in 38% of those with drug dependence and 7% of those with alcoholism. HCV was found in 85% of those with drug dependence and 14% of those with alcoholism.

Prevalence of Alcoholism and Drug Dependence in Persons with HIV, Hepatitis and TB

Among 201 patients with HIV (mean age 26.6 ± 0.6 years; 125 males, 76 females) assessed in the Botkin In-

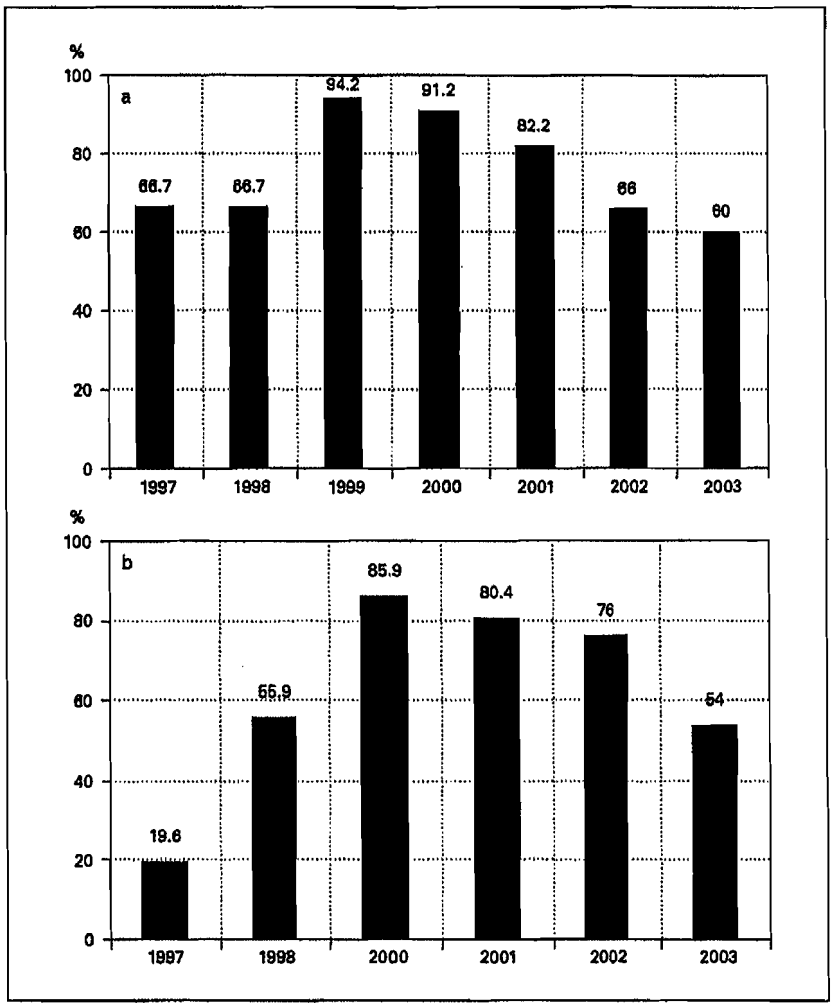


Fig. 3. Prevalence of injection drug users among new HIV-positive individuals registered within a year in the LR (**a**) and SP (**b**).

fectious Diseases Hospital, 47% were drug abusers or addicts; 38% were drug abusers or addicts *and* alcohol abusers or addicts; 10% had alcohol abuse or dependence; and only 5% showed no evidence of alcohol or drug abuse or dependence [12].

Of the 201 patients who tested positive for HBV and/or HCV (mean age 29.0 ± 0.9 years; 162 males, 39 females) and were treated in the Botkin hospital, 12.4% suffered from drug abuse or dependence; 18.5% suffered from drug abuse or dependence *and* alcohol abuse or dependence; 26.8% suffered from alcohol abuse or dependence, and 42.3 % did not have a substance use disorder. Drug abuse or dependence was seen in 7% of patients with hepatitis B, 40% of patients with hepatitis C, and 48% of patients with hepatitis B and C. A diagnosis of alcohol abuse or dependence was established in 24% of patients with hepatitis B, 34% of patients with hepatitis C, and 19% of patients with hepatitis B and C.

Among the 160 patients treated in the TB hospital of SP (mean age 39.8 ± 1.0 years; 109 males, 51 females), 4.3% were drug abusers or addicts; 10.1% were drug abusers or addicts *and* alcohol abusers or addicts; 53.1% were alcohol abusers or addicts, and 32.5 % did not have a substance use disorder.

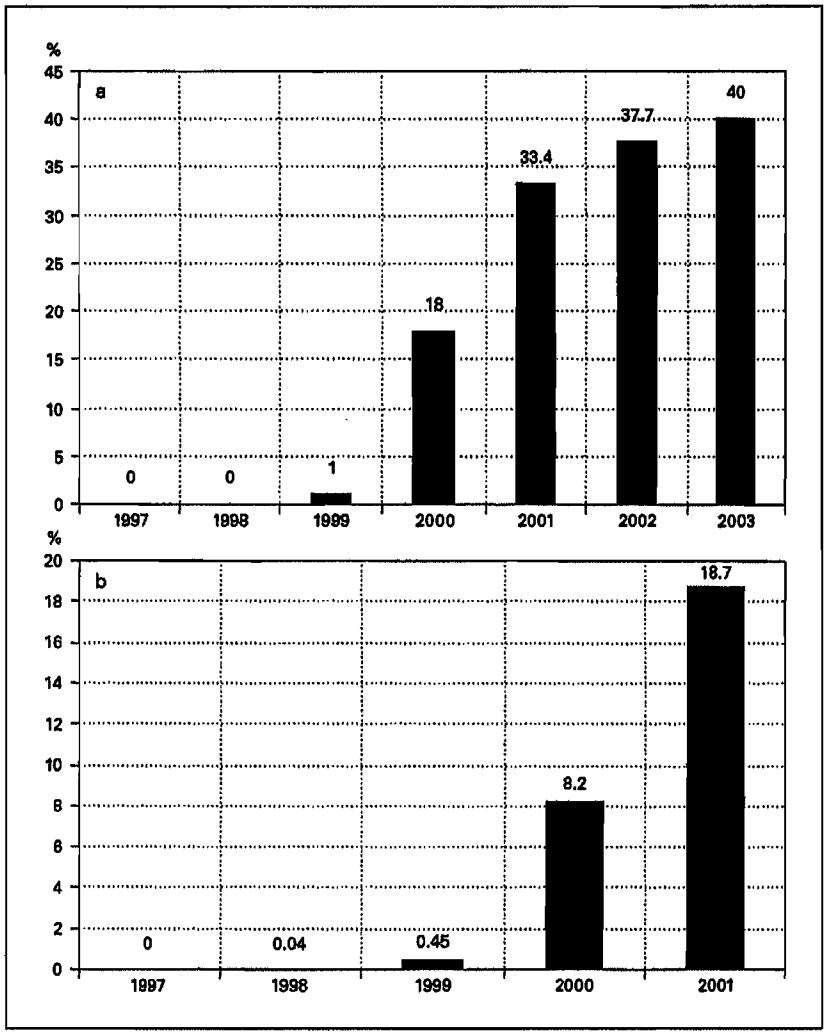


Fig. 4. Prevalence of HIV-positive individuals among drug dependent individuals treated within a year in the addiction hospitals of the LR (**a**) and SP (**b**).

Discussion

These data clearly show a high co-morbidity between substance use disorders, HIV, hepatitis and TB in the northwest region of Russia. The high co-morbidity between IDU, HIV and hepatitis is consistent with findings from many studies [5, 13, 14], as is the co-morbidity between alcoholism and TB [6]. Unlike data from the USA and Africa, where unprotected sex has been the main way to propagate HIV, these data clearly document that the HIV epidemic in SP and the LR has been almost entirely associated with injection drug (heroin) use. A similar situation is taking place in other cities of the Russian Federation [15], where more than 90% of all new cases of HIV in 1998 and 1999 were registered among injection drug users [1, 16].

The drop in new cases of IDU and HIV that began in 2002/2003 is, we think, related to a decreasing popularity of heroin among young people in the region, combined with a decrease in drug trafficking (associated with less availability of heroin). We have no other hypotheses since there were no significant changes in health care policies, logistics, and the overall opportunity to be tested for HIV or treated for alcoholism and drug dependence over these time periods. The drop in the proportion of injection drug users with HIV in 2001 may reflect a new stage of the epidemic, with the virus spreading into the general population through sexual contact. This latter possibility em-

phasizes the importance of an aggressive education and prevention program to better inform the general public about the virus and how to prevent its spread. So far, no such program exists. We should also mention that substitution therapy (methadone and buprenorphine) is forbidden by law in Russia. Changing this law to permit maintenance therapy is one option to explore in the fight to prevent HIV and hepatitis B and C infections.

The very high rate of hepatitis C in patients with drug dependence is extremely likely to be caused by sharing injection equipment, given that injection is the most common way of heroin consumption in SP and the LR. The problem of hepatitis C infections is likely to become very significant over the next 10–20 years for several reasons: it becomes chronic in 75% or more of those infected; many patients with HIV also have hepatitis C, which accelerates the course of AIDS; alcohol use is a long-standing problem in Russia and accelerates hepatitis C-related liver damage, and hepatitis C-related cirrhosis and liver failure occur in about 10% who are chronically infected. We did not have the resources to genotype the hepatitis C virus to determine if it was the type most likely to respond to antiviral therapy. However, the costs of such treatment, or of liver transplants for patients with end-stage liver disease, are likely to be another very serious problem in the future.

The prevalence of HBV and HCV infection in alcoholics was also quite high, which might be related to the higher rate of unsafe sexual contacts in persons with alcohol dependence [17, 18]. Alcohol- and other substance-related, unprotected sexual contacts are likely to contribute to the spread of HIV and hepatitis from substance abusers into the general population. This may be occurring already, as suggested by the data showing a high proportion of patients in the LRCA who have been infected with hepatitis B and C but have no history of a substance use disorder. These data indicate that education about how HIV and hepatitis B and C is spread, and widespread vaccination against hepatitis B of persons at high risk (especially those with drug and alcohol dependence and their sexual contacts) could be a very important public health measure. So far, no such program exists, mainly due to concerns about compliance with the vaccination regimen and financial constraints. Finally, the high prevalence of alcohol dependence among patients with TB indicates that assessment and treatment of alcohol dependence should be part of the standard care in a TB hospital since it appears that alcoholism puts individuals at an especially high risk of TB.

These data, taken together, indicate that substance use disorders and HIV, hepatitis B and C, and TB are parallel, overlapping epidemics. The impact of alcohol and drug use on initial infection and subsequent medical care are key issues, as is the impact of infectious diseases on the recovery of persons with alcohol and drug dependence. Education of the general public about how these diseases are transmitted is urgently needed, along with prevention and treatment programs for both substance use disorders and infectious diseases that cross the lines of specific disease boundaries. The Russian health care system may need re-organizing to address these co-morbidities.

Acknowledgments

This work was supported by NIH grants P30-AI 045, P60-DA 051861, U10-DA 13043 and K05-DA 17009 (G.W.), and the Department of Veterans Affairs. We would like to thank Daniel Langleben, MD, who reviewed the Russian contents used in the diagnostic interview parts of this study for compliance with the English versions.

References

1 Adelekan M: Reviewing the evidence of IDU epidemics. Proceedings of the Global Research Network on HIV Prevention in Drug-Using Populations: HIV and Drug Use. The Global Situation. Bethesda: National Institute on Drug Abuse, 2003, pp 12–20.

2 Dehne KL, Grund JP, Khodakevich Y, Kobyshcha Y: The HIV/AIDS epidemic among drug injectors in Eastern Europe: Patterns, trends and determinants. J Drug Issues 1999; 29:729–776.

3 Federal AIDS Centre of Russia: HIV Infection. Information Bulletin No. 20. Moscow: Federal AIDS Centre of Russia, 2001 (in Russian).

4 Grassly NC, Lowndes CM, Rhodens T, et al: Modeling emerging HIV epidemics: The role of injection drug use and sexual transmission in the Russian Federation, China, and India. Int J Drug Policy 2003;14:25–43.

5 Garfein R, Vlahov D, Galai N, et al: Viral infections in short-term drug users: The prevalence of the hepatitis C, hepatitis B, human immunodeficiency, and human T-lymphotropic viruses. Am J Public Health 1996;86:655–661.

6 Farmer PE, Kononets AS, Borisov SE, Goldfarb A, Healing T, McKee M: Recrudescent tuberculosis in the Russian Federation. Global Impact of Tuberculosis. Cambridge, MA: Harvard Medical School/Open Society Institute, 1999, pp 41–83.

7 Krupitsky E, Zvartau E, Karandashova G, Horton NJ, Schoolwerth KR, Bryant K, Samet JH: The onset of HIV infection in the Leningrad Region of Russia: A focus on drug and alcohol dependence. HIV Med 2004;5:30–33.
8 Herrera JL: Serological diagnosis of viral hepatitis. S Med J 1994;87:677–684.
9 Allen JP (ed): Assessing Alcohol Problems: A Guide for Clinicians and Researchers. Bethesda, NIAAA, 1995.
10 McLellan AT, Luborsky L, Woody GE, O'Brien CP: An improved diagnostic evaluation instrument for substance abuse patients: The Addiction Severity Index. J Nerv Ment Dis 1980; 168:26–33.
11 Sobell LC, Sobell MB: Timeline follow-back: A technique for assessing self-reported alcohol consumption; in Litten R, Allen J (eds): Measuring Alcohol Consumption. Totowa, Human Press, 1992, pp 41–72.
12 Krupitsky E, Zvartau EE, Lioznov D, Horton NJ, Williams ES, Samet JH: Alcohol use and HIV risk-behaviors among HIV-infected inpatients in St. Petersburg, Russia. Alcohol Clin Exp Res 2003;27(suppl):69.
13 Cohn JA: HIV-1 infection in injection drug users. Infect Dis Clin North Am 2002;16:745–770.
14 Sorensen JL, Mason CL, Perlman DC: HIV/hepatitis prevention in drug abuse treatment programs. Science and Practice Perspectives, NIDA 2002;1:4–12.
15 Fedotova TT, Zemerov VB, Efimova OS: Development of HIV infection epidemic among intravenous drug users in the Sverdlovsky region (in Russian). Zhurnal Mikrobiol EpidemiolImmunobiol 2003:86–89.
16 Ball A, Stimson GV, Kobyshcha Y, Fitch C, Pokrovsky V: HIV infection associated with drug injecting in the newly independent states, Eastern Europe: The social and economic context of epidemics. Addiction 1999;94:1323–1336.
17 Cooper ML: Alcohol use and risky sexual behavior among college students and youth: Evaluating the evidence. J Stud Alcohol Suppl 2002;(14):101–117.
18 Scheidt DM: HIV risk behavior among alcoholic inpatients before and after treatment. Addict Behav 1999;24:725–730.

The American Journal of Drug and Alcohol Abuse, 32: 435–452, 2006

ISSN: 0095-2990 print/1097-9891 online
DOI: 10.1080/00952990600753958

informa
healthcare

Emergency Department and Hospital Utilization Among Alcohol and Drug-Dependent Detoxification Patients without Primary Medical Care

Mary Jo Larson, Ph.D., M.P.A.,[1] Richard Saitz, M.D., M.P.H.,[2,3] Nicholas J. Horton Sc.D.,[4] Christine Lloyd-Travaglini M.P.H.,[5] and Jeffrey H. Samet, M.D., M.A., M.P.H.[2]

[1]New England Research Institutes, Watertown, Massachusetts, USA
[2]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts, USA
[3]Department of Epidemiology and Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts, USA
[4]Department of Mathematics, Smith College, Northampton, Massachusetts, USA
[5]Data Coordinating Center, Boston University School of Public Health, Boston, Massachusetts, USA

Abstract: Utilization of emergency department (ED) services and hospitalization among a cohort of substance abusers are described based on structured research interviews with 470 adults without primary care admitted to an urban residential detoxification program. Cross-sectional analysis of baseline data of subjects found nearly 19% of subjects went to an ED on 2 or more occasions in the 6 months prior to detoxification and 14% were admitted for an overnight hospitalization. Upon further analysis of past 6-month ED utilization, the following factors were independently associated with increased odds of ED use: White race; at least one month homeless in the past 5 years chronic health condition; injury in past 6 months; and subject perception that their substance abuse interfered with seeking care from a regular doctor. Subjects with cocaine as a primary problem had lower odds of ED utilization than a reference group with alcohol as a primary problem.

Address correspondence to Mary Jo Larson, New England Research Institutes, 9 Galen St., Watertown, MA, 02472, USA. E-mail: mjlarson@neriscience.com.

Keywords: Alcoholism, access, detoxification, drug abuse, emergency department

BACKGROUND

Every day, over 23,000 persons receive substance abuse care in US detoxification centers (1). Admissions to detoxification services, while brief, are an important source of care for substance abuse problems; detoxification admissions exceed the number of residential treatment and hospital stays for substance abuse. Twenty-two percent of alcohol admissions, 46% of heroin admissions, and 22% of cocaine admissions were for detoxification care in 1999. Receipt of detoxification care represents an important opportunity to address not only alcohol and drug problems but also to assess the need for other medical and social services.

Despite the widespread use of detoxification care among persons with alcohol and drug problems, limited information exists on the types of medical care that substance abusers seeking detoxification receive and the factors that lead to such medical care use among these individuals. For some, detoxification may be the only care they seek. Onsite health services are not routinely incorporated into drug abuse treatment programs (2). Of particular interest is the group of substance abusers without any identified source of regular medical care. These persons may be using services in an uncoordinated fashion without any ongoing clinical management (3–5).

There are medical and social reasons that may result in substance abusers placing high demands on the medical system. Emergency departments (EDs) may be perceived by vulnerable populations as the most accessible source of medical care, even for non-urgent reasons (6). Cocaine users present to EDs with nontraumatic chest pain (7), respiratory complications (8), and other adverse health consequences (9, 10). Heroin users present with bacterial endocarditis (11), cellulitis, pneumonia (12), and overdose (10). Chronic illicit drug users are 30% more likely to use ED care than occasional or non-drug using persons (6, 13). Psychiatric patients with comorbid substance use disorders have 2.8 times the odds as other psychiatric patients of visiting the ED repeatedly (14). Additionally, alcohol abusers among the mentally ill are at increased risk of hospitalization (15), psychiatric rehospitalization (16), and are heavy medical service utilizers (17). It is likely that some of the increased utilization is associated with circumstances resulting from serious health problems among drug users rather than simply inappropriate use (18–22).

The present study is a secondary analysis of baseline data collected in a randomized trial of a multidisciplinary intervention to link adults

in a residential detoxification unit to primary medical care (21). The intervention involved evaluation in a clinic onsite at the detoxification unit including a referral to primary care (23). Given the focus of the trial, and a desire to understand a group receiving the least coordinated care, the study enrolled only those detoxification patients who had no primary care physician. The primary outcome of interest of the randomized trial was attendance at an off-site primary care appointment within 12 months of detoxification care. Prior studies of medical care use among substance abusers have either excluded or not commented on this population of adults seeking detoxification care without established primary medical care.

The present study sought to establish baseline rates of intensive medical care (hospitalization and emergency department use) among adults with no regular physician with acute alcohol and drug use problems. In addition, we sought to determine factors associated with increased ED utilization. We hypothesized that utilization of emergency and inpatient medical services was high for this population and that even accounting, in part, for medical need, addiction related factors and self-perceived barriers to routine medical care would be associated with increased use of these services prior to detoxification.

METHODS

Subjects

All subjects were recruited from a single residential detoxification unit in the Boston, Massachusetts area between February 1997 and April 1999. Patients who met the following inclusion criteria were invited to participate in the study after they had been in the unit for a minimum of 24 hours: 1) alcohol, heroin, or cocaine was the first or second drug of choice; 2) age greater than 17 years; 3) present or planned residence in the study institution's catchment area where primary care was to be arranged, or homelessness. The following exclusion criteria were assessed at the time of initial screening for study recruitment: 1) an established relationship with a primary care physician that the patient intended to continue (self-report); 2) mental status deficiencies making the subject unable to provide pertinent history or informed consent; 3) specific plans to leave the Boston area in the next 12 months; 4) inability to provide three contact names; 5) pregnancy; and 6) not fluent in English or Spanish. Mental status deficiencies were assessed by administering the Mini-Mental State Examination (24). Individuals scoring less than 21 of 30 points on this standard test of cognition were excluded from the study. Further description of the study methods has been previously

reported (21). The study was approved by the Institutional Review Board at the study's institution and a certificate of confidentiality was secured from the federal government to further protect subjects' privacy.

Subject Assessment at Baseline

After enrollment, all subjects received an interviewer-administered baseline assessment, typically lasting one hour, from a trained research associate. The baseline instrument included the following instruments or assessed the following issues: demographics; homelessness (had at least one month in homeless shelter in past 5 years); mental and physical health status in the past 4 weeks (the SF-36 Health Survey) (25); chronic and episodic medical problems (26); substance use history and current problems as measured by the Addiction Severity Index (27); health care utilization; self-perceived barriers to accessing medical services from a regular doctor from a list of potential barriers; and health insurance anytime in the past 6 months. Health insurance included Medicaid, Medicare, health insurance from a job or family member's job, or any other health benefits plan that paid for medical care expenses.

Measurement of Utilization and Barriers

Data reported by subjects were used to determine ED and hospital utilization in the 6-month period prior to the interview. The specific questions were: "In the last 6 months, did you go to a hospital emergency room for medical care?" and "How many times in the last 6 months did you go to a hospital emergency room?" Patients were further asked to list the names of the EDs visited. Patients were asked: "In the last 6 months, have you been a patient in a hospital, overnight or longer? (not including a detoxification program or an emergency room stay only)". Patients were further asked how many times and how many nights in the last 6 months and the names of the hospitals. Hospitalizations included both medical and psychiatric. Subjects were shown a card with a list of 12 statements about reasons for not having a regular doctor and asked for a "yes/no" response to each response to the question "Which of the following statements, if any, are reasons why you don't have a regular medical doctor?" Statements read to subjects included payment barriers ("I cannot pay for services"), system/structural barriers ("The office or clinic hours are inconvenient"), help-seeking attitudes or beliefs ("I'm afraid others will find out about a health problem I have"), social disadvantage/hardship ("I do not have a babysitter or someone to watch my child"), and substance abuse ("My substance abuse interferes..."). This list was generated by the study

authors based on experience in clinical practice and participation in Substance Abuse and Mental Health Services Administration's (SAMHSA) cooperative agreement study on managed care for vulnerable populations. Similar system factors (28) and attitudinal factors (29) have been reported as access barriers to needed medical and mental health care.

Data Analysis and Conceptual Framework

Descriptive statistics were used to describe hospitalization and ED utilization in the 6 months prior to detoxification and study enrollment. Chi-square analysis of barrier statements compared the perceptions of subjects with insurance (primarily Medicaid) to subjects without insurance. Two or more visits to the ED was chosen for the outcome in the regression analyses to try to focus on a pattern of ED use that was potentially preventable or perhaps an indicator of over-reliance on episodic care.

Exploratory analysis assessed the bivariate association of theoretically relevant demographic, health status and other risk factors for repeat (>2) ED visits. The independent variables explored with these analyses were conceptualized focusing on the individual and vulnerability determinants of the behavioral model of health care utilization (30–32). Predisposing determinants included demographic characteristics. Enabling determinants included subjects' perceived barriers to having a regular doctor, insurance status, and unemployment. Need/illness determinants had two components: (1) vulnerability factors including substance use and homelessness, and (2) injury and health status. Since serious injuries and drug overdoses were common reported problems, inclusion of these variables in an analysis of utilization need was adopted.

The results from these bivariate analyses were used to reduce the list of variables included in the subsequent multivariable analyses. Variables with nonsignificant bivariate associations with theoretical importance that did not contribute to the final analysis include indicators of jail time in the past 6 months and lifetime injection drug use status. The significance of differences between the subject groups with and without repeat ED visits was determined by the chi-square statistic, or t-tests in the case of continuous variables.

To create a parsimonious multivariable model, when two or more variables were highly correlated (Pearson's correlation, $r > .40$), the authors chose the most clinically relevant factor for inclusion in subsequent analyses. To identify the relative importance of different contributing factors, a series of logistic regressions incorporating 4 groups of variables was used to predict 2 or more ED visits in the prior 6 months. The first group, demographic, and predisposing variables, included

gender, age, and an indicator for white race. Enabling variables, the second group entered, included having insurance, being unemployed, and five indicator variables for barriers to having a regular doctor. The third group entered focused on indicators of medical need from substance use vulnerability, and included having heroin or cocaine as a primary drug (reference group alcohol), and an indicator of a history of homelessness. Medical need, the fourth group entered, included measures of ever having an overnight hospitalization, the SF-36 mental component summary scales, and being told by a physician of a chronic medical condition. The contribution of each risk factor is reported as an adjusted odds ratio, with all analyses conducted using SAS (33).

RESULTS

General Characteristics

Of 2,062 patients screened at the detoxification unit, 1,420 did not meet study eligibility criteria; the major reasons were self-report that they already had an established primary care physician (n = 980, 69%), and residence not in proximity to the referral primary care clinic area (n = 204, 14%). Of 642 eligible subjects, 470 (73%) provided informed consent and participated in this study; the remainder left the detoxification unit before being approached for informed consent, could not be scheduled on the clinic day, or refused informed consent. Refusal to participate was unusual. More commonly, a subject found eligible was discharged from the program before informed consent could be administered or before one of the two weekdays of clinic operation.

Characteristics of subjects are shown in Table 1. Subjects were generally young (mean age 35.8 years), 39% unemployed, 75% male, and ethnically diverse, with 13% born outside the US. Characteristics reflect a disadvantaged status, with a high proportion homeless. Alcohol was the first or second drug of choice for 63% of subjects; recent cocaine use was endorsed by 65% of subjects, heroin use by 38%.

Health Status

A high proportion of these detoxification patients reported lifetime chronic illness and recent infectious disease, indicating need for medical attention. Nearly one-half (47%) stated that a physician had previously told them that they had at least one condition from a list of chronic conditions: asthma/emphysema or chronic lung disease, high blood pressure, chronic liver disease, seizures/epilepsy, chronic arthritis/osteoarthritis,

Table 1. Demographic characteristics and health status of detoxification subjects (n = 470)

Characteristic	N	Percent
Male	359	76
Race		
Black	218	46
White	173	37
Hispanic	51	11
Other	28	6
Homeless shelter past 5 years, at least one month	219	47
Unemployed	181	39
Any Insurance past 6 months	187	40
English 1st Language	418	89
Age, mean years (SD)	36 (8)	
SF-36 standardized mental component summary, mean score (SD)[a]	32 (8)	
General health		
Excellent/very good	186	40
Good	143	30
Fair	110	23
Poor	31	7
Chronic medical condition (told by doctor)	222	47
Shot, stabbed or traffic accident, past 6-months	52	11
Drug incident requiring ED attention[b]	68	14
Self-reported drug of choice (primary drug)		
Alcohol	185	39
Cocaine	156	33
Heroin	128	27
Medical problems in past 30 days	287	61
Mean number of days (SD), if any	13.3 (11.0)	

[a]N = 468.
[b]Chest pain after cocaine, overdose; ED = emergency department.

peripheral neuropathy, HIV/AIDS, heart attack (myocardial infarction), other heart disease requiring medication, diabetes, cancer, heart failure (congestive heart failure), and stroke. Table 1 presents other summary information on subjects' health at time of detoxification. Traumatic injury from gunshot, stabbing or traffic accident was also common. In the prior 6 months, 14% of all subjects reported overdose or chest pain after cocaine use requiring an emergency department visit, 12% physical assaults by a known individual, and 24% other serious injuries (34). Other common acute illnesses included pneumonia (27%), vomiting

Table 2. Health services used by detoxification patients without primary care 6 months prior to detoxification (n = 470)

Type of care	N	Percent	Range
Overnight hospitalization	68	14	
Mean admissions (SD), if any	1.5 (1.3)		1–10
Mean nights (SD), if any	8.5 (12.6)		1–69
Prior detoxification	223	47	
Mean prior admissions (SD), if any	2.5 (2.5)		1–21
Mean nights (SD), if any admission	13.6 (12.6)		0–82
Visited ED	223	47	
Mean visits (SD), if any	1.8 (1.6)		1–16

blood (19%), abdominal or stomach pain (16%), and ulcers (10%). Additional discussion of the health status of these detoxification patients without primary medical care has been reported (26).

Medical Care Utilization

The majority of subjects were utilizers of a variety of medical services in the 6 months prior to detoxification. Table 2 presents the type and frequency of services used. Over 14% had been hospitalized overnight in the prior 6 months. Among the hospitalized subjects, the mean number of admissions was 1.5 (SD = 1.3) and mean number of nights of care 8.5 (SD = 12.6).

Nearly one-half of subjects reported at least one visit to a hospital ED (47%); 19% went 2 times or more; range 1–16, with a mean among hospital ED users of 1.8 (SD = 1.6). When asked the name and location of the ED they visited during the prior 6 months, subjects reported locations throughout the metropolitan area, reflecting the complex web of service delivery in the Boston area. Some ED visits were out-of-state, perhaps reflecting an injury, unanticipated medical care event, or unstable living situation.

Perceived Barriers to Medical Care

When read a list of possible reasons perceived as barriers to medical care, the majority of subjects agreed with at least one statement related to difficulties paying for regular medical care and to at least one statement related to help-seeking attitudes or beliefs about medical care (Table 3). Among the group with no insurance, statements about payment barriers received the highest level of endorsement (87%). While a smaller proportion of those with insurance agreed with statements about payment

Table 3. Self-reported barriers to having a regular doctor by insurance status last 6 months (n = 470)

	No insurance (n = 283) Percent (n)	Have insurance (n = 187) Percent (n)
Payment barriers	87 (246)	59 (111)
Could not pay for services	85	57
Not eligible for free care	23	18
System or structural barriers	37 (105)	46 (86)
Can't get to services... transportation problems	23	35
Clinic hours are inconvenient	16	16
Do not want to lose my job	15	10
Help-seeking attitudes or beliefs	60 (168)	59 (111)
Do not know where to go	41	34
Afraid others... find out about a health problem I have	10	13
Do not feel I need a regular doctor	22	20
Other reason	2	6
Social Disadvantage/hardship	5 (14)	8 (15)
Do not speak ...English well enough	2	3
Do not have someone... to watch children	3	5
My substance abuse interferes...	61 (173)	65 (121)

barriers, a majority still reported difficulties in paying for services of a regular doctor (59%). The statement "my substance abuse interfered" was the most frequent barrier endorsed by the group with insurance (65%), and an equally common reason among the uninsured group (61%). A large minority of those without insurance (41%) reported they did not know where to go for a regular doctor and, the majority (60%) of uninsured agreed with at least one help-seeking attitude or belief statement. Overall, help-seeking attitudes and beliefs were equally important among those with and without insurance, although the distribution of specific reasons appeared slightly different. Also very common in both groups was mention of a system or structural barrier to having a regular doctor, most frequently difficulty getting to the office; at least one system or structural barrier was mentioned by a substantial minority of subjects with (46%) and without (37%) insurance. Other evidence of social disadvantage such as language barriers or unmet childcare needs was relatively infrequently noted.

Repeated Visits to an Emergency Department

In the multivariable logistic regression analysis (Table 4), subjects who self-identified as white had greater odds of repeat ED use in the past 6 months relative to non-white subjects (OR = 2.2, CI = 1.25–3.83). The odds ratio for white race was statistically significant in all model specifications, although its magnitude was attenuated when vulnerability and need measures were included. Age and gender, the other predisposing measures, were not significant in any of the models.

Among the enabling variables, a self-perception that substance abuse interfered with seeking care from a regular doctor significantly increased the odds of repeat ED use (OR = 1.9, CI = 1.04–3.64). Again, though it remained statistically significant, the magnitude of the coefficient was reduced when medical need variables were added to the final model. This may indicate that association with need moderated some of the influence of this enabling determinant. Subjects with insurance were more likely to seek repeat ED care (OR = 1.6, CI = 0.94–2.87, p = 0.08 in the final model); however, its impact appeared to be moderated by health needs. Other enabling variables (recent employment status, other perceived barriers to regular doctor) were not associated with repeat ED visits.

Homeless subjects had increased odds of repeat ED visits (OR = 2.1, CI = 1.18–3.68). Subjects who perceived cocaine as the primary drug problem had reduced odds of repeat ED visits (OR = 0.5, CI = 0.22–0.97), even after controlling for other medical needs, relative to subjects with alcohol as a primary problem. Heroin users had reduced odds relative to alcohol users but the coefficient was not statistically significant.

DISCUSSION

Detoxification patients reported a notable burden of lifetime chronic illness, recent infectious diseases, and trauma indicating a vulnerability to illness episodes that might require frequent medical contact to manage symptoms, maintain treatment, or respond to acute problems (26, 34). In this study, hospital and ED utilization were common. While we did not query the specific reasons for ED visits, it is likely that not all ED use was inappropriate. Given the medical needs of this sample, some ED use was undoubtedly warranted. To better understand preventable or over-reliance on the ED, we focused the final investigation on repeat ED use. Despite this approach, it is possible, even likely, that this measure of ED use may include truly needed ED care.

The present research was largely exploratory because only one prior study has reported ED use among substance abusers undergoing

Table 4. Characteristics (Predisposing, enabling, and type of illness) associated with emergency department utilization (2 or more visits in past 6 months) among detoxification patients without primary care (adjusted odds ratios) (n = 465)

	Model			
Characteristics	Demographics	+ Enabling	+ Addiction Need	+ Medical Need
Female	1.5 (0.89–2.59)	1.4 (0.80–2.51)	1.8 (0.98–3.240)	1.6 (0.88–3.05)
Age (years)	1.0 (0.99–1.06)	1.0 (0.99–1.05)	1.0 (0.98–1.04)	1.0 (0.97–1.04)
White	**2.8 (1.75–4.56)**	**2.9 (1.78–4.79)**	**2.2 (1.27–3.72)**	**2.2 (1.25–3.83)**
Payment barrier		0.8 (0.42–1.40)	0.7 (0.37–1.27)	0.7 (0.38–1.37)
System/structural barrier		1.1 (0.65–1.83)	1.0 (0.60– 1.73)	0.9 (0.49–1.50)
My SA interferes (barrier)		**2.5 (1.42–4.47)**	**2.3 (1.25–4.10)**	**1.9 (1.04–3.64)**
Help-seeking belief or attitude barrier		0.9 (0.55–1.50)	0.9 (0.55–1.53)	0.9 (0.52–1.52)
Has insurance		**1.8 (1.09–3.12)**	**1.8 (1.04–3.0)**	1.6 (0.94–2.87)
Unemployed		1.1 (0.66–1.89)	1.0 (0.58–7.71)	1.1 (0.60–1.86)
Heroin the primary drug[a]			0.6 (0.35–1.20)	0.7 (0.38–1.37)
Cocaine the primary drug[a]			**0.4 (0.20–0.84)**	**0.5 (0.22–0.97)**
Homeless[b]			**1.9 (1.08–3.20)**	**2.1 (1.18–3.68)**
Injury past 6 months				**2.7 (1.26–5.60)**
Hospitalized in lifetime				1.0 (1.00–1.08)
SF-36 MCS				1.0 (0.97–1.01)
Any chronic condition				**2.1 (1.17–3.63)**

Bold signifies $p < 0.05$.
SA = substance abuse; MCS = standardized mental component summary.
[a]alcohol as the primary drug is reference group.
[b]at least one month in shelter, past 5 years.

detoxification (43). Comparable to the 47% of the present research cohort reporting emergency department use in the past 6 months (Table 2), 63% of the prior study's drug users (n = 77) had at least one emergency department visit in the prior year. Hospitalization rates were less comparable, with the present research showing 14% of patients being hospitalized in the past 6 months, compared to the prior study's finding of 46% of patients undergoing hospitalization in the past year. That study also found that drug users were 2.3 times as likely to have at least one visit to an emergency department in the past year and 6.7 times as likely to be hospitalized as non-drug users who sought care at a general hospital walk-in clinic. However, the previous study did not distinguish those patients with a primary care relationship from those patients without such care. A second study of addicted patients seeking treatment or detoxification found that 55% of 5,824 patients had recent ED use (past 3 months) and recent ED use was not significantly associated with having a primary care physician. In that study, 41% did not have a primary care physician (44). Another study showed that among the homeless, those with repeat ED use were significantly more likely to be substance users (45).

In logistic regression models, indicators of medical needs (i.e., being told of a chronic medical condition, having an injury in the past 6 months) were associated with repeat ED use, as hypothesized and consistent with other utilization studies (35, 36, 38, 39, 46). Homelessness, an indicator of social vulnerability, also was associated with increased odds of repeat ED use. Unanticipated results included the finding that subjects who were white had increased odds of 2 or more visits to the ED in a 6-month period. This finding is not explained by age, medical need, medical insurance, perceptions of barriers to regular doctors, or substance use. Also somewhat surprisingly, the mental health status measure, the SF-36 mental component summary, was not a significant determinant. This measure of mental health reflected a two-week window prior to the interview, whereas the study's utilization measure reviewed a 6-month period. It is still possible that mental health is associated with ED utilization. If current mental health status was not sufficiently correlated with prior mental health status, this measure of mental health needs may have been inadequate.

We studied the health care utilization of substance abusers for the immediate period before they were admitted for detoxification care. While this is a period of particularly high healthcare needs, previous utilization studies rarely focus on this vulnerable period for a substance abusing population (5). Since substance abusers often seek detoxification during a crisis, the detoxification stay may also be a potentially "teachable" moment. Furthermore, nationally, a high proportion of all admissions for substance abuse care take place in detoxification programs. Some substance abusers rely on detoxification programs as their only source of

addictions care. Again, this demonstrates that detoxification care may provide a limited opportunity to link substance abusers to other services they need.

This study focuses on substance abusers reporting no relationship with a regular doctor, a group of particular interest to policymakers interested in reaching the hardest to engage group. In Boston (the study area), people with chronic or episodic serious medical conditions have various opportunities to obtain free medical care through health clinics and a state hospital-based free care program. Also, a substantial proportion of low-income individuals can enroll in an expanded Medicaid program in Massachusetts and be assigned a primary care doctor. Indeed, a substantial proportion of this study sample reported some health insurance (primarily Medicaid) and thus had some payment method for regular doctors. Despite opportunities to establish a relationship with a primary care physician, utilization of episodic health services was high. Multiple indicators of high medical needs and the notable frequency of ED use and hospitalizations underscore the importance that a primary care relationship may hold. In a Medicaid population it was previously reported that continuity of care with a provider was associated with a decreased future likelihood of hospitalization (37).

Since care-seeking is a complex decision process, it is influenced by the individual's self-perceived distress, own attitudes, encouragement from others, cost of care, and ease of access (40). Our hypothesis that barriers to primary care would increase ED use is informed by service studies. A study of a nationally representative sample of adults reported that dissatisfaction with the usual source of care and perceived barriers to a usual source of care were associated with having an ED visit for non-urgent care (47). A second community-based study of injection drug users, other drug users, and nonusers found that injection and other drug users were less satisfied with access to care than non-drug users (41).

Despite pressing medical and addictions issues, a substantial proportion of the sample in this study reported they "didn't need a regular doctor", or had another fear or notion about medical care that may have precluded seeking care from a regular doctor, consistent with a prior study of drug users who had not sought health care (42). These attitudes were sometimes accompanied by other frustrations: Difficulty seeking care at a convenient time or location. Even among the insured, there was a perception that they could not afford to have a regular doctor. Nevertheless, most self-reported attitudes and barriers to a regular doctor were not associated with repeated use of the ED as a site of care. Analogously, a study of the homeless in Los Angeles found that competing priorities affected use of regular medical care, but were not associated with increased hospitalization rates (32).

There are several limitations associated with the present study. Detoxification patients who had primary care were excluded; the study was not able to compare subjects to a comparison group of detoxification patients who did have established primary care. Rather, the notable high use of ED and other utilization services was compared to findings from substance abusing cohorts reported in the literature. Furthermore, subjects in the current study lived in the catchment area of a hospital that provided care to all patients regardless of health insurance status. Communities without this resource might have different utilization for similar patients.

Utilization data obtained by self-report is a limitation. Self-report data can lead to underreporting because of difficulty in recalling distant medical events or over-reporting because of "telescoping" or reporting an event that actually occurred in a prior period. However, such issues were addressed by the nature of this study's design: a) the recall period is relatively brief, only 6 months; b) ED and hospitalization were chosen as the focus of this study, as these important events are easier to recall than outpatient visits; and, c) the outcome analyzed is a dichotomous measure rather than a less reliable count measure. A related limitation is that self-reported ED and hospitalization utilization were not verified by medical records. However, for other analyses not reported here we conducted a sub-study involving the search of computerized administrative data on primary care visits from two major sites of primary care for this population. We discovered that of subjects with self-report data who were determined by administrative data to have linked to primary care, 81% reported linkage (kappa = 0.41) (21). It is likely that the reliability of reporting ED and hospitalization would be even higher.

The results present one reason to hypothesize that establishment of primary medical care would reduce the episodic pattern of care seen here. Subjects that acknowledged their substance abuse interfered with getting care from a regular doctor were more likely to have repeated ED visits. Prior research with adults with an identified health problem found that chronic drug users were more likely than non-drug users to not desire medical treatment or delay seeking needed services (42). Yet, attitudinal reasons for not getting care are not queried in all access studies and deserve further study by health services researchers (28). This tendency to delay needed care may in part explain higher ED utilization among a subgroup of the detoxification patients. We speculate that both the awkwardness of not being able to maintain abstinence, and the disorganized life associated with continued substance use, hinder a relationship with one doctor and may promote delays in seeking care at the first sign of symptoms. Attending to these issues in detoxification or other substance abuse programs, for example, by facilitating the establishment of primary medical care, may be beneficial in reducing reliance on episodic sources of care.

Understanding and effectively addressing perceptions about the value of care may also address the difference in ED utilization found between White subjects and subjects from other racial/ethnic groups. This type of association is suggested by a study of drug users, where African American, Hispanic, and white drug users reported different attitudes towards the value of getting drug treatment and different perceptions of their need for drug treatment (48).

Evidence also is provided that it is meaningful in health services studies to take into consideration the primary drugs of abuse. In this sample of substance abusers requiring detoxification, cocaine users were less likely than alcohol users to require repeat emergency care, a finding consistent with different types of medical events in the two groups and perhaps suggesting an overall propensity of alcohol abusers to seek episodic care (40, 49).

Generally, these findings suggest that policymaking that addresses payment and structural barriers to doctor office care alone might not be enough to change reliance upon the ED as an important source of care in this population if substance abuse issues continue to interfere. Other approaches, attentive to the special characteristics of substance abusers, that actually link persons in detoxification programs to medical care could be fruitful (36). This study's findings may be very generalizable to the common situation in the U.S. in which alcohol and drug dependent individuals do not have good access to primary medical care. Addressing a broader range of health issues in the one healthcare setting in which these challenging to engage individuals present for care, a residential detoxification unit, provides novel opportunities for behavioral administrators and clinical managers. The substance abuse treatment setting is an opportunity to address the health needs and health expenditures for this high utilizing population with medical and mental health co-morbidities. Interventions that address some of the attitudes that were common in this group might have success at establishing a sustained primary care relationship that ultimately changes utilization patterns.

ACKNOWLEDGEMENTS

This work is supported in part by the General Clinical Research Center grant M01 RR00533 from the National Center for Research Resources and by grants from the National Institute of Alcohol Abuse and Alcoholism (R01 AA 10870) and the National Institute on Drug Abuse (R01 DA 10019). An abstract of an earlier version of this article was presented at the Annual Meeting of the Association for Health Services Research, June 2000. The authors thank Jessica Richardson for assistance with manuscript preparation.

REFERENCES

1. Substance Abuse and Mental Health Services Administration. *Treatment Episode Data Set (TEDS): 1994–1999, National Admissions to Substance Abuse Treatment Services*. Office of Applied Studies. DASIS Series: S-14, DHHS Publication No. (SMA) 01-3550, Rockville, MD, 2001.
2. Friedmann PD, Alexander JA, Jin L, D'Aunno TA. On-site primary care and mental health services in outpatient drug abuse treatment units. J Behav Health Serv Res 1999; 26(1):80–94.
3. Samet JH, Saitz R, Larson MJ. A case for enhanced linkage of substance abusers to primary medical care. Substance Abuse 1996; 17(4):181–192.
4. Samet JH, Friedmann P, Saitz R. Benefits of linking primary medical care and substance abuse services: Patient, provider, and societal perspectives. Arch Intern Med 2001; 161(1):85–91.
5. McGeary KA, French MT. Illicit drug use and emergency room utilization. Health Serv Res 2000; 35(1 Pt 1):153–169.
6. Falik M, Needleman J, Wells BL, Korb J. Ambulatory care sensitive hospitalizations and emergency visits: Experiences of Medicaid patients using federally qualified health centers. Med Care 2001; 39(6):551–561.
7. Lange RA, Hillis LD. Cardiovascular complications of cocaine use. N Engl J Med 2001; 345(5):351–358.
8. Perper JA, Van Thiel DH. Respiratory complications of cocaine abuse. Recent Dev Alcohol 1992; 10:363–377.
9. Patkar AA, Sterling RC, Gottheil E, Weinstein SP. A comparison of medical symptoms reported by cocaine-, opiate-, and alcohol-dependent patients. Substance Abuse 1999; 20(4):227–235.
10. Substance Abuse and Mental Health Services Administration. *Emergency Department Trends from the Drug Abuse Warning Network, Preliminary Estimates January–June 2001 with Revised Estimates 1994 to 2000*. Office of Applied Studies. DAWN Series D-20, DHHS Publication No. (SMA) 02-3634. Rockville, MD, 2002.
11. Nahass RG, Weinstein MP, Bartels J, Gocke DJ. Infective endocarditis in intravenous drug users: A comparison of human immunodeficiency virus type 1-negative and -positive patients. J Infect Dis 1990; 162(4):967–970.
12. Samet JH, Shevitz A, Fowle J, Singer DE. Hospitalization decision in febrile intravenous drug users. Am J Med 1990; 89(1):53–57.
13. French MT, McGeary KA, Chitwood DD, McCoy CB. Chronic illicit drug use, health services utilization and the cost of medical care. Soc Sci Med 2000; 50(12):1703–1713.
14. Curran GM, Sullivan G, Williams K, Han X, Collins K, Keys J, Kotrla KJ. Emergency department use of persons with comorbid psychiatric and substance abuse disorders. Ann Emerg Med 2003; 41(5):659–657.
15. Sullivan G, Wells KB, Morgenstern H, Leake B. Identifying modifiable risk factors for rehospitalization: A case-control study of seriously mentally ill persons in Mississippi. Am J Psychiatr 1995; 152(12):1749–1756.

16. Kent S, Yellowlees P. Psychiatric and social reasons for frequent rehospitalization. Hosp Comm Psychiatr 1994; 45(4):347–350.
17. Booth BM, Blow FC, Cook CA, Bunn JY, Fortney JC. Relationship between inpatient alcoholism treatment and longitudinal changes in health care utilization. J Stud Alcohol 1997; 58(6):625–637.
18. Neumark YD, Van Etten ML, Anthony JC. "Alcohol dependence" and death: Survival analysis of the Baltimore ECA sample from 1981 to 1995. Subst Use Misuse 2000; 35(4):533–549.
19. Neumark YD, Van Etten ML, Anthony JC. "Drug dependence" and death: Survival analysis of the Baltimore ECA sample from 1981 to 1995. Subst Use Misuse 2000; 35(3):313–327.
20. Hwang SW, Orav EJ, O'Connell JJ, Lebow JM, Brennan TA. Causes of death in homeless adults in Boston. Ann Intern Med 1997; 126(8):625–628.
21. Samet JH, Larson MJ, Horton NJ, Doyle K, Winter M, Saitz R. Linking alcohol and drug dependent adults to primary medical care: A randomized controlled trial of a multidisciplinary health evaluation in a detoxification unit. Addiction 2003; 98:509–516.
22. Saitz R, Larson MJ, Horton NJ, Winter M, Samet JH. Linkage with primary medical care in a prospective cohort of adults with addictions in inpatient detoxification: Room for improvement. Health Serv Res 2004; 39(3):587–606.
23. Sweeney LP, Samet JH, Larson MJ, Saitz R. Establishment of a multidisciplinary Health Evaluation and Linkage to Primary care (HELP) clinic in a detoxification unit. J Addict Dis 2004; 23(2):33–45.
24. 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(3):189–198.
25. Ware Jr. JE. SF-36 Health Survey: Manual and Interpretation Guide. Boston: The Health Institute – New England Medical Center, 1993.
26. 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(1):33–45.
27. McLellan AT, Kushner H, Metzger D, et al. The fifth edition of the addiction severity index. J Subst Abuse Treat 1992; 9(3):199–213.
28. Strunk BC, Cunningham PJ. Treading Water: Americans' Access to Needed Medical Care, 1997–2001. Washington, DC: Center for Studying Health System Change, March 2002.
29. Roper Starch Worldwide Inc. America's Mental Health Survey 2001. New York, NY: The National Mental Health Association, May 2001. Roper Number: CNT505.
30. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. Milbank Mem Fund Q Health Soc 1973; 51(1):95–124.
31. Phillips KA, Morrison KR, Andersen R, Aday LA. Understanding the context of healthcare utilization: Assessing environmental and provider-related variables in the behavioral model of utilization. Health Serv Res 1998; 33(3 Pt 1): 571–596.

32. Gelberg L, Gallagher TC, Andersen RM, Koegel P. Competing priorities as a barrier to medical care among homeless adults in Los Angeles. Am J Public Health 1997; 87(2):217–220.
33. SAS/STAT Software, Version 8.2 of the SAS System for Windows [computer program]. Version 8.2. Cary, NC: SAS Institute Inc., 2001.
34. Rees VW, Horton NJ, Hingson RW, Saitz R, Samet JH. Injury among detoxification patients: Alcohol users' greater risk. Alcohol Clin Exp Res 2002; 26(2):212–217.
35. Wenzel SL, Bakhtiar L, Caskey NH, et al. Homeless veterans' utilization of medical, psychiatric, and substance abuse services. Med Care 1995; 33(11):1132–1144.
36. Laine C, Hauck WW, Gourevitch MN, Rothman J, Cohen A, Turner BJ. Regular outpatient medical and drug abuse care and subsequent hospitalization of persons who use illicit drugs. J Am Med Assoc 2001; 285(18):2355–2362.
37. Gill JM, Mainous AG, 3rd. The role of provider continuity in preventing hospitalizations. Arch Fam Med 1998; 7(4):352–357.
38. Padgett D, Struening EL, Andrews H. Factors affecting the use of medical, mental health, alcohol, and drug treatment services by homeless adults. Med Care 1990; 28(9):805–821.
39. Koegel P, Sullivan G, Burnam A, Morton SC, Wenzel, S. Utilization of mental health and substance abuse services among homeless adults in Los Angeles. Med Care 1999; 37(3):306–317.
40. Mechanic D. Removing barriers to care among persons with psychiatric symptoms. Health Aff 2002; 21(3):137–147.
41. Chitwood DD, Comerford M, McCoy HV. Satisfaction with access to health care among injection drug users, other drug users, and nonusers. J Behav Health Serv Res 2002; 29(2):189–197.
42. McCoy CB, Metsch LR, Chitwood DD, Miles C. Drug use and barriers to use of health care services. Subst Use Misuse 2001; 36(6–7):789–806.
43. Stein MD, O'Sullivan PS, Ellis P, Perrin H, Wartenberg A. Utilization of medical services by drug abusers in detoxification. J Subst Abuse 1993; 5(2): 187–193.
44. Saitz R, Mulvey KP, Samet JH. The substance-abusing patient and primary care: Linkage via and addiction treatment system. Substance Abuse 1997; 18(4):187–195.
45. Kushel MB, Perry S, Bangsberg D, Clark R, Moss AR. Emergency department use among the homeless and marginally housed: Results from a community-based study. Am J Public Health 2002; 92(5):778–784.
46. Thornquist L, Biros M, Olander R, Sterner S. Health care utilization of chronic inebriates. Acad Emerg Med 2002; 9(4):300–308.
47. Sarver JH, Cydulka RK, Baker DW. Usual source of care and nonurgent emergency department use. Acad Emerg Med 2002; 9(9):916–923.
48. Longshore D, Hsieh SC, Anglin MD, Annon TA. Ethnic patterns in drug abuse treatment utilization. J Ment Health Adm 1992; 19(3):268–277.
49. Weisner C, Mertens J, Tam T, Moore C. Factors affecting the initiation of substance abuse treatment in managed care. Addiction 2001; 96(5):705–716.

American Journal of Gastroenterology
© 2006 by Am. Coll. of Gastroenterology
Published by Blackwell Publishing

ISSN 0002-9270
doi: 10.1111/j.1572-0241.2006.00616.x

Hepatitis C Infection Is Associated with Depressive Symptoms in HIV-Infected Adults with Alcohol Problems

Howard Libman, M.D.,[1] Richard Saitz, M.D., M.P.H.,[2,3] David Nunes, M.D.,[4] Debbie M. Cheng, Sc.D.,[2,5] Jessica M. Richardson, B.A.,[2] John Vidaver, M.A.,[6] Julie K. Alperen, Dr.PH.,[2] and Jeffrey H. Samet, M.D., M.A., M.P.H.[2,7]

[1]*Divisions of General Medicine and Primary Care and Infectious Diseases, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts;* [2]*Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts;* [3]*Youth Alcohol Prevention Center and the Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts;* [4]*Section of Gastroenterology, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts;* [5]*Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts;* [6]*DM-STAT, Inc., Malden, Massachusetts;* [7]*Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts*

OBJECTIVES: Depression is common in persons with HIV infection and with alcohol problems, and it has important prognostic implications. Neurocognitive dysfunction has been reported with chronic hepatitis C virus (HCV) infection. We hypothesized that HCV infection is associated with more depressive symptoms in HIV-infected persons with a history of alcohol problems.

METHODS: We performed a cross-sectional analysis of baseline data from a prospective cohort study of 391 HIV-infected subjects with a history of alcohol problems, of whom 59% were HCV antibody (Ab) positive and 49% were HCV RNA-positive. We assessed depressive symptoms (Center for Epidemiologic Studies Depression [CES-D]) and past month alcohol consumption. In the primary analysis, we evaluated whether there were more depressive symptoms in HCV Ab-positive and RNA-positive subjects in unadjusted analyses and adjusting for alcohol consumption, gender, age, race, CD4 count, homelessness, drug dependence, and medical comorbidity.

RESULTS: Mean CES-D scores were higher in subjects who were HCV Ab-positive compared with those who were HCV Ab-negative (24.3 *vs* 19.0; $p < 0.001$). In adjusted analyses, the difference in CES-D scores between HCV Ab-positive and Ab-negative subjects persisted (24.0 *vs* 19.0; $p < 0.001$). Unadjusted mean CES-D scores were also significantly higher in HCV RNA-positive subjects compared with those who were RNA-negative, and the difference remained significant (24.6 *vs* 19.3; $p < 0.001$) in adjusted analyses.

CONCLUSIONS: HCV/HIV coinfected persons with a history of alcohol problems have more depressive symptoms than those without HCV, and this association is unexplained by a variety of population characteristics. These data suggest that HCV may have a direct effect on neuropsychiatric function.

(Am J Gastroenterol 2006;101:1804–1810)

INTRODUCTION

Depression in HIV-infected patients is a common but underdiagnosed condition with important prognostic implications (1). Depressive symptoms have been associated with poor medication adherence, more rapid HIV disease progression, and increased mortality (2). Better understanding of the factors contributing to depression and its detrimental effect on the course of HIV infection may be gained by studying the effect of significant comorbidities.

Alcohol use and hepatitis C virus (HCV) infection are also common in HIV-infected patients, particularly those with a history of injection drug use. Alcohol use is clearly associated with depression and may exacerbate it (3, 4). Chronic HCV infection has been associated with neurocognitive symptoms, perhaps mediated by a direct effect on the central nervous system (5). Interferon, a component of the treatment regimen for HCV infection, can worsen depressive symptoms (6).

In order to better understand the relation between HCV infection and depressive symptoms in the context of HIV

disease, we studied a cohort of HIV-infected patients with a history of alcohol problems. We tested the hypothesis that HCV infection is associated with more depressive symptoms in these HIV-infected subjects.

METHODS

Subject Recruitment

Study subjects were participants in the HIV-LIVE (HIV-Longitudinal Interrelationships of Viruses and Ethanol) study, a prospective, observational cohort study of HIV-infected patients with past or current alcohol problems. The present study is a cross-sectional analysis of data collected at entry into the HIV-LIVE cohort.

A total of 401 subjects were recruited from several different sources including: (1) a previous cohort study of people with HIV and alcohol problems (N = 154, 38%) (7); (2) the Diagnostic Evaluation Unit (DEU), an intake clinic for HIV-infected patients at Boston Medical Center (BMC) (N = 88, 22%) (8); (3) the HIV Primary Care and Specialty Clinics at Beth Israel Deaconess Medical Center (BIDMC) (N = 31, 8%); and (4) additional health care centers, homeless shelters, drug treatment programs, other studies, subject referrals, and flyers (N = 128, 32%). Enrollment began in August 2001, and ended in July 2003.

Eligibility criteria for the study included the following:

1. Documented HIV Ab test by ELISA and confirmed by Western blot (medical record or tested at enrollment).
2. Two or more affirmative responses to the CAGE alcohol screening questionnaire (9, 10) or physician-investigator diagnosis of alcoholism.
3. Ability to speak English or Spanish.
4. At least one contact person who was likely to know the subject's whereabouts.

Exclusion criteria included: (1) scoring <21 on the 30-item Folstein Mini-Mental State Examination (MMSE) (11); and (2) a trained interviewer assessment that the patient was incapable of comprehending informed consent or of answering the interview questions.

If an eligible individual agreed to participate in this study, a research associate scheduled an appointment for the first interview at BMC's General Clinical Research Center (GCRC) or BIDMC's Clinical Research Center (CRC). All subjects who met the eligibility criteria and wished to participate in the study provided written informed consent prior to enrollment. The Institutional Review Boards of BMC and BIDMC approved this study. Additional privacy protection was secured by the issuance of a Certificate of Confidentiality by the Department of Health and Human Services to protect subjects from release of their research data even under a court order or subpoena.

Subject Assessment

After enrollment, subjects received an interviewer-administered assessment. The assessment included questions on the following: demographics; depressive symptoms (Center for Epidemiologic Studies Depression [CES-D] scale) (12); medical comorbidity by a validated interview measure (13); current and lifetime alcohol use and dependence (Composite International Diagnostic Interview [CIDI]) (14); current drug dependence (CIDI Short Form); and HIV risk behaviors (Risk Assessment Battery [RAB], modified version) (15). Past month alcohol consumption was assessed using a validated calendar method (16). Heavy alcohol consumption was defined as more than 14 drinks per week or more than 4 drinks on any one occasion for men aged 65 yr and younger; or more than 7 drinks per week or more than 3 drinks on any one occasion for women and anyone over the age of 65 yr. Moderate use was defined as 1 or more drinks in the past 30 days but less than the "heavy" category. Abstinent was defined as no drinks in the past 30 days. Homelessness 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.

All subjects in this cohort were Ab tested for HCV infection. Those who were Ab-positive had HCV testing by RNA measurement using polymerase chain reaction testing to verify the presence of active infection.

Primary Outcome

The primary study outcome was depressive symptoms, which were assessed using the CES-D (12). The CES-D is a short self-report tool intended to assess depressive symptoms in the general population. It consists of 20 questions concerning mood and behavior over the past week with results reported as rarely or none of the time (<1 day), some or a little of the time (1–2 days), occasionally or a moderate amount of the time (3–4 days), or most or all of the time (5–7 days). CES-D scores can range from 0 to 60. Higher CES-D scores reflect the presence of more depressive symptoms.

Primary Independent Variable

The main independent variable was HCV status, which was defined in two ways: (1) HCV Ab-positive *versus* Ab-negative and (2) HCV RNA-positive *versus* RNA-negative. Examination of the independent variable in this manner was deemed important to identify a potential biologic effect of HCV infection on depressive symptoms. For the purpose of the analysis, HCV Ab-negative subjects were assumed to be HCV RNA-negative (17).

Statistical Analyses

χ^2 and Wilcoxon rank sum tests were used to compare subject characteristics by HCV serologic status. Multiple linear regression models were used to assess the cross-sectional association between HCV infection and depressive symptoms. Separate analyses were performed for each method of defining HCV status. Covariates examined included alcohol consumption (abstinent *vs* moderate *vs* heavy) (18), gender, age, race (black *vs* white *vs* Hispanic *vs* other), CD4 cell count, homelessness (yes *vs* no), diagnosis of drug dependence (yes *vs* no), and medical comorbidity. Self-reported information

was available on whether subjects ever used injection drugs. However, this variable was highly correlated with HCV infection status, whereas drug dependence diagnosis was not. Thus, drug dependence diagnosis was included as the covariate in regression analyses to avoid potential collinearity.

Secondary analyses were conducted modeling CES-D as a binary outcome (CES-D $\geq$ 23 *vs* CES-D $<$ 23) and also modeling CES-D as a continuous outcome excluding those questions (1, 5, 7, 11, 20) that reflect somatic symptoms. Additional analyses were conducted to assess the following potential confounders: MMSE; ever received interferon therapy; educational level (high school *vs* not); employment status (yes *vs* no); income level (above *vs* below median); and current injection drug use (within 6 months). To assess the potential bias from including subjects who were previously on interferon therapy, the primary analysis was repeated excluding those subjects.

All analyses were conducted using two-sided significance tests defining $p < 0.05$ as statistically significant. Analyses were performed using SAS software (version 8.2; SAS Institute, Cary, NC).

RESULTS

Of the 401 HIV-infected subjects with current or past alcohol problems enrolled in the HIV-LIVE cohort, 391 had available HCV Ab test results. Of these 391 subjects, 231 (59%) were HCV Ab-positive (Fig. 1). Of the 213 HCV Ab-positive subjects who were tested for RNA, 183 (86%) had a detectable level. One additional subject did not have available HCV Ab results, but tested HCV RNA-negative and was included only in the HCV RNA analyses. Only one HCV-infected subject in the study was receiving interferon therapy at baseline, and only 17 of 231 (7.4%) HCV Ab-positive subjects had received interferon therapy ever. Of those 17 subjects, 10 had received it for more than 3 months, and 6 had received it for more than 6 months.

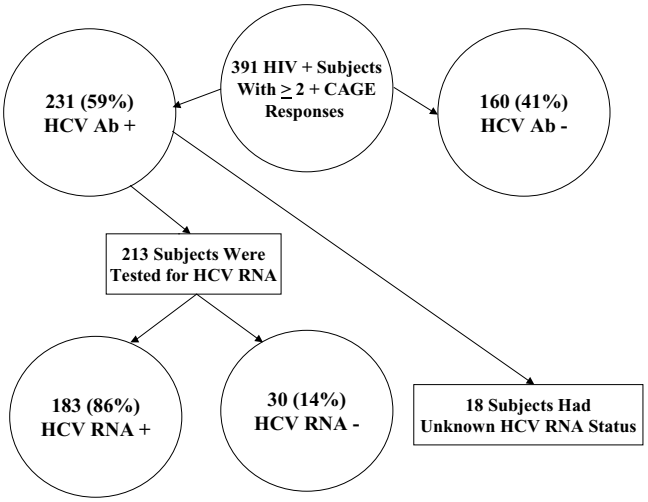


Figure 1. Hepatitis C serologic status of a cohort of HIV-infected subjects with current or past alcohol problems.

Table 1. Characteristics of HIV-Infected Subjects with Current or Past Alcohol Problems

Characteristic	HCV Ab-Positive (N = 231)	HCV Ab-Negative (N = 160)
Male, N (%)*	165 (71%)	130 (81%)
Median (IQR) age*	44.4 (39.7, 48.1)	39.9 (35.6, 45.7)
Race, N (%)		
Black	87 (38%)	75 (47%)
White	77 (33%)	52 (33%)
Hispanic	53 (23%)	21 (13%)
Other	14 (6%)	12 (7%)
Homelessness, N (%)*	69 (30%)	29 (18%)
Drug dependence, N (%)	101 (44%)	65 (41%)
Alcohol consumption, N (%)*		
Abstinent	145 (63%)	78 (49%)
Moderate	18 (8%)	23 (14%)
Heavy	68 (29%)	58 (37%)
Primary HIV risk behavior, N (%)*		
Injection drug use	156 (75%)	12 (8%)
Men sex with men	9 (4%)	75 (50%)
Other	44 (21%)	64 (42%)
Median (IQR) CD4 cell count*	362 (232, 546)	472 (291, 698)
Median (IQR) HIV log RNA	3.0 (0, 4.1)	2.9 (0, 4.1)
Median (IQR) medical comorbidity*	2 (0, 6)	1 (0, 6)
Receiving interferon therapy, N (%)	1 (0.4%)	0 (0.0%)
Mean (SD) MMSE score*	26.94 (2.34)	27.64 (2.14)

*$p < 0.05$.
IQR = interquartile range; SD = standard deviation.

Characteristics of the cohort reflected the urban setting of this study: 75% were men with a median age of 42 yr; 41% were black, 33% white, and 19% Hispanic; 25% were homeless; and 43% met criteria for current drug dependence (past 12 months). Thirty-two percent reported heavy alcohol consumption, 11% had moderate alcohol consumption, and 58% were abstinent in the past 30 days. The median CD4 cell count was 402/mm^3 (interquartile range 241–624/mm^3), and the median HIV log RNA was 2.9 copies/mL (interquartile range 0.0–4.1 copies/mL). The median number of medical comorbidities was 1 (interquartile range 0–6).

Characteristics of subjects who were HCV Ab-positive *versus* those who were HCV Ab-negative are listed in Table 1. HCV Ab-positive subjects were more likely to be men, older, homeless, abstinent from alcohol, have injection drug use as their primary HIV-risk behavior, and have a lower MMSE score. HCV Ab-positive subjects also had a lower median CD4 cell count and more medical comorbidity.

Unadjusted mean CES-D scores were higher in the 231 subjects who were HCV Ab-positive compared with the 160 who were Ab-negative (24.3 *vs* 19.0; $p < 0.001$) (Table 2). In adjusted analyses, the difference in CES-D scores between HCV Ab-positive and Ab-negative subjects remained significant (24.0 *vs* 19.0; $p < 0.001$). Unadjusted mean CES-D scores were significantly higher in the 183 HCV

Table 2. Bivariate and Multivariable Analysis of the Impact of HCV Antibody Status on Depressive Symptoms

	Mean Depressive Symptoms (SE) Unadjusted	Mean Depressive Symptoms (SE) Adjusted*
HCV Ab-positive (N = 231)	24.3 (0.88)	24.0 (0.86)
HCV Ab-negative (N = 160)	19.0 (0.90)	19.0 (1.04)
P-value	<0.001	<0.001

Bivariate and multivariable analysis of the impact of HCV antibody status on depressive symptoms is measured by CES-D score.
*Adjusted for alcohol consumption, gender, age, race, CD4 count, homelessness, drug dependence, and medical comorbidity.

RNA-positive subjects compared with the 191 who were RNA-negative (24.8 *vs* 19.2; $p < 0.001$) (Table 3). The difference in CES-D scores remained significant (24.6 *vs* 19.3; $p < 0.001$) in adjusted analyses.

In order to assess the effect of using a clinically relevant CES-D threshold, we repeated our primary analysis with CES-D < 23 *vs* CES-D ≥ 23 as the dependent variable and still found a significant association between HCV status and depressive symptoms (HCV RNA, adjusted analysis: OR 2.78 [1.73, 4.49]). In order to determine whether somatically focused CES-D questions may have influenced the results, we repeated our primary analysis excluding those five questions and still found a comparable significant association between HCV status and depressive symptoms (Table 4).

In order to assess whether MMSE was a confounder, we repeated our primary analysis adjusting for MMSE score and still found a significant association between HCV status and depressive symptoms (Table 4). In order to determine whether prior interferon therapy may have affected the results, we repeated our primary analysis excluding subjects who had ever received interferon therapy and still found a significant association between HCV status and depressive symptoms (Table 4).

We also examined whether sociodemographic factors may have influenced the results. Our primary analysis was repeated adjusting for educational level, employment status, and income level and still showed a significant difference in CES-D scores by HCV status (Table 4). In addition, we repeated our primary analysis adjusting for self-reported current injection drug use and still found a significant difference in CES-D scores by HCV status (Table 4).

Table 3. Bivariate and Mutivariable Analysis of the Impact of HCV RNA Status on Depressive Symptoms

	Mean Depressive Symptoms (SE) Unadjusted	Mean Depressive Symptoms (SE) Adjusted*
HCV RNA-positive (N = 183)	24.8 (1.00)	24.6 (0.95)
HCV RNA-negative (N = 191)	19.2 (.85)	19.3 (0.92)
P-value	<0.001	<0.001

Bivariate and mutivariable analysis of the impact of HCV RNA status on depressive symptoms is measured by CES-D score.
*Adjusted for alcohol consumption, gender, age, race, CD4 count, homelessness, drug dependence, and medical comorbidity.

Table 4. Adjusted Mean Difference in CES-D Score in Secondary Analyses

Secondary Analysis	HCV Ab+ *vs* Ab−	HCV RNA + *vs* RNA−
CES-D without somatic questions as the dependent variable	3.4	3.8
Primary model adjusting for MMSE	4.7	4.9
Primary model excluding interferon users	5.2	5.2
Primary model adjusting for educational level	4.0	4.5
Primary model adjusting for employment status	4.3	4.6
Primary model adjusting for income level	4.2	4.6
Primary model adjusting for current drug use	5.1	5.3

Primary model examines CES-D as continuous variable and is adjusted for alcohol consumption, gender, age, race, CD4 count, homelessness, drug dependence, and medical comorbidity.
All adjusted mean differences $p < 0.05$.

DISCUSSION

In this cohort of HIV-infected subjects with current or past alcohol problems, depressive symptoms were significantly more frequent in those coinfected with HCV. Other population characteristics, including alcohol consumption, gender, age, race, CD4 count, homelessness, drug dependence, and medical comorbidity, did not account for this observed difference. Significant differences in depressive symptoms between HCV-infected and uninfected subjects were still noted when using a clinically relevant CES-D threshold, excluding CES-D questions with somatic content, adjusting for MMSE score, excluding subjects who had ever received interferon therapy, and adjusting for additional sociodemographic factors and current injection drug use.

Depression is common in patients with chronic HCV infection, and most authors have attributed it to a psychological response to a chronic progressive medical condition or drug use itself (19–26). In one blinded study of 309 injection drug users, 57.2% of subjects with HCV infection had significant depressive symptomatology based on CES-D test results compared with 48.2% of HCV-negative controls (27). None of the HCV-infected subjects were receiving interferon therapy. However, another study comparing 295 injection drug users who were HIV+/HCV− (N = 81), HIV−/HCV+ (N = 62), and HIV−/HCV− (N = 152) found no differences in psychological morbidity on several affective scales among these groups (28).

Several studies have described an association between chronic HCV infection and neurocognitive dysfunction that

appears independent of liver disease severity. Forton *et al*. used a computer-based cognitive battery to demonstrate selective impairments of attention, concentration, and psychomotor speed in patients without significant disease on liver biopsy (5). Fatigue, depression, or a history of drug abuse did not account for these findings. Hilsabeck *et al*. described neuropsychological impairment in 49% of HCV-infected patients without cirrhosis (29). McAndrews *et al*. evaluated a cohort of HCV-infected subjects, screened to exclude relevant comorbidities, with neuropsychological tests (30). Compared to controls, subjects with HCV infection were observed to have somewhat poorer learning ability. Ryan *et al*. compared coinfected subjects with advanced HIV disease to subjects without HCV infection using neurocognitive testing and psychiatric interviews (31). Forty-two percent of each group met criteria for major depression, but coinfected subjects exhibited diminished neurocognitive capabilities.

Neuroradiologic and neurophysiologic studies have indicated the possibility of an underlying biological mechanism for neurocognitive dysfunction in HCV-infected patients (32, 33). Forton *et al*. showed altered brain metabolism using proton magnetic resonance spectroscopy in patients with chronic HCV infection (32). Kramer *et al*. demonstrated mild abnormalities on neuroelectrophysiologic testing in this patient population not attributable to drug or alcohol use or cirrhosis (33). HCV genomic sequences have also been detected in postmortem brain tissues along with evidence of viral replication (34, 35). Thus, in addition to epidemiological evidence for an association between HCV infection and depressive symptoms, neuropsychological testing and physiological data from neuroimaging provide support for a potential biologic basis for this observation.

Depression affects half of the HIV-infected population at some time in the course of their disease, occurring twice as frequently as in seronegative persons (1, 36–39). However, because of the clinical focus on other complications, it may not always be diagnosed (39, 40). Depression in this patient population has been associated with decreased adherence to medical therapy (41) and increased mortality (2, 42). Recognition that HIV-infected patients who also have hepatitis C may be prone to more depressive symptoms has important management implications.

This study has several limitations. While the CES-D is a well-validated scale for depressive symptoms, use of other instruments, such as the Beck Depression Inventory, might have yielded different results. Interpretations of the importance of differences in CES-D scores vary. However, the observed differences in this study have generally been considered clinically important (2). Whether these study findings are applicable to other populations with chronic HCV infection would need to be confirmed. The cross-sectional, observational nature of this study limits our ability to establish a causal link between HCV infection and depressive symptoms. In addition, we cannot distinguish between the effects of injection drug use and HCV serostatus on depressive symptoms. An alternative, but less compelling, explanation for these findings would be that depressed persons are more likely to inject drugs, which leads to HCV infection.

In summary, HCV infection appears to be associated with more depressive symptoms in patients with HIV infection who have a history of alcohol problems. Recent literature suggests that HCV infection has a direct effect on the central nervous system, which may be responsible for this observation.

Clinicians should be alert for depressive symptoms in HIV/HCV coinfected patients and initiate treatment for depression when appropriate. Institution of antidepressant therapy may enhance medical adherence, which is key to successful antiretroviral management, and the patient's ability to tolerate treatment for HCV infection. Researchers should focus future efforts on understanding the potential biologic reasons for the association observed in this study. Further research may better delineate the contributions of HCV and other factors in the development of depressive symptoms in HIV-infected patients.

ACKNOWLEDGMENTS

The authors appreciate the contributions of the staff researchers on the project (Rachel Levison, Elizabeth Curley, Eileen O'Connor) and data management assistance (Vincent Faber). Support for this study came from the following grants from the National Institute of 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 MO1 RR00533, and the Clinical Research Center at Beth Israel Deaconess Medical Center, USPHS Grant MO1 RR01032.

STUDY HIGHLIGHTS

What is Current Knowledge

- Depression is common in HIV-infected persons and persons with alcohol problems and has important prognostic implications.
- Neurocognitive dysfunction has been reported with chronic hepatitis C virus (HCV) infection.

What is New Here

- HCV/HIV co-infected persons with a history of alcohol problems have more depressive symptoms than those without HCV.
- This association is unexplained by a variety of population characteristics.
- HCV may have a direct effect on neuropsychiatric function.

Reprint requests and correspondence: Howard Libman, M.D., Beth Israel Deaconess Medical Center, 330 Brookline Avenue SCC 621A, Boston, MA 02215.

Received October 14, 2005; accepted February 10, 2006.

REFERENCES

1. Fairfield KM, Libman H, David RB, et al. Detecting depression: Providing high quality primary care for HIV-infected patients. Am J Med Qual 2001;16:71–4.
2. Ickovics JR, Hamburger ME, Vlahov D, et al. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: Longitudinal analysis for the HIV Epidemiology Research Study. JAMA 2001;285:1466–74.
3. Sullivan LE, Fiellin DA, O'Connor PG. The prevalence and impact of alcohol problems in major depression: A systematic review. Am J Med 2005;118:330–41.
4. Saitz R. Unhealthy alcohol use. N Engl J Med 2005;352:596–607.
5. Forton DM, Taylor-Robinson DS, Thomas HC. Cerebral dysfunction in chronic hepatitis C infection. J Viral Hepat 2003;10:81–6.
6. Scalori A, Pozzi M, Bellia V, et al. Interferon-induced depression: Prevalence and management. Dig Liver Dis 2005;37:102–7.
7. Samet JH, Horton NJ, Traphagen ET, et al. Alcohol consumption and HIV disease progression: Are they related? Alcohol Clin Exp Res 2003;27:862–7.
8. Samet JH, Libman H, LaBelle C, et al. A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. Arch Intern Med 1995;155:1629–33.
9. Mayfield D, McLeod G, Hall P. The CAGE questionnaire: Validation of a new alcoholism screening instrument. Am J Psychiatry 1974;131:1121–3.
10. Buchsbaum DG, Buchanan RG, Centor RM, et al. Screening for alcohol abuse using CAGE scores and likelihood ratios. Ann Intern Med 1991;115:774–7.
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;12:189–98.
12. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1977;1:385–401.
13. 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.
14. Robins LN, Wing J, Wittchen HU, et al. The composite international diagnostic interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 1988;45:1069–77.
15. Navaline HA, Snider EC, Petro CJ, et al. An automated version of the Risk Assessment Battery (RAB): Enhancing the assessment of risk behaviors. AIDS Res and Hum Retroviruses 1994;10:S281–3.
16. Sobell LC, Sobell MB. Alcohol timeline followback (TLFB). In: American Psychiatric Association, ed. Handbook of psychiatric measures. Washington, DC: American Psychiatric Association, 1996:477–9.
17. Thio CL, Nolt KR, Astemborski J, et al. Screening for hepatitis C virus in human immunodeficiency virus-infected individuals. J Clin Microbiol 2000;38:575–7.
18. Helping patients who drink too much: A clinician's guide. www.niaaa.nih.gov/publications/practitioner/guide.pdf.
19. Lee DH, Jamal H, Regenstein FG, et al. Morbidity of chronic hepatitis C as seen in a tertiary care medical center. Dig Dis Sci 1997;42:186–91.
20. Dwight MM, Kowdly KV, Russo JE, et al. Depression, fatigue, and functional disability in patients with chronic hepatitis C. J Psychosom Res 2000;49:311–7.
21. Fontana RJ, Hussain KB, Schwartz SM, et al. Emotional distress in chronic hepatitis C patients not receiving antiviral therapy. J Hepatol 2002;36:401–7.
22. Goulding C, O'Connell P, Murray FE. Prevalence of fibromyalgia, anxiety and depression in chronic hepatitis C virus infection: Relationship to RT-PCT status and mode of acquisition. Eur J Gastroenterol Hepatol 2001;13:507–11.
23. Kraus MR, Schafer A, Csef H, et al. Emotional state, coping styles, and somatic variables in patients with chronic hepatitis C. Psychosomatics 2000;41:377–84.
24. Forton DM, Thomas HC, Murphy CA, et al. Hepatitis C and cognitive impairment in a cohort of patients with mild liver disease. Hepatology 2002;35:433–9.
25. Lehman CL, Cheung RC. Depression, anxiety, posttraumatic stress, and alcohol-related problems among veterans with chronic hepatitis C. Am J Gastroenterol 2002;97:2640–6.
26. McDonald J, Jayasuriya J, Bindley P, et al. Fatigue and psychological disorders in chronic hepatitis C. J Gastroenterol Hepatol 2002;17:171–6.
27. Johnson ME, Fisher DG, Fenaughty A, et al. Hepatitis C virus and depression in drug users. Am J Gastroenterol 1998;93:785–9.
28. Grassi L, Mondardini D, Pavanati M, et al. Suicide probability and psychological morbidity secondary to HIV infection: A control study of HIV-seropositive, hepatitis C virus (HCV)-seropositive and HIV(HCV-seronegative injecting drug users. J Affect Disord 2001;64:195–202.
29. Hilsabeck RC, Perry W, Hassanein TI. Neuropsychological impairment in patients with chronic hepatitis C. Hepatology 2002;35:440–6.
30. McAndrews MP, Farcnik K, Carlen P, et al. Prevalence and significance of neurocognitive dysfunction in hepatitis C in the absence of correlated risk factors. Hepatology 2005;41:801–8.
31. Ryan EL, Morgello S, Isaacs K, et al. Neuropsychiatric impact of hepatitis C on advanced HIV. Neurology 2004;62:957–62.
32. Forton DM, Allsop JM, Main J, et al. Evidence for a cerebral effect of the hepatitis C virus. Lancet 2001;358:38–9.
33. Kramer L, Bauer E, Funk G, et al. Subclinical impairment of brain function in chronic hepatitis C infection. J Hepatol 2002;37:349–54.
34. Forton DM, Karayiannis P, Mahmud N, et al. Identification of unique hepatitis C virus quasispecies in the central nervous system and comparative analysis of internal translational efficiency of brain, liver, and serum variants. J Virol 2004;78:5170–83.
35. Radkowski M, Wilkinson J, Nowicki M, et al. Search for hepatitis C virus negative-strand RNA sequences and analysis of viral sequences in the central nervous system: Evidence of replication. J Virol 2002;76:600–8.
36. Bing EG, Burnam MA, Longshore D, et al. Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. Arch Gen Psychiatry 2001;58:721–8.
37. Ciesla JA, Roberts JE. Meta-analysis of the relationship between HIV infection and risk for depressive disorders. Am J Psychiatry 2001;158:725–30.
38. Kilbourne AM, Justice AC, Rabeneck L, et al. General

medical and psychiatric comorbidity among HIV-infected veterans in the post-HAART era. J Clin Epidemiol 2001;54(suppl 1):22–8.
39. Penzak SR, Reddy YS, Grimsley SR. Depression in patients with HIV infection. Am J Health Syst Pharm 2000;57:376–86.
40. Asch SM, Kilbourne AM, Gifford AL, et al. Underdiagnosis of depression in HIV: Who are we missing? J Gen Intern Med 2003;18:450–60.
41. Starace F, Ammassari A, Trotta MP, et al. Depression is a risk factor for suboptimal adherence to highly active antiretroviral therapy. J Acquir Immune Defic Syndr 2002;31(suppl 3):S136–9.
42. Leserman J, Jackson ED, Petitto HM, et al. Progression to AIDS: The effects of stress, depressive symptoms, and social support. Psychosom Med 1999;61:397–406.

CONFLICT OF INTEREST

Guarantor of the article: Howard Libman, M.D.
The authors declared no conflicts of interest.

In the Minority: Black Physicians in Residency and Their Experiences

Jane M. Liebschutz, MD, MPH; Godwin O. Darko, MD, MPH; Erin P. Finley; Jeanne M. Cawse; Monica Bharel, MD; and Jay D. Orlander, MD, MPH

Washington, District of Columbia; Atlanta, Georgia; and Worcester, Boston and Jamaica Plain, Massachusetts

Financial support: Dr. Liebschutz was supported in part by the Robert Wood Johnson Foundation Generalist Faculty Scholar Award Program (RWJ #045452).

This work was presented at the Society of General Internal Medicine Annual Meeting, Boston, MA, May 2000.

Objective: To describe black residents' perceptions of the impact of race on medical training.

Materials and Methods: Open-ended interviews were conducted of black physicians in postgraduate year ≥2 who had graduated from U.S. medical schools and were enrolled in residency programs at one medical school. Using Grounded Theory tenets of qualitative research, data was culled for common themes through repeated readings; later, participants commented on themes from earlier interviews.

Results: Of 19 participants 10 were male, distributed evenly among medical and surgical fields. Four major themes emerged from the narratives: discrimination, differing expectations, social isolation and consequences. Participants' sense of being a highly visible minority permeated each theme. Overt discrimination was rare. Participants perceived blacks to be punished more harshly for the same transgression and expected to perform at lower levels than white counterparts. Participants' suspicion of racism as a motivation for individual and institutional behaviors was tempered by self-doubt. Social isolation from participants' white colleagues contrasted with connections experienced with black physicians, support staff and patients, and participants strongly desired black mentors. Consequences of these experiences varied greatly.

Conclusions: Black physicians face complex social and emotional challenges during postgraduate training. Creating supportive networks and raising awareness of these issues may improve training experiences for black physicians.

Key words: racism ■ grounded theory ■ residency

 From the Section of General Internal Medicine (Liebschutz, Darko, Finley, Cawse, Orlander), Boston Medical Center (Liebschutz, Darko, Finley, Cawse, Bharel, Orlander), Boston University Schools of Medicine (Liebschutz, Darko, Finley, Cawse, Bharel, Orlander) and Public Health (Liebschutz); Health Care for Homeless (Bharel); and VA Boston HealthCare System, Boston, MA (Orlander); General Internal Medicine, Washington Hospital Center, Washington, DC (Darko); Department of Anthropology, Emory University, Atlanta, GA (Finley, doctoral student); University of Massachusetts School of Medicine, Worcester, MA (Cawse, medical student). Send correspondence and reprint requests for *J Natl Med Assoc.* 2006;98:1441–1448 to: Dr. Jane Liebschutz; Section of General Internal Medicine, Boston Medical Center, Boston University Schools of Medicine and Public Health, Boston, MA 02118; phone: (617) 414-7399; fax: (617) 414-4676; e-mail: jliebs@bu.edu

INTRODUCTION

Black students enter medical training at half the expected rate compared to their representation in the U.S. population, and have higher attrition rates from medical schools.[1-5] The available literature identifies factors that predict success and describes programs supporting enrollment and retention of underrepresented minorities in medical school and other graduate programs.[3,6,7] Higher test scores, undergraduate locale and performance, educational level of parents and enthusiasm of the admissions committee are predictive of medical school success.[7,8] Complex environmental and psychological variables, such as integration into the academic environment, predict success in black graduate students.[9]

In medical school and residency, blacks have the universal stressors of the need to improve knowledge and skills, demanding hours, and limited social and personal pursuits. Financial pressures can be more intense for blacks who disproportionately come from less wealthy backgrounds than their white classmates.[10,11]

Medical students and residents of all races perceive a high rate of mistreatment, such as harassment, verbal abuse and personal denigration, during training.[12-15] Racial minorities, particularly blacks, report the highest rates of mistreatment.[12,14,16-19] Blacks perceive more stress in medical school than whites, stemming from their minority status and racial discrimination experi-

enced in training.[8,20] In a survey of physicians practicing in Massachusetts, more than half of the nonwhite respondents reported experiencing some form of discrimination on the job. For this group, racism tended to manifest itself in the disrespect of coworkers and perceived difficulty in advancing professionally.[21] A recent literature review did not reveal any studies focused specifically on black residents.

Existing literature does not clearly describe the impact of racially motivated discrimination or abusive treatment on individuals. Most studies present data on the prevalence of individual discriminatory acts without detailing the range or consequences of such experiences. This lack of detail impedes our understanding of the impact of racial discrimination on the education, professional choices and personal lives of medical trainees.

Our study employed qualitative research methods to understand how black race affects the experience of medical training. While racial discrimination is an important focus, we also sought to understand how race impacts medical training more broadly. In our study, we defined "racial discrimination" as differential treatment based on race rather than individual merit. This definition is supported by the published work from the Committee on National Statistics, National Research Council, an effort to bring a national consensus for the definition of racial discrimination in social science research.[22]

METHODS

Study Design

Qualitative methodology allows the researcher to formulate hypotheses from narrative data that explores the attitudes, feelings, beliefs and behaviors of the target group. Authors analyzed in-depth interviews to gain insight into the experiences of black medical residents and utilized a standard qualitative method—Grounded Theory—to organize the analysis and derive conceptual meaning.[23] As explained below, Grounded Theory requires development of themes and ideas during data collection that can be substantiated or refuted by subsequent participants.

Study Subjects

Participants were recruited from all postgraduate training programs affiliated with a medical school in the northeast United States, and identified by the minority recruitment office and word of mouth. The trainees rotated at a variety of healthcare institutions, including tertiary care hospitals, community hospitals, veterans administration hospitals and numerous outpatient settings (community health centers, private physician offices, hospital affiliated clinics). Eligibility requirements included: self-identification as black or African-American, graduation from a U.S. medical school and completion of ≥2 postgraduate year (PGY). Twenty eligible residents were identified, and 19 agreed to participate. Interviewers told participants this was a study on medical training experiences of black doctors in residency. Each participant was interviewed one time. Authors obtained written informed consent and provided no compensation. The institutional review board approved the study.

Table 1. Characteristics of subjects

	Women (n=9)	Men (n=10)	Total (n=19)
Medical specialty	5	6	11
Surgical specialty	4	4	8
Postgraduate year 2 or 3	7	8	15
Postgraduate year 4+	2	2	4

Interview Technique

Investigators conducted semistructured, in-depth interviews between January 1998 and February 2000. An Asian-Indian female chief resident conducted the first three interviews; a black male physician who completed residency training at the same institution conducted the remaining interviews. The interviews started with an open-ended question on experiences related to racial background. This initial question was later changed to one emphasizing positive experiences to encourage participants to think broadly about the impact of race on their medical training experience and not limit themselves to incidents of stereotypical discrimination. Interviewers probed responses to encourage narrative descriptions of particular incidents (Figure 1). As qualitative research is iterative, interviewers asked later participants about concepts raised by earlier participants to clarify, confirm or refute theories developed by the investigators. Interviews lasting 60–90 minutes were audiotaped in private locations convenient to participants.

Analysis

A professional transcriptionist transcribed each interview. The interviewer listened to each tape while reviewing the transcript to ensure accuracy. To maintain confidentiality, the transcripts included altered versions of the names and locations mentioned on the audiotapes. The authors then read each transcript multiple times to clarify meaning and identify common narrative themes. According to the tenets of Grounded Theory, the interviewer then clarified and expanded these themes in subsequent interviews. This cycle of theme identification and clarification continued until there was no further new information revealed in the interviews.

Four of the authors (JL, GD, JC, EF) independently reviewed transcripts to identify prominent themes and concepts. They developed a preliminary coding scheme

while discussing each of the first 10 interviews. The authors then coded the remaining interviews using this scheme. For each interview, the authors compared coding and resolved differences in interpretation through close reading of the text. The authors adopted a final coding scheme once all the interviews were analyzed. Investigators then entered the interviews into NUD*ST, a software program for qualitative research, where they were electronically coded. Two people coded each interview and compared it for consistency. Investigators resolved differences through discussion of the text. Finally, the authors grouped the coding categories by larger themes to develop a coherent analysis of the reported experiences.

RESULTS

Ten participants were male and nine female. We categorize 11 as training in medical (family medicine, internal medicine, pediatrics) and eight in surgical specialties (emergency medicine, general surgery, obstetrics/gynecology and surgical subspecialties). Fifteen were in PGY2 or 3, the rest in PGY ≥4 (Table 1). To preserve participant anonymity, we do not report individual specialty by gender or year of training. Other demographic information was not collected systematically.

Four themes characterize the study's major findings: discrimination, differing expectations, social isolation/social support and consequences/coping strategies. Key to each is the ubiquitous experience of being part of a highly visible minority. Every participant repeatedly mentioned being "one of a few" black physicians in almost any medical or training setting. According to one resident, "You're in medicine, you're black, you're rare." Another reported,

> *At least once every time at the conference, I look around the room and I realize I'm the only person of color in this room and just by the basis of that I feel different.*

For many participants, this visibility accentuated a sense of vulnerability at the hands of more senior physicians. Although the sense of vulnerability decreased somewhat as residents gained experience, awareness of visibility persisted. For the rest of the paper, we weave this concept of visibility into the reporting of other themes.

Discrimination

Participants report various kinds of discrimination, only some of which were recognizable as overt discrimination. In fact, only six subjects described incidences of blatant discrimination in medical training: a nurse referring to black residents as "you people," a chief resident reporting that a black resident needed to be watched carefully despite stellar evaluations, black residents being conspicuously ignored by teachers in a small class

Figure 1. Interview questions

Initial Question

- Have you been treated differently during your medical training because of your racial background? (initial eight interviews)

OR

- Have you had any positive experiences during your medical training because of your racial background? (last 11 interviews)

Probe Topics (if not spontaneously mentioned by participant)

Experience and impact of the following:

- Positive experiences because of racial background
- Negative experiences because of racial background
- Interactions with students, peers, staff, faculty
- Interactions with patients
  - Pride*
- Future career choices
  - Sense of representing race as motivation*
- Local city environment
- Medical school or college experiences
- Childhood experiences of discrimination
- Suggestions for improving experience of blacks in medicine
- Representation of blacks in medicine*
- Mentorship*
- Disciplinary actions*
- Evaluations of performance*
- Being mistaken for nonphysician*

* Added for later interviews to clarify ideas noted earlier

setting, patients requesting nonblack physicians (≥1 reported this) and black residents being called derogatory names.

One participant describes an experience in which a patient with mental status changes said, "look at that nigger," in a low voice to the entire treating team, including the attending. Unsure if she had heard correctly, the participant verified the slur "by the look on the attending's face," who responded, "okay, okay" to the patient. Afterwards, the team dispersed in different directions and the participant did not discuss the incident with colleagues or the attending for fear of making anybody uncomfortable.

Another common example of discrimination involved being mistaken for a nonphysician. Fifteen participants reported being mistaken for nurses, food service workers, orderlies and housekeepers on a regular basis. One male resident noted:

> *If you walk into a patient's room and she says, "Oh, are you here to get my tray?" That's a clue of racial interaction—you're in a white coat, a certain tie, you have your nametag on, and that's the only thing that was different. So, that's absolutely clear.*

Female participants attributed being mistaken for nurses to both their gender and race. Participants grew to expect this type of mistaken identity. Many compensated by taking extra care with their dress, so as to be identified as physicians, including always wearing white coats, displaying their nametags prominently and introducing themselves as "Dr. X." One participant went so far as to wear a tie for the duration of night duty, despite the pervasive culture in the residency of wearing hospital scrubs for this rotation.

Study participants also reported more subtle forms of discrimination but often tempered these descriptions with qualifying language. Such reports fell under two broad categories: 1) instances in which the participant suspected discrimination occurred but used nonconfrontational, diplomatic or sarcastic language to qualify the description; and 2) instances in which the participant doubted whether the experience involved discrimination. An example of the former is from a male resident:

> *A case in point is that if you look at the X program here at this institution, it's just a matter of fact—I'm not making any sort of implications about anything—that the residents who've been dismissed, in the last six years, have been 100% black. Now, maybe that's a coincidence. Maybe it's not a coincidence.*

By saying "a matter of fact" and "100% black," this participant made the case for discrimination. However, he was careful to add "I'm not making any sort of implications about anything," suggesting he was hesitant to accuse this program of discrimination.

In contrast, a female resident described a similar situation in which a black resident was dismissed from a training program:

> *People say they had it in for him, as a black man. I don't know. I honestly don't know. I don't know the details. I've heard different stories. I've never talked to him. Obviously, the truth lies somewhere in the middle.*

She felt unable to judge whether this was a discriminatory act because she lacked first-hand knowledge of the situation. On the other hand, she clearly took part in conversations where others implied the resident was not given a fair chance.

Differing Expectations

Inconsistent expectations and unequal treatment of minority and majority trainees were common narrative themes. The two examples given above represent one aspect of these differing expectations: any transgression by a black trainee seemed to be more harshly punished than one by a majority trainee. Thirteen participants described unfair punishment meted out to themselves or colleagues. The most potent examples involved peers dismissed from training. Participants perceived punishment to be the default response for dealing with blacks struggling in training and believed sanctions were disproportionately greater for minorities. This comment was from a male medical resident:

> *I think if he was a white resident, I don't think it would have gotten to that point. I think someone would have stopped and said, "You need to do this, this and that." And given him support and help instead of more and more bad reports. Finally, he just got kicked out.*

Participants' perceptions led them to believe they had fewer chances to make mistakes than their white counterparts. In many cases, participants felt pressure to perform without errors and to make sure to address all the details; these concerns were expressed in terms of not showing weakness or vulnerability.

Many participants believed others had lower expectations of black students' performances. One participant explained:

> *First of all, they anticipate that they're not as good from the get-go. Then you have to prove yourself. And I think you just have to work harder at it than the majority do.*

These lowered expectations caused some participants to wonder whether they were recruited for their minority status or their abilities. A male medicine resident described this self-doubt:

Are you here because you're very good or because they'd like to recruit more people of color? So, you always think, in the back of your mind, you have to prove that you are as good or better than your colleagues, [to prove] that you deserve to be there.

Participants also mentioned instances of perceived discrimination when others' words and actions undermined their skills, knowledge and authority. Students, residents, nurses or faculty challenged participants' assessments in situations where majority colleagues would not be questioned. One female medical resident reported an incident where a student challenged her physical examination of a patient with hepatomegaly.

If I had been a big shot white male attending, he would have agreed with me. Or if I had been a white male resident, he would have agreed with me. But since I'm this small black female, he had to point out to the whole team, "the liver is small."

The patient was subsequently diagnosed with fulminant liver failure. Instead of feeling vindicated, the resident felt angry and humiliated.

Social Isolation/Social Support

Many participants felt like outsiders in the social interactions that accompanied medical training with colleagues, supervisors and supervisees. This social isolation included specific moments of discomfort in social conversations, not being invited to outings with peers as well as a generalized discomfort with not being surrounded by black faces and peers. Some subjects expressed feeling they do not have anything in common with their colleagues, which inhibited making small talk or sharing personal experiences despite the intense amount of time spent with their "team." Some reported the connections among majority physicians resulted in their white colleagues receiving guidance and support from each other to the exclusion of black members of the group. One male resident described this:

A lot of times, you'll find those individuals talking about things that you're not accustomed to and you don't find yourself being a part of. Like golf or wines, things that I, myself, have not been exposed to in my sort of social experience. I don't have the same bond or kinship that they have.

However, most participants also reported sources of very strong social support, including black patients, black support staff in the hospital, black colleagues and dedicated mentors (both black and white). Study participants uniformly reported black patients took special pride in having a black physician. From a medical resident:

I've had really good feedback from my black patients in clinic. A lot of them are like, "Oh, we think you're great;" "It's so good to see a black woman." I feel like I'm helping them. In that way, there's a good bond and I feel a sense of community.

The majority of subjects felt black support staff gave them similar encouragement. This support included expressions of admiration and support as well as concern for their personal welfare. For example, a few participants noted getting extra food in the cafeteria line or from the nursing staff. A number experienced a comforting sense of familiarity in seeing black faces among support staff.

Another source of social support was the peer network of other black physician trainees. Participants who did not attend historically black medical schools remarked on the support provided by their fellow black students. In residency, there were fewer black peers, and participants described seeking out other black physicians. Interestingly, many of the participants knew one another despite being in different years and different training disciplines in a large institution with multiple teaching hospitals. Participants remarked on situations in which black trainees were dismissed from their residency program, even if the study participants were not in the same field as the dismissed trainee. This suggested a functioning network of communication, shared experience and support.

Our subjects yearned for mentoring by black faculty, describing two special benefits gained by having black mentors: having a source of aid in case of trouble, and more importantly, possessing a guide to success. One participant explained:

I think it's more important to have black mentors because ... if you're gonna be in an academic setting, there's so much politics that you need to know. I think any mentor's gonna help you tremendously, but I think someone who's black, who's been through all that bullshit can help you a little more to navigate exactly what you need to do and who you need to talk to and how you need to present things.

Based on participants' own experiences of needing to do more to get the same recognition, they believed a person of color successful in academic medicine must have broken through more barriers and accomplished more than someone who was white. Some reported

black mentors were more critical of them than nonblack mentors in an effort to make sure they "made it." In addition, several described important relationships with white mentors who directly addressed the race issue and made promoting the participant's career a priority.

Consequences and Coping Styles

Because the experiences of black trainees were individual and multifaceted, it was difficult to trace which consequences arose from race-related encounters. Two participants, however, contemplated leaving clinical medicine due to the strain of racial tension, one by leaving medicine completely and the other by changing to pathology. Several participants noted emotional consequences such as a damaged sense of self and a lack of confidence in professional roles, as well as feelings of being on guard at all times, doubting themselves and being frustrated with the system that surrounds them. Career consequences were also varied. Five participants described choosing their training institution based on racial diversity among faculty. Some thought they would have difficulty advancing into leadership and faculty positions because of bad evaluations or a lack of systemic support.

Participants reported different coping strategies to overcome the issues described above. One important coping strategy was strengthening social networks by seeking rapport with majority colleagues and with black colleagues and faculty, as described above. A female resident who actively sought out other black students as roommates stated:

> *Even if going to class every day was a drag, or everybody I ever saw at the hospital was all white, at least I could go home and feel comfortable.*

Participants felt they needed to stand up for themselves and their race because they had few advocates. One resident made a point of correcting people who used culturally insensitive language. Another mode of assertiveness involved making sure that strengths and accomplishments were recognized properly.

For varied reasons, participants believed they have to perform at a higher standard than their majority colleagues. Participants spoke of working harder, going the extra mile, attending more carefully to details and being more aggressive in group settings to ensure their own success and to protect themselves from punishment.

A final coping strategy was to diminish the importance of potential problems. Many emphasized the positive aspects of generally negative experiences. Others reminded themselves of their own self-worth. As one participant remarked:

> *I don't preoccupy my mind with "am I being discriminated against or not?" I just do what I gotta do. I don't worry about it too much.*

DISCUSSION

Black residents from across disciplines at an academic medical center in the northeastern United States shared several common perceptions and experiences of medical training. Their experiences stemmed from interpersonal racial discrimination, underrepresentation in a predominantly white training institution and cultural differences. The impact of these experiences ranged from negligible ("ignore it"), to positive (motivation to work harder), to destructive (leaving clinical medicine). From a recent review of the literature, this is the first study of black resident physicians.

A number of our findings echoed conclusions from studies of black students in medical and graduate school and black physicians in private practice.[20,21,24] In our study, while participants reported rare instances of blatant racism and frequent doubt as to individual subtle acts, they overwhelmingly endorsed the existence of regular instances of subtle forms of racial discrimination. Post and Weddington reported similar findings in a qualitative study of 10 practicing African-American family physicians in Ohio,[24] as did Griffith and Delgado in observations of training in psychiatry.[25]

Participants in our study appeared most affected by their perception of lowered expectations and harsher punishment, which induced both anger at the system and increased motivation to work harder. Similarly, in a qualitative study of 31 black medical students, Bullock and Houston reported 25 subjects believed faculty members perceived black students as intellectually inferior.[26] Of 19 participants who believed racism impacted their studies, 10 felt motivated to work harder and nine had more difficulty as a result.

Social isolation from white peers and supervisors encountered by our participants paralleled other studies measuring perceptions of medical school by black students and graduates.[8,24,26,27] A comparison of 148 African-American and white medical school graduates found equivalent career satisfaction and achievement but a lasting dissatisfaction with the social environment of medical school among the African-American alumni. Griffith and Delgado suggested the socialization experience of residency training occurs within a group context.[25] Lack of shared experience between white and black colleagues may lead to differences in responding to the group process and then to the frustration and conflicts some of our participants report.

Participants in this study not only surmounted the obstacles of medical school but also made it into or beyond the second postgraduate year to qualify for entry into this study. Residency programs are generally smaller than medical school classes; thus, our participants were generally one of a very few in their training programs. Also, the position of resident is unique because it is a training position and a position of leadership (medical students and younger residents) and authority

(physician privileges). This unique position may have increased frustration for some participants because they felt neither protected by the attending physician nor respected by those over whom they had authority. For the most part, neither participants nor their programs explicitly considered their unique needs. Black residents' high visibility and role as community representatives may have intensified pressure to perform while increasing isolation. This motivated some to perform at higher levels while it increased self-doubt and dissatisfaction with the medical field in others.

Black residents in this study did seek support and rapport with majority colleagues but found relationships with black colleagues, mentors, support staff, family and community to be better sources of support. These relationships provided day-to-day psychological comfort, professional development and help with career planning.

Some of these experiences and perceptions are likely shared among other underrepresented minority groups and possibly women in fields with low representation of women. In fact, this is supported by studies of discrimination in groups of trainees.[13,15,18] However, these studies also showed that blacks reported higher frequency of racial discrimination than other groups. Future studies can examine the differences between varied minority groups to understand common and unique themes. Nonetheless, we feel that this information is useful to understand the black experience during residency.

This report comes from a single institution in the northeast United States and may not represent experiences in other parts of the country or in other institutions. However, the programs included different types of hospitals (private, public, tertiary care, community based and veterans administration) and likely reflected similar institutions in other urban areas. The heterogeneity in the training specialties may have obfuscated themes particular to individual specialties and should be explored with a national sample. Likewise, the small number of participants may have limited important themes that may have arisen from a larger group of participants.

Implications

Unique challenges confront black trainees, and others have commented on larger institutional changes and programmatic interventions that may enhance the representation of minorities in clinical medicine.[3,28] Further discussion on these broader societal issues is beyond the scope of this report. We can, however, assert a few lessons to share broadly with faculty and educational leaders of clinical programs.

Majority faculty and leaders need to remember that black trainees experience additional pressures and burdens during their medical training. Faculty members must be sensitive to the complex ways in which discrimination is manifest in our culture and how this discrimination and concomitant pressure systematically impact black residents. There is a continual need for support, rapport and connection among black residents to faculty and peers to sustain the growth and well-being of the black physician community. Support must come from both majority and minority physicians. While trainees of all races require support, a support network is less available to black residents. Regardless of the source, any degree of social integration is likely to impact residents' work and well-being. Hence, even small residency programs may benefit from establishing formal social and professional support networks for black residents.

We believe the most immediately applicable result of our findings is the initiation of dialogue by majority faculty with residents, both individually and in small groups. As a result of this work, two white study authors (JL, JO), have altered our interactions with black residents. We now ask residents how race impacted their experience here and elsewhere and have used attending rounds and conference times to lead open discussions on the issue. For us personally, these encounters have been rewarding and well received by the trainees involved. This approach resonates with the precepts of learner-centered learning by putting forth an issue that is critically important to the trainee but often ignored or easily overlooked by majority faculty.

ACKNOWLEDGEMENTS

Thanks to Elliot Mishler and the Narrative group for comments on data interpretation. Thanks also to Julien Dedier, MD; Cathy Riessman, PhD; Tracy Battaglia MD; and Tali Averbuch, MPP for critical reading of the manuscript.

REFERENCES

1. Lloyd SM Jr, Miller RL. Black student enrollment in U.S. medical schools. *JAMA*. 1989;261(2):272-274.

2. Brotherton SE, Simon FA, Etzel SI. U.S. graduate medical education, 2001–2002: changing dynamics. *JAMA*. 2002;288(9):1073-1078.

3. Tekian A. A thematic review of the literature on underrepresented minorities and medical training, 1981–1995: securing the foundations of the bridge to diversity. *Acad Med*. 1997;72(suppl 10):S140-S146.

4. Tekian A. Attrition Rates of Underrepresented Minority Students at the University of Illinois at Chicago College of Medicine, 1993–1997. *Acad Med*. 1998;73(3):336-338.

5. Castillo-Page L, Zhang K, Steinecke A, et al. *Minorities in Medical Education: Facts & Figures 2005*. Washington, DC: Association of American Medical Colleges; 2005.

6. Shields PH. A survey and analysis of student academic support programs in medical schools. Focus: underrepresented minority students. *J Natl Med Assoc*. 1994;86(5):373-377.

7. Calkins EV, Willoughby TL. Predictors of black medical student success. *J Natl Med Assoc*. 1992;84(3):253-256.

8. Strayhorn G, Frierson H. Assessing correlations between Black and white students' perceptions of the medical school learning environment, their academic performances, and their well-being. *Acad Med*. 1989;64:468-473.

9. DeFour DC, Hirsch BJ. The adaptation of black graduate students: a social network approach. *Am J Community Psychol*. 1990;18(3):487-503.

10. Frierson HT, Jr. Black medical students' perceptions of the academic environment and of faculty and peer interactions. *J Natl Med Assoc*. 1987;79(7):737-743.

11. Kornitzer B, Ronan E, Rifkin MR. Improving the adjustment of education-

ally disadvantaged students to medical school: the Summer Enrichment Program. *Mt Sinai J Med.* 2005;72(5):317-321.

12. Sheehan H, Sheehan DV, White K, et al. A pilot study of medical student 'abuse': student perceptions of mistreatment and misconduct in medical school. JAMA. 1990;263(4):533-537.

13. VanIneveld CHM, Cook DJ, Kane S-LC, et al. Discrimination and Abuse in Internal Medicine Residency. *J Gen Intern Med.* 1996;11(7):401-405.

14. Kassebaum DG, Cutler ER. On the culture of student abuse in medical school. *Acad Med.* 1998;73(11):1149-1158.

15. Baldwin DCJ, Daugherty SR, Rowley BD. Residents' and Medical Students' Reports of Sexual Harassment and Discrimination. *Acad Med.* 1996;71(suppl 10):S25-S27.

16. Richardson DA, Becker M, Frank RR, et al. Assessing medical students' perceptions of mistreatment in their second and third years. *Acad Med.* 1997;72(8):728-730.

17. Reitzes DC, Elkhanialy H. Black students in medical schools. *J Med Educ.* 1976;51:1001-1005.

18. Baldwin DCJ. Racial and ethnic discrimination during residency: Results of a national survey. *Acad Med.* 1994;69(10 suppl):S19-S21.

19. Crowley S, Fuller D, Law W, et al. Improving the climate in research and scientific training environments for members of underrepresented minorities. *Neuroscientist.* 2004;10(1):26-30.

20. Strayhorn G. Perceived stress and social supports of black and white medical students. *J Med Educ.* 1980;55:619-620.

21. Coombs AA, King RK. Workplace discrimination: experiences of practicing physicians. *J Natl Med Assoc.* 2005;97(4):467-477.

22. Blank R, Dabady M, Citro C, eds. *Measuring Racial Discrimination.* Washington, DC: National Academies Press; 2004.

23. Strauss A, Corbin J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Newbury Park, CA: Sage Publications; 1990.

24. Post DM, Weddington WH. Stress and Coping of the African American Physician. *J Natl Med Assoc.* 2000;92(2):70-75.

25. Griffith EEH, Delgado A. On the professional socialization of black residents in psychiatry. *J Med Educ.* 1979;54(6):471-476.

26. Bullock SC, Houston E. Perceptions of racism by black medical students attending white medical schools. *J Natl Med Assoc.* 1987;79(6):601-608.

27. Gartland JJ, Hojat M, Christian EB, et al. African American and white physicians: a comparison of satisfaction with medical education, professional careers, and research activities. *Teach Learn Med.* 2003;15(2):106-112.

28. Dennis GC. Racism in medicine: planning for the future. *J Natl Med Assoc.* 2001;93(3 suppl):1S-5S. ■

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 Health Literacy on Depressive Symptoms and Mental Health-related Quality of Life Among Adults with Addiction

Alisa Lincoln, PhD MPH,[1,7] Michael K. Paasche-Orlow, MD,[3] Debbie M. Cheng, ScD,[4] Christine Lloyd-Travaglini, MPH,[4] Christine Caruso, MA, MPH,[1] Richard Saitz, MD, MPH,[2,3,5] Jeffrey H. Samet, MD, MA, MPH[1,3]

[1]Department of Social and Behavioral Sciences, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [2]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA; [3]Section of General Internal Medicine (Clinical Addiction Research and Education Unit), Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [4]Department of Biostatistics, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [5]Department of Epidemiology, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [6]Data Coordinating Center, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [7]Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA.

BACKGROUND: Health literacy has been linked to health status in a variety of chronic diseases. However, evidence for a relationship between health literacy and mental health outcomes is sparse.

OBJECTIVE: We hypothesized that low literacy would be associated with higher addiction severity, higher levels of depressive symptoms, and worse mental health functioning compared with those with higher literacy in adults with alcohol and drug dependence.

METHODS: The association of literacy with multiple mental health outcomes was assessed using multivariable analyses. Measurement instruments included the Rapid Estimate of Adult Literacy in Medicine (REALM), the Center for Epidemiologic Studies-Depression (CES-D) scale, the Mental Component Summary scale of the Short Form Health Survey, and the Addiction Severity Index for drug and alcohol addiction. Subjects included 380 adults recruited during detoxification treatment and followed prospectively at 6-month intervals for 2 years. Based on the REALM, subjects were classified as having either low ($\leq$ 8th grade) or higher ($\geq$ 9th grade) literacy levels.

RESULTS: In longitudinal analyses, low literacy was associated with more depressive symptoms. The adjusted mean difference in CES-D scores between low and high literacy levels was 4 ($P<.01$). Literacy was not significantly associated with mental health-related quality of life or addiction severity.

CONCLUSIONS: In people with alcohol and drug dependence, low literacy is associated with worse depressive symptoms. The mechanisms underlying the relationship between literacy and mental health outcomes should be explored to inform future intervention efforts.

KEY WORDS: literacy, alcohol dependence, drug dependence, addiction, mental health, depressive symptoms.
DOI: 10.1111/j.1525-1497.2006.00533.x
J GEN INTERN MED 2006; 21:818–822.

Approximately 90 million American adults have low health literacy and lack the basic literacy skills to function in society.[1–3] A growing body of work exists linking low literacy with an array of negative outcomes.[4] These outcomes include more frequent hospitalization,[5,6] higher rates of health services utilization, and worse prevention practices in people with diabetes,[7,8] asthma,[9,10] cancer,[11–13] and other chronic illnesses.[14,15] To date, few studies have examined the role of health literacy in addiction and mental illness.

Several studies have evaluated the association between low literacy and mental illness. While most of these studies demonstrate associations between low literacy and depression, all of them relied on cross-sectional analyses and are thus limited in terms of causal inferences. Research on health literacy and substance abuse has been similarly limited, with cross-sectional studies suggesting a high prevalence of low literacy in people with addiction. Understanding mechanisms responsible for the relationship between limited literacy and mental health outcomes is critical in informing the development of future interventions for mental health and addictive disorders. We examined the relationship between health literacy and addiction severity, depressive symptoms, and mental health functioning among people with drug and alcohol dependence over a 2-year period. We hypothesized that low health literacy would be associated with higher addiction severity, more depressive symptoms, and lower mental health-related quality of life (MHQOL). A secondary hypothesis was that utilization of mental health services would be a mediator of these relationships.

METHODS

This study was a prospective cohort analysis of data collected in the Health Evaluation and Linkage to Primary care (HELP) study. The HELP study, conducted in a 35-bed inner-city short-term inpatient detoxification unit, was a randomized-controlled trial of a multidisciplinary clinical assessment designed to link substance abusing persons to primary medical care. The results of this intervention have been reported elsewhere.[16] Subjects randomized to the HELP intervention from 6/97 through 3/99 received a 90-minute clinical session with a physician, nurse, and social worker, along with an appointment and referral letter for primary care, before leaving the detoxification unit. Control subjects did not receive this intervention but were treated similarly in all other respects. The usual length of stay for a detoxification admission was 6 days for heroin dependence and 4 days for alcohol dependence. The HELP trial eligibility criteria included inpatient detoxification admission, age greater than 17 years, and report of alcohol, heroin, or cocaine as the substances of first or second choice. Exclusion criteria were as follows: having a primary care provider and having seen that provider on at least 1 occasion in the preceding 2 years; pregnancy; Mini-Mental State examination[17] score less than 21; lack of fluency in either English or Spanish; less than 3 contacts available to facilitate follow-up; or specific plans to leave the Boston area within 24 months.

None of the authors have any conflicts of interest to declare.

Address correspondence and requests for reprints to Dr. Lincoln: BUSPH, 715 Albany Street, T244W, Boston, MA 02118 (e-mail: alincoln@bu.edu).

Trained research associates identified 642 trial-eligible subjects within 24 to 48 hours of admission for detoxification, of whom 470 (73%) consented to be in the HELP trial (see Fig. 1). Research associates performed a 60 to 90 minutes interview with all subjects before their discharge. Follow-up interviews were conducted with subjects every 6 months after the baseline interview for up to 24 months. Follow-up rates, computed as the proportion of the original cohort alive at each follow-up period, were 54% at 6 months, 46% at 12 months, 54% at 18 months, and 59% at 24 months. During these interviews, depressive symptoms, mental health-related quality of life, and addiction severity were evaluated. Interviews were conducted in English or Spanish; however, only subjects who completed the English version of the interview were included in these analyses, as the Rapid Estimate of Adult Literacy in Medicine (REALM) instrument is only available in English (n=453). Of these, 3 subjects refused to complete the REALM instrument. Thus, baseline analyses were conducted on the final sample of 450 subjects who completed the interview, including the REALM, in English. Of these subjects, 380 (84%) completed at least 1 follow-up interview. Longitudinal analyses are based on this subset of 380 subjects. Subjects were compensated in the form of supermarket certificates for their time, $20 for the initial interview, and $30 for subsequent ones. The Boston University Medical Campus Institutional Review Board approved this research. A certificate of confidentiality from the federal government provided additional protection for subjects' privacy.

Primary Independent Variable

Health literacy was measured at study entry using the REALM. The REALM is a screening test used in public health and primary care and other settings to assess a patient's reading level.[18] This instrument was chosen for its validity and brevity. It provides reading grade estimates for patients who read below the 9th-grade level, and thus is particularly useful at helping to identify potential problems with reading among a population hypothesized to have low levels of health literacy. Two health literacy categories were defined based on the REALM: low ($\leq$ 8th grade: REALM score 0 to 60) versus higher ($\geq$ 9th grade: REALM score 61 to 66).

Outcome Variables

There were 4 outcomes of interest measured at study entry (baseline), 6, 12, 18, and 24 months. The Addiction Severity Index drug scale (ASI-Drug) and the Addiction Severity Index alcohol scale (ASI-Alc) assess drug and alcohol addiction severity, respectively, with composite scores for each ranging from 0 to 1, higher scores indicating greater severity.[19] The Mental Component Summary (MCS) of the Short Form Health Survey (SF-36) assesses mental health-related quality of life, scores ranging from 0 to 100, with higher scores indicating better quality of life.[20,21] Finally, the Centers for Epidemiologic Studies-Depression (CES-D) scale measures depressive symptoms with higher CES-D scores indicating greater levels of distress. Scores range from 0 to 60 with a score $\geq$16 traditionally interpreted as a clinically significant level of distress.[22]

Analyses

Baseline characteristics of subjects (n=380) with at least 1 follow-up interview were compared between the low and higher health literacy groups. Preliminary analyses were conducted to assess the baseline cross-sectional associations between health literacy and mental health and addiction severity measures. For the primary analyses, generalized linear models for longitudinal data were used to model each outcome variable. The longitudinal regression models account for the correlation due to repeated observations on the same subject. An unstructured covariance structure was used to fit the longitudinal regression models. The primary analyses controlled for factors known to be related to both literacy and mental health and addiction outcomes by including them as covariates in the regression models. The covariates included in all regression models were as follows: gender, age, race (black, white, Hispanic, or other), years of formal education, income ($\leq$ $19,000, $20,000 to $49,000, or $\geq$ $50,000), primary language (English vs other), primary substance of choice (alcohol, cocaine, or heroin), cognitive functioning measured using the Mini-mental Status Examination (MMSE), time of follow-up interview (number of months since baseline), and randomization group. The baseline values for addiction severity, MHQOL,

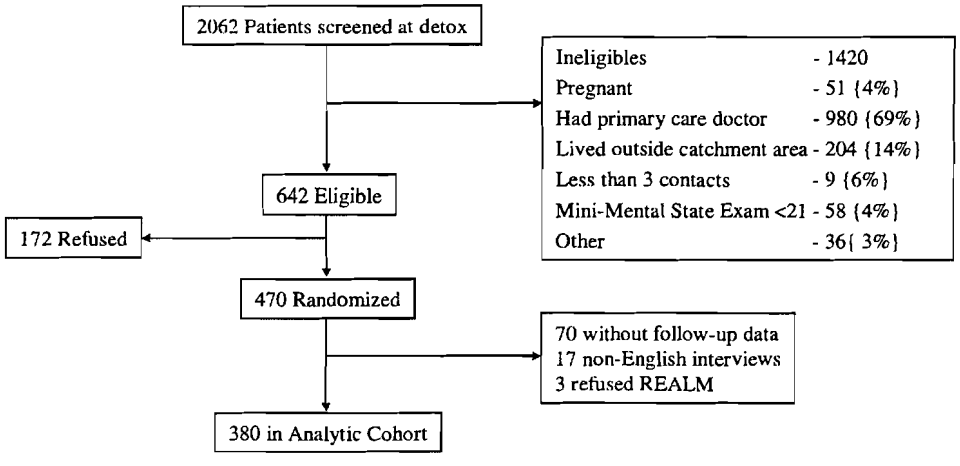


FIG. 1. Selection of study subjects.

Table 1. Description of Study Cohort at Baseline (*N*=380)

	Total Sample	Low Literacy, *N*=174 (46%)	Higher Literacy, *N*=206 (54%)
Mean age (SD)	36 (7.64)	36 (7.84)	36 (7.48)
Race/ethnicity, *N* (%)			
Black	201 (53)	122 (70)	79 (38)
White	132 (35)	21 (12)	111(54)
Hispanic	5 (6)	18 (10)	5 (2)
Other	11 (6)	13 (8)	11 (5)
Male, *N* (%)	287 (76)	135 (78)	152 (74)
Primary language, *N* (%)			
English	354 (93)	150 (86)	204 (99)
Other language	26 (7)	24 (14)	2 (1)
Mean years of formal education (SD)	11.98 (1.98)	11.43 (1.94)	12.44 (1.90)
Income, *N* (%)			
<$19,000	211 (58)	104 (62)	107 (54)
$20,000 to $49,000	123 (34)	54 (32)	69 (35)
>$50,000	33 (9)	9 (5)	24 (73)
Mean CES-D (SD)	33.03 (12.56)	30.91 (11.26)	34.82 (13.32)
Mean ASI-Alc (SD)	0.47 (.34)	0.46 (.34)	0.48 (.34)
Mean ASI-Drug (SD)	0.26 (.14)	0.26 (.13)	0.26 (.15)
Mean MCS (SD)	31.19 (12.75)	33.02 (12.97)	29.67 (12.39)
Primary substance of choice, *N* (%)			
Alcohol	141 (37)	51 (30)	90 (44)
Cocaine	135 (36)	76 (44)	59 (29)
Heroin	103 (27)	46 (27)	57 (28)

ASI, Addiction Severity Index drug scale; ASI-Alc, Addiction Severity Index alcohol scale; CES-D, Center for Epidemiologic Studies-Depression; MCS, Mental Component Summary.

and depressive symptoms were also included as covariates in regression models.

Finally, we assessed whether service utilization mediated the relationships between low literacy and outcome measures, by including it as an additional covariate in regression models and assessing whether the association between literacy and mental health outcomes changed. Service utilization was defined as any self-reported use of either medical services or behavioral health services (mental health and addiction services) in the past 6 months before the study interview. These variables, assessed at baseline and every 6 months during follow-up, were included in the regression model as a time-varying covariate. All analyses were conducted using 2-sided tests and a significance level of .05. Analyses were completed using SAS/STAT software, Version 8.2.

RESULTS

Subjects' demographic characteristics reflect the nature of the population served in a public urban residential detoxification unit (Table 1). Among the 174 subjects with low literacy, 52 (30%) had REALM scores indicating a reading level ≤6th grade, while 122 (70%) had a 7th to 8th-grade reading level. In unadjusted analyses of baseline data, low literacy was associated with higher MCS (P=.02) and lower CES-D (P=.004) scores but was not significantly associated with ASI-Drug (P=.98) or ASI-Alc (P=.38) (Table 2). In adjusted analyses of baseline data, no associations were detected between health literacy and ASI-Drug, ASI-Alc, MCS, or CES-D.

In longitudinal regression analyses, baseline health literacy was not significantly associated with mental health-related quality of life (MCS) or alcohol or drug addiction severity (ASI). However, low health literacy was associated with higher levels of depressive symptoms, consistent with the study hypotheses. In a longitudinal model adjusting for the full set of potential confounders, the adjusted mean CES-D was 26.7 for the low as compared with 22.7 for the higher health literacy group ($P \leq .01$) (Table 3). For both low and higher health literacy groups, the level of depressive symptoms appeared to improve following the baseline interview. Higher levels of depression for

Table 2. Baseline Relationship of Literacy and Mental Health Outcomes (*N*=380)

	ASI-Drug* (Range=0 to 1)		ASI-Alc* (Range=0 to 1)		CES-D† (Range=0 to 60)		MCS‡ (Range=0 to 100)	
	Unadjusted	Fully Adjusted§	Unadjusted	Fully Adjusted§	Unadjusted	Fully Adjusted§	Unadjusted	Fully Adjusted§
Low (≤8th grade)	0.25	0.25	0.46	0.47	30.9	31.4	32.9	30.4
Higher (≥9th grade)	0.25	0.28	0.49	0.48	34.3	33.8	30.1	29.2
P-value	.98	.11	.38	.88	.004	.09	.02	.42

**Higher scores indicate greater severity.*
†Higher scores indicate greater depressive symptoms.
‡Higher scores indicate greater quality of life.
§Models adjusted for sex, age, race, education, income, primary language, primary substance of choice, randomization group, and mini-mental status examination.
ASI, Addiction Severity Index drug scale; ASI-Alc, Addiction Severity Index alcohol scale; CES-D, Center for Epidemiologic Studies-Depression; MCS, Mental Component Summary.

Table 3. Longitudinal Models of Literacy and Mental Health Outcomes (N=380)

	ASI-Drug*		ASI-Alc*		CESD†		MCS‡	
	Minimally Adjusted§	Fully Adjusted‖	Minimally Adjusted§	Fully Adjusted‖	Minimally Adjusted§	Fully Adjusted‖	Minimally Adjusted§	Fully Adjusted‖
Low (≤ 8th grade)	0.13	0.13	0.23	0.26	24.3	26.7	40.7	39.1
Higher (≥ 9th grade)	0.12	0.12	0.26	0.25	21.0	22.7	42.0	41.2
P-value	.28	.35	.18	.86	.01	<.01	.22	.14

**Higher scores indicate greater severity.*
†Higher scores indicate greater depressive symptoms.
‡Higher scores indicate greater quality of life.
§Models adjusted for baseline measure of outcome variable.
‖Models adjusted for time, sex, age, race, education, income, primary language, primary substance of choice, randomization group, mini-mental status examination, and baseline measure of outcome variable.
ASI, Addiction Severity Index drug scale; ASI-Alc, Addiction Severity Index alcohol scale; CES-D, Center for Epidemiologic Studies-Depression; MCS, Mental Component Summary.

the low literacy group compared with the higher literacy group were observed only after the initial assessment (see Fig. 2).

We examined reported service utilization for both low and higher literacy groups at each time point and found no significant differences between groups, with an average of 78% of low-literacy and 82% of higher-literacy subjects using any health or addiction services over a 6-month period. Accordingly, the addition of service utilization to the other covariates listed in Table 3 did not attenuate the association between health literacy and depressive symptoms. The mean CES-D scores were 26.6 and 22.8 for the low and higher health literacy groups, respectively, after adjusting for self-reported service utilization (P<.01).

DISCUSSION

The recent AHRQ evidence review, "Literacy and Health Outcomes: Evidence Report/Technology Assessment,"[23] cites 5 studies that evaluate the association between a marker of health literacy and a marker of mental illness. Four of these studies report statistically significant associations between low literacy and higher prevalence of depression; however, not all of these associations remained significant in adjusted analyses. Each of these studies and more recent work by Wolf et al.[24] have relied on cross-sectional analyses, and thus do not allow for consideration of causal inferences. For example, Gazmarian et al.[25] found that 13% of new Medicare recipients had depression according to the geriatric depression scale. Subjects with low health literacy were 3 times more likely to have depression. However, after controlling for demographics, social support, health behavior, and health status, health literacy did not remain an independent risk factor for depressive symptoms.

In this study, we examined the relationship between health literacy and several behavioral health outcomes over time. These data provide evidence supporting an important longitudinal relationship between literacy skills and depressive symptoms among adults with addiction. We did not detect an association between health literacy and addiction severity or mental health-related quality of life. These are some of the first data to examine prospectively and longitudinally the relationship of health literacy and mental health outcomes.

Subjects in this study were recruited at a residential detoxification unit and accordingly had severe symptoms, at a nadir for both addiction and mental health. Depressive symptoms improved for both the low and higher health literacy groups over time; however, subjects with low health literacy had higher levels of depressive symptoms during follow-up. Prior research provides evidence that the differences in CES-D scores found in our study are clinically significant. Pandya et al.[26] reported an increase of approximately 6% in the prevalence of major depressive disorder among people with multiple sclerosis when CES-D scores increased from 22 to 26.

These data do not allow adequate examination of potential mechanisms to understand the relationship between health literacy and depressive symptoms. We explored health literacy as a barrier to service utilization, as this has been suggested as an important possible mechanism for the relationship between health literacy and chronic disease outcomes,[27] but did not find this to explain the observed association. Although service utilization can offer important insights into the relationship

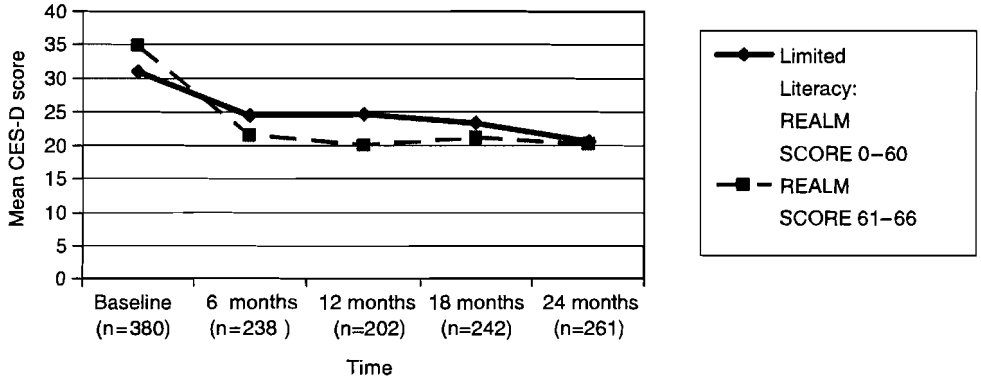


FIG. 2. Literacy and depressive symptoms across time.

between health literacy and chronic disease outcomes, other potential mechanisms remain to be explored. One of these, which may be especially salient in examining mental health outcomes, is the high levels of shame associated with limited reading ability.[28] This shame may be in the causal chain of increased depressive symptomatology.

There are several limitations of this study. First, the generalizability of these results is limited by the specific nature of the population studied, a cohort of patients recruited during inpatient detoxification. In addition, we cannot assess the extent to which the level of literacy or depressive symptoms may have influenced study participation. Second, this secondary analysis of prospectively collected data only allowed the use of CES-D, MCS, and ASI as measures of mental health and addiction outcomes. While CES-D is a good measure of depressive symptoms, future work should examine a broader spectrum of mental health and addiction outcomes.

Also, the REALM was only administered at baseline and we are unable to evaluate changes in literacy over time. While it is certainly possible that subjects may have acquired additional literacy skills through adult education, it is unlikely that significant changes occurred in literacy scores over time in our cohort. While the REALM is merely a test of word-recognition, it had been shown to correlate highly with tests that evaluate other domains of literacy such as numeracy, reading comprehension, and document literacy.[29]

Finally, our preliminary analyses indicate that level of mental health service utilization does not explain the influence of health literacy on depressive symptoms. These findings are limited by our use of a single dichotomous, self-report measure of health or addiction service use and should be interpreted cautiously.

CONCLUSION

Little attention has been paid to health literacy in substance abuse and mental health care settings. These data indicate that health literacy may be an important factor in the course and outcomes of depressive symptoms. While health literacy was not found to be a predictor of mental heath functioning or addiction outcomes, clearly this study represents an initial exploration of this area, one that merits further investigation.

The important role of low health literacy in people's lives has been inadequately addressed. Future work should examine the role of low health literacy among subjects with a full range of psychiatric and addiction disorders. In addition, potential mechanisms to explain these relationships should be examined in order to inform the development of interventions designed to reduce the burden of low health literacy among people with addictive and mental health diseases.

This work was supported by a grant from the National Institute on Alcohol Abuse and Alcoholism (RO1-AA10870) and the National Institute on Drug Abuse (RO1-DA10019) (R25-DA13582). This work was supported in part by the Boston University General Clinical Research Center from the National Center for Research Resources (MO1-RR00533).

REFERENCES

1. **Kefalides PT.** Illiteracy: the silent barrier to health care. Ann Intern Med. 1999;130:333–6.
2. **Nielsen-Bohlman LT, Panzer AM, Hamlin B, et al.** Health Literacy: A Prescription to End Confusion. Committee on Health Literacy, Board on Neuroscience and Behavioral Health. Washington, DC: National Academies Press; 2004.
3. **Paasche-Orlow MK, Parker RM, Gazmararian JA, et al.** The prevalence of limited health literacy. J Gen Intern Med. 2005;20:175–84.
4. **DeWalt DA, Berkman ND, Sheridan S, et al.** Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med. 2004;19:1228–39.
5. **Baker DW, Parker RM, Williams MV, et al.** Health literacy and the risk of hospital admission. J Gen Intern Med. 1998;13:791–8.
6. **Baker DW, Gazmararian JA, Williams MV, et al.** Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health. 2002;92:1278–83.
7. **Rothman RL, DeWalt DA, Malone R, et al.** Influence of patient literacy on the effectiveness of a primary care-based diabetes disease management program. JAMA. 2004;292:1711–6.
8. **Schillinger D, Grumbach K, Piette J, et al.** Association of health literacy with diabetes outcomes. JAMA. 2004;288:475–82.
9. **Williams MV, Baker DW, Honig EG, et al.** Inadequate literacy is a barrier to asthma knowledge and self-care. Chest. 1998;114:1008–15.
10. **Paasche-Orlow M, Brancati F, Rand C, et al.** Education of patients with asthma and low literacy. J Gen Intern Med. 2003;18(Suppl):S227.
11. **Davis TC, Arnold C, Berkel HJ, et al.** Knowledge and attitude on screening mammography among low-literate, low-income women. Cancer J. 1996;78:1912–20.
12. **Sharp LK, Zurawski JM, Roland PY, et al.** Health literacy, cervical cancer risk factors, and distress in low-income African-American women seeking colposcopy. Ethnicity Disease. 2002;12:541–6.
13. **Scott TL, Gazmararian JA, Williams MV, et al.** Health literacy and preventive health care use among Medicare enrollees in a managed care organization. Med Care. 2002;40:395–404.
14. **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.
15. **Kalichman SC, Rompa D.** Functional health literacy is associated with health status and health-related knowledge in people living with HIV-AIDS. J Acquir Immune Defic Syndr. 2000;25:337–44.
16. **Samet JH, Larson M, Horton NJ, et al.** Linking alcohol and drug dependent adults to primary medical care: a randomized controlled trial of a multidisciplinary health intervention in a detoxification unit. Addiction. 2003;98:509–16.
17. **Folstein MF, Folstein SE, McHugh PR.** Mini-Mental State: a practical method for grading the state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
18. **Davis TC, Jackson RH, George RB, et al.** Reading ability in patients in substance misuse treatment centers. Int J Addict. 1993;28:571–82.
19. **McLellan AT, Kushner H, Metzger D, et al.** The fifth edition of the Addiction Severity Index. J Subst Abuse Treat. 1992;9:199–213.
20. **Ware JE.** SF-36 Health Survey. Manual and Interpretation Guide. Boston, MA: The Health Institute; 1993.
21. **Ware JE.** SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: The Health Institute; 1994.
22. **Rafloff LS** The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measure. 1977;1:385–401.
23. **Berkman ND, DeWalt DA, Pignone MP, et al.** Literacy and Health Outcomes. Evidence Report/Technology Assessment No. 87 (Prepared by RTI International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication No. 04-E007-2. January 2004. Rockville, MD: Agency for Healthcare Research and Quality.
24. **Wolf MS, Gazmarian JA, Baker DW.** Health literacy and functional health status among older adults. Arch Intern Med. 2005;165: 1946–52.
25. **Gazmararian J, Baker D, Parker R, Blazer DG.** A multivariate analysis of factors associated with depression: evaluating the role of health literacy as a potential contributor. Arch Intern Med. 2000;160:3307–14.
26. **Pandya R, Metz L, Patten S.** Predictive value of the CES-D in detecting depression among candidates for disease-modifying multiple sclerosis treatment. Psychosomatics. 2005;46:131–4.
27. **Baker DW.** Reading between the lines: deciphering the connections between literacy and health. J Gen Intern Med. 1999;14:315–7.
28. **Parikh NS, Parker RM, Nurss JR, et al.** Shame and health literacy: the unspoken connection. Patient Educ Couns. 1996;27:33–9.
29. **Davis TC, Long SW, Jackson RH, et al.** Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med. 1993b; 25:391–5.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Biomedical Informatics 39 (2006) 468–481
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

How to write health dialog for a talking computer

Jeffrey P. Migneault *, Ramesh Farzanfar, Julie A. Wright, Robert H. Friedman

Medical Information Systems Unit, Department of Medicine, Boston University and Boston Medical Center, USA

Received 15 August 2005
Available online 23 February 2006

Abstract

Automated dialogue systems delivered over the telephone offer a promising approach to delivering health-related interventions to populations of individuals at low-cost. Over the past two decades, an automated telephone system called Telephone-Linked Care or TLC has been successfully designed and evaluated by the authors and their colleagues. This work has resulted in over twenty systems for various health-related conditions and lifestyle behaviors. This paper describes our approach to developing and writing dialogue for these automated telephone systems, including determining the program objectives, defining the target population, and selecting a theory of behavior change to guide the intervention. Both macro and micro issues are considered in constructing dialogue systems that are engaging for the target population, easy to use, and effective at promoting positive health behaviors and outcomes.

Keywords: Dialogue systems; Behavioral change systems; Telephone-Linked Care

1. Introduction

Automated dialogue systems are increasingly being used in health care to provide information, advice, counseling, disease monitoring, clinical problem identification, as well as enhancing patient-provider communication. They are also being used with the general population (consumers) to improve health-related lifestyle behaviors. We have 20 years of experience developing and evaluating completely automated telephone-based conversational systems that interact with patients and consumers to improve health outcomes and the delivery of health services. In this paper, we will present our experiences designing these systems, which we call Telephone-Linked Care or TLC systems. We will describe the process of developing conversational structure and content as well as the process of implementation.

The systems we have developed and evaluated have focused on (1) positively influencing a person's health behavior by modifying behavioral risk factors for disease (e.g., smoking, diet quality, etc.) and promoting disease-related self-care behaviors (e.g., taking prescribed medications regularly, attending scheduled clinical office visits, etc.), and (2) monitoring patient's health conditions outside of clinical settings and identifying potential acute, subacute, and chronic medical care issues that are then communicated to the person's responsible health professional(s). For this paper, our focus will be on the dialogues necessary for the first of these objectives, namely improving health behaviors of patients and consumers, although most of the considerations also apply to the second objective of disease monitoring, an integral component of disease management programs. For the improvement of lifestyle behaviors, we have developed TLC systems for several aspects of diet [1–3], physical activity [4–6], and cigarette smoking [7]. For disease related self-care behaviors, there are TLC programs for promoting medication-regimen adherence [3,8–10], scheduled office visit attendance [10], appropriate disease screening behavior [11], and use of home self-measurement devices [8]. TLC chronic disease management programs exist for hypertension [8], angina pectoris, chronic obstructive lung disease [12,13], asthma [14], diabetes mellitus, depression [10], and patients with multiple chronic diseases.

* Corresponding author. Fax: +1 617 638 5580.
E-mail address: jpm@bu.edu (J.P. Migneault).

1532-0464/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2006.02.009

Over time our program objectives have evolved from being single behavioral objectives (e.g., improve medication adherence) to multi-behavioral interventions (medication adherence, diet, and exercise). This evolution reflects the increasing functionality of TLC programs and a recognition that improving health or self-care of chronic conditions (e.g., hypertension) is a multi-behavioral enterprise.

Evaluations of TLC systems have demonstrated efficacy in improving target health behavior in multiple studies. Evaluations of TLC systems have compared TLC to "usual care" and to mock-interventions that used TLC technology to control for the attention-placebo effect of the contact itself and the novelty of the TLC intervention. Using the standard metric of effect sizes (behavior change expressed in terms of standard deviations on the outcome variables), TLC systems have demonstrated medium to large effect sizes ranging from .29 to .69 [15]. More recently, a TLC physical activity promotion program was compared to both an assessment-only control and a human telephone counselor program [16]. The results demonstrated that both the human and TLC groups were effective compared to the control and were not significantly different from each other.

2. Automated telephone systems or telephony

The use of the telephone as the communication channel of health behavior change programs has a number of advantages. The telephone is almost universally available in the US, and in both developing and developed countries. In fact, with the spread of mobile telephone technology globally, many developing countries have built national mobile telephone systems while their landline telephone systems remain rudimentary. Because telephones are ubiquitous and have been a mainstay of communication between people, typically two people at a time, using a phone for conversation with another person is perceived as natural and easy. Our research has shown that users actively project human qualities onto the automated voice and this lends credibility to the system [17,18]. Although it has not been empirically tested, we believe that this projection is more likely to occur in telephone encounters than in programs that use other technological channels of communications. Interestingly, this effect might not be entirely positive, as in some situations individuals may be more honest with less personal (computers vs. live interviewers) assessment methods [19,20].

Unlike human-to-human interactions, the TLC system's conversations with patients and consumers are entirely automated. During TLC telephone conversations, the system speaks to individuals over the telephone using digitally recorded human voices. Users communicate with TLC by pressing buttons on their telephone keypad, or in newer systems, by speaking into the telephone receiver.

The interaction is designed to resemble a typical conversation between a health counselor and a patient/client. The use of speech recognition technology helps make the interaction "feel" like a natural conversation. At the core of these systems is the dialogue, the exact words that the system 'speaks' to the user during TLC conversations. We refer to the written version of this dialogue as the 'script.' The script contains the conversational segments to be digital recorded by voice consultants (actors) into separate sound files as well as the rules for assembling these segments into the resulting, often unique, conversation. This includes any questions that are going to be asked during the conversation and the possible answers that the system will accept. How skillfully the dialogue is designed and written largely determines the nature and effectiveness of telephony systems.

Our current TLC programs are comprised of a computer system that combines an interactive voice response (IVR) subsystem for generating speech using prerecorded audio message segments, a speech recognition subsystem for recognizing what the user is saying, a database management subsystem for storing and managing system and user data, and a conversation control subsystem that controls the content and flow of individual TLC conversations with users. TLC systems are finite state machines in which programmed decision rules are used to select, combine and play pre-recorded sound files of conversational segments. These rules determine which sound files to play based on logic that is applied at each step of the conversation and data the system has about the user and their progression through the program. The data on user characteristics might be known before a contact is initiated or collected from the user during the call.

We design TLC systems so that they are tailored to the user's 'state' using current behavioral theories to structure the design of the systems and the strategies used to promote change. Since behavioral theories do not usually delineate the specific tactics that should be used by the systems to affect the behavioral strategies, we incorporate the heuristics of experienced clinicians in the tactical design and in crafting the specific words uttered by TLC. Given the complexity of these design considerations, the variety of component behaviors a program might be addressing, and the number of user states given the factors we are tailoring on, TLC systems are quite large and complex. The printed representations of TLC 'scripts' are usually in the range of 150–600 pages. We know of no other evaluated health dialogue systems that approach TLC programs in their size, breadth, depth, complexity of content or in the degree of incorporation of behavioral change theories and heuristics of health professionals. This paper will describe the most salient attributes of these systems and our process for developing them.

3. Initial considerations

Before beginning to write dialogue for health behavior change systems, a number of parameters should be carefully examined and clearly delineated, including the objectives of the program, the characteristics of the population that

will be using the system, and the intervention approach that will be employed. Having these parameters well defined will make the development process more efficient and lessen the chance that changes will be needed after the program has been initiated, which is quite costly for automated systems.

3.1. Program objectives

Although any intervention program will at the outset have an overall objective, it is important for an automated health behavior intervention program to have objectives that are carefully considered and explicitly delineated. The program's objectives need to be kept in mind throughout the design and writing phases of intervention development. Adjustments to the objectives often occur during the developmental process and should be explicitly documented. This process is imperative when multiple developers are working on the system. The objectives should be defined not just in terms of the overall goal (e.g., promoting physical activity) but also include: (1) what aspects of a behavior or behaviors will be changed, (2) what are the ultimate goals and what are the intermediate goals, (3) in what order will the goals be accomplished, and (4) how will the achievement of the goals be assessed.

The first of these considerations is important because most lifestyle behaviors are quite complex and any one program is likely to focus on only some aspects of the behavior. For example, in one of our current exercise programs, we are specifically targeting levels of moderate intensity aerobic activity such as brisk walking. We are not trying to improve strength, flexibility, or promote vigorous intensity exercise although incidental improvements in these will certainly not be discouraged. In this same program, specifications of ultimate and intermediate goals (#2 from above) would include the ultimate goal of getting users to reach and maintain recommended levels of moderate or greater intensity exercise of 30 min per day most days of the week (or 150 min per week) [21]. Intermediate goals are often different for different users depending on their characteristics. For example, intermediate goals may include increasing the level of motivation for users with low motivation, developing specific plans for those users who are motivated but do not have a viable exercise plan, increasing total minutes of exercise for those who are getting less than 150 min per week in small but steady increments, and preventing relapse for those who achieve this level of exercise. The third consideration, the order of goals, is especially important in programs that target more than one health behavior, but can also be important when intermediate goals do not have an inherent order. The last aspect of the objectives to be defined (#4) is to specify how they are going to be measured. We assess the targeted behavior throughout the program for two reasons: First, to provide data to individualize intervention messages as they are being delivered, and, second, for program evaluation. Assessment procedures for these two purposes can be independent or overlapping. There are a number of study design issues to be considered, but this topic is beyond the scope of this paper.

3.2. Population characteristics

The second important parameter is understanding the characteristics of the population that will use the program. This knowledge is necessary so that the dialogue will be engaging to all or nearly all users. Furthermore, as an integral aspect of these systems is to individualize or tailor the intervention to each user, knowledge of the population can assist in identifying the critical individual and group differences to take into account in dialogue construction. A non-exhaustive set of important characteristics to consider in dialogue writing include education level, gender, age, culture or ethnicity, income, and geographic characteristics of the home neighborhood. Since the dialogue is designed for the whole population, understanding the range of values on a characteristic is more important than knowing the mean value for the population. For example, knowing the prevalence of low educational levels in a target population for an automated health behavior program is more helpful than understanding what the average level is. Interventions adjusted for low education level (or low health literacy) are generally effective even for those with higher levels of education. That said, decisions also might be made *not* to adapt the dialogue to the extremes of a dimension. If 98% of the population has at least a sixth grade education, it might not make sense to adapt the intervention to the 2% of the population that has lower education levels. In this case, a decision would need to be made whether or not to exclude individuals with this level of education.

A new area of research for our group is investigating interventions tailored to specific racial or ethnic groups [3]. The ethnic identity of the user provides the primary context in which the lifestyle behaviors occur. For instance, one's culture has a large effect on dietary preferences and the role of food in self-concept and social interactions. In our work on culture, we use Resnicow and colleagues' [22] framework, which emphasizes understanding both 'surface' aspects of culture (e.g., food preferences) as well as 'deep' aspects (e.g., role of food in maintaining interpersonal relationships).

In addition to accessing data on distributions across variables, we have found that two other sources provide invaluable information on populations we serve. The first is direct experience, usually in clinical settings, with members of the target population. If this experience does not exist with the design team, it is advisable to find outside consultants to review program design and the dialogue writing approach. The second source of information is from focus groups of individuals from the target population. In an ongoing project to improve self-care for hypertensive African-Americans, we ran a series of focus groups on self-care behaviors, relationships with care givers and the experience of hypertension. We listened carefully not

only to what was said but to how it was said. Idiosyncratic terminology used by a sub-group can be used to adapt the intervention scripts. Once an in depth understanding of the population is achieved, the writing of the dialogue will be better informed producing more effective and engaging interactions.

3.3. Intervention approach: using theory to guide the development of a dialogue system

The third set of decisions to be made concerns the intervention approach that will be used to change the behavior. The chosen approach should be explicitly stated and used as a guide to dialogue writing. We believe the most important of these intervention approach decisions is selecting a behavioral change theory to guide the construction of the dialogue system.

Some of the most effective health behavior change programs use theory as a framework for designing and implementing the intervention. Theory is defined as a set of interrelated concepts, definitions, and propositions that present a systematic view of events or situations by specifying relations among variables to explain and predict these events and situations [23,24]. In health behavior research, theory is essentially a framework that determines the approach to and the components of the intervention. There are numerous theories available to researchers building health behavior interventions, e.g., Social Cognitive Theory [25,26], Health Belief Model [27], Theory of Reasoned Action [28], Theory of Regulation and Self-control [29], Theory of Subjective Culture and Interpersonal Relations [27–31], Transtheoretical Model of Behavior Change [32–34], Protection Motivation [35], and Precaution Adoption Process [36]. Our laboratory has structured the majority of our health behavior programs using established behavioral theories including the Transtheoretical Model and Social Cognitive Theory (SCT). The purpose of this paper is not to propose the use of any one specific theory to guide the dialogue system; our purpose is to recommend that dialogue systems for health behavior change be based on a theory. We believe this is important for two reasons. First and most immediately, it is likely to lead to more useful programs because good theory leads to more comprehensive and consistent intervention designs and provides effective strategies for behavior change. The second reason is that evaluating theory-based programs provides important data to the field on how to improve theories of health behavior change which ultimately lead to more effective interventions.

In choosing a theoretical basis for a specific dialogue system, the system developer should evaluate candidate theories on a few main criteria that we have found to be important. The first is whether the theory has any empirical evidence related to improving the target health outcomes. The second is to determine if the theory has a framework that it is easily translatable into a dialogue conversation. Since TLC conversations are aimed at the individual user, it is important that the theory is able to explain an individual's behavior as opposed to explaining population or group behavior. Third, the theory should provide clear guidance on the timing and content of the intervention, i.e., what to say and when to say it, so computer algorithms to control the conversation can be specified.

One theory that we have used many times in designing our TLC automated dialogue systems is the Transtheoretical Model of Behavior Change (TTM). The TTM integrates a set of constructs related to how individuals change a health-related behavior; a detailed description of the model can be found in an article by Prochaska and Velicer [37]. From our perspective, the TTM meets all three criteria delineated above. First, there is extensive empirical evidence on the effectiveness of the TTM for health behavior change [32,33]. The precepts of the TTM model have been used to structure a variety of health behavior change intervention programs, and these programs have been shown in well designed evaluation studies to change health behavior in the ways predicted by the model. Second, as the theory was developed to specifically explain the process of behavior change, its conceptual structure facilitates its translation from a theory to an intervention. The theory consists of four interrelated constructs, each of which is easily translated into dialogue. The third quality that makes the TTM an appropriate basis for an automated intervention is that it provides guidance on what content is to be delivered and in what situations; thus, computer algorithms can be easily rendered.

3.4. An example of how to use theory to guide the development of a dialogue system

The following section uses information on one of our physical activity programs (TLC-PA), developed using the TTM as a theoretical framework, as an example of what a dialogue system might include. Based on current health recommendations, a typical goal of our physical activity programs is to assist users to achieve and maintain 150 or more minutes per week of moderate-intensity physical activity [21]. Four TTM constructs were used to design TLC-PA, namely: stage of change, processes of behavior change, decisional balance, and self-efficacy. The first of these represents an individual's progression through a change process and provides an organizing structure for the integration of the other three constructs; therefore, we design TLC-PA to provide separate dialogues for each of the five stages of change: Precontemplation, Contemplation, Preparation, Action, and Maintenance [38]. The user's stage is assessed at the beginning of each call through a series of three to five branched questions. If they have changed stage since the previous call, they are given appropriate feedback on this movement in stage. The caller will then hear content based on the other three theoretical constructs but tailored both by their current stage of change and what content they have progressed through previously. For example, a person in the Contemplation Stage would

engage in a conversation to clarify their decision to exercise or not (the decisional balance construct) and one part of this would be about the many benefits of doing regular moderate physical activity for example: "*O.K., I will read a list of several benefits to you. You can pick the topic most important to you and I will then tell you how exercise can help.*" The processes of change are a set of overt (actions) or covert (thinking/feeling) activities that foster positive behavior change and are employed in all stage dialogues. The process that is most beneficial at any one time is largely determined by which stage the user is in. For example, the process of consciousness raising might be used in the Preparation Stage by encouraging a user to learn more about the types of physical activity they might consider. The conversation for this user might then move on to building the user's confidence that they can successfully engage in regular exercise (the construct of self-efficacy). The introduction to this section includes: "*As you move to becoming and staying regularly active, there may be times when you have doubts whether you can make these changes. If you would like to hear how to feel more confident about being active, say 'confident,' if you would like to skip ahead, say 'skip' now.*" Finally, this conversation would end with TLC-PA negotiating a stage-specific goal for the next week. For instance, for those in Preparation who have begun to increase their level of physical activity, the system has a set of algorithms that considers the overall physical activity goal of the program, the person's previously stated intermediate goals, the person's previous levels of physical activity and heuristics that consider appropriate changes in goals over time. For example, after asking the caller about their goal, TLC might say, "*Exercise experts and physicians agree that people can and should slowly increase their amount of exercise activity; even if it is only by 5 minutes each day. Previously, you decided to exercise on* [*previous planned days*][1] *days for* [*previous planned minutes*][1] *minutes each day. I think it's reasonable for you to try to do 5 more minutes of exercise on each day that you exercise this coming week. Would you like to re-consider your activity plans for the coming week? Say 'Yes' or 'No' now.*" Alternatively, if the user tries to set an intermediate goal that is much higher than their current exercise level, the TLC system will encourage them to adjust their next goal downward to a level where success is more likely. This strategy process is in keeping with the theory and empirical findings on self-efficacy, namely, that it is increased by success on intermediate goals. In summary, the TTM was used throughout the design of this intervention, its use made success more likely, and studies using this system are a good test of the usefulness of this theory in the design of effective interventions.

[1] [previous planned days] and [previous planned minutes] represent variables that contain the number of days and minutes the users chose as their plan during the previous TLC call one week earlier.

4. System specifications

One of the next steps in the process of designing an automated TLC program is determining the systems specifications. Specifications include the number of contacts, the duration of the program, the schedule of contacts, and the duration of individual TLC conversations. These attributes are specific to each application and are determined by the objectives, population characteristics, and the demands of the intervention approach but are also tempered by practical considerations. The total contact time designed into a program is determined by these parameters with each dimension being independently important. TLC programs have varied greatly on all three of these dimensions.

4.1. Number of user–system contacts

We have developed TLC systems for a range of contacts, from as few as a single contact to systems that can handle an indefinite number of contacts. An example of a system designed for one contact is our TLC program to promote screening mammography in women. This system contacts the woman a month before her annual mammogram is due. The one-call design fits the behavioral objective (to have the woman take a single action) and is sufficient to present the content hypothesized to bring about this effect. The intervention both assists the user directly in the scheduling of the screening and addresses whatever barriers to getting a mammogram that the woman might have. The system allows the user to choose from a set of 22 barriers to hear advice that is applicable to her. The system also allows the women to call back if they were interested in hearing more barriers or wanted information repeated. In a recent study, we found that few women utilized this call back option which was indirect support for the 1-contact design [11]. Also, the mean call duration was $11\frac{1}{2}$ min and the range of durations was from $3\frac{1}{2}$ to 27 min indicating that women chose to hear very different amounts of content.

Given that many health behaviors (e.g., eating, being physically active/inactive, drinking alcohol, smoking cigarettes, etc.) are inherently complex and often resistant to change, they require much more intensive programs than the mammography program just described. The majority of our TLC systems have used multiple contacts over time. In fact, we have designed systems with the potential to provide on-going contacts for an indefinite amount of time in order to foster life-time maintenance of health behaviors.

4.2. Duration of the program

We have found that the duration of a behavioral program should be long enough for users to receive the full content of the program at a reasonable rate for them, have time to practice and integrate new skills or perspectives into their lifestyle both between and across contacts, have

the opportunity to repeat difficult sections of the program, and achieve success and maintain it for at least a short period of time. The majority of our programs to date have been designed to be delivered over a 6 month period which has resulted in positive health outcomes [16]. One of our TLC physical activity programs was delivered over the course of a year. It had positive effects after 6 months but there were no further increases between 6 and 12 months [16], suggesting that 6 months of TLC was sufficient at least for initiation of behavior change.

Most of our work to date has been in the initiation of behavior change, whereas maintenance of behavior change might call for different program designs, such as those with durations longer than a year or those which are provided on an unlimited basis. As mentioned earlier, one of the outstanding advantages of automated systems is that once developed, the incremental cost of use is low. Unfortunately, few data exist to inform design decisions in this area given that few studies have systematically investigated the use of automated systems to promote the long term maintenance of health behavior change. We are presently completing two TLC systems that expressly target the long-term maintenance of dietary improvement, a very difficult problem. These two systems are based on completely different theories of behavior change, and they will be compared both to each other and to a control group. Although dialogue systems allow for program designs of indefinite length, more research is needed to determine the optimal duration for the various objectives of these systems.

4.3. Schedule of contacts

Not only does the duration of a program vary, but the schedule of TLC contacts within this time period varies as well. Although the most common frequency of contacts is one call per week, other options are possible. One clear advantage of regular (daily or weekly) calls is that it is easier for a user to get into the habit of having a call at that time, improving call adherence. We are currently involved in the design of a system that will use very short daily calls to just gather information on two behaviors (physical activity and glucose testing), and then this information will be combined and used in the more usual weekly behavior change conversations. Other TLC programs start off with weekly calls, and then after the main content has been completed, reduce frequency to bi-weekly and then monthly calls that become more of a check-in and review than presentation of new content. This variation in call frequency has led to 6 month programs with a low of nine contacts (weekly for 1 month, followed by monthly for 5 months) to those with 26 weekly calls. Again, these decisions are made based on the amount of content the program needs to cover and our understanding of what might be optimal given the program objectives and population of users. However, we have not directly compared outcomes of programs with different durations or contact frequency.

4.4. Duration of TLC conversations

The duration of each TLC conversation varies across the programs, across different calls within a program, and among the users of a particular program. The average duration of TLC calls was as short as 4 min for the original TLC-Hypertension system [8] and as long as 25 min for an intervention to increase physical activity. In general, we have found that calls of 10–15 min tend to be optimal for most users. Given that it usually takes a couple of minutes to initiate the call and end the call with a short review and confirmation of the next contact time, there is between 8 and 13 min for the main content of the intervention per call. In some systems, the decision whether or not to go on to a new section is made during the call. If the elapsed time is past a limit, say 10 min, the system might decide to not start a new section or, alternatively, ask the user if they want to go on or start the section in the next call. This takes some extra programming but can be a nice way to give the user more control over the conversation.

We have limited the duration of individual TLC calls for a number of reasons. First, we wished to minimize the user's perceived burden as it is likely to be associated with attentiveness during the calls, rates of call completion, and adherence to the prescribed use schedule. Second, we wished not to overload the user, but rather maximize the likelihood that the user comprehends and retains the information and counseling delivered. Often when adapting a non-automated intervention to a dialogue system, the content of what might be delivered in an ideal assessment and intervention contact will have to be divided into the smallest viable units and delivered across multiple TLC conversations. For example, the total amount of time to deliver one complete dietary assessment and intervention in our TLC "healthy eating" dietary program [1] would be about 1 h. We divided the material into six conversations of about 10 min each, covering all of the content. These six conversations were called a cycle. When one cycle was completed, there was a review call, and then a new cycle of calls would begin.

4.5. Other important system specifications

There are other design decisions that although more limited in scope, can have large effects on intervention effectiveness. These specifications include how the user-TLC conversation is initiated, and the automated options related to user–system interactions and response options.

4.5.1. Initiation of the TLC conversation

Automated telephony technology permits both inbound and outbound calls, or in other words, either the user can call the system (inbound), or the system can call the user (outbound). From a programming perspective, the inbound contact systems are much simpler to design and operate. Users initiate the contacts by dialing a telephone number and entering a password that identifies them to

the system, which then confirms their name and accesses stored data. All of our early systems were inbound systems. Outbound systems have to account for the increasingly complex set of possibilities of whom or what answers the phone. The details of this are beyond the scope of this paper, but a difficult challenge is designing the systems so it reliably can differentiate between a live person and an answering machine or voice mail system answering the call. In addition, protocols have to be specified to handle a number of situations including scheduling call backs if the line is busy, not answered, if the user is not home or home but not available to take the call. TLC, for example, will leave a short message on an answering machine, but although it might call back repeatedly, it will only leave one message per day and no more than a few per week. TLC also produces daily reports on calls completed, calls partially completed, and the results of other call attempts to be reviewed by project staff. Other protocols specify when a staff person is to contact the user by phone or mail to trouble shoot connection problems.

Beyond technological and practical issues, there are pros and cons of each method of initiating a TLC conversation. An inbound system gives the user more control over the initiation of an interaction; however, we have found that inbound systems result in fewer than prescribed contacts. Outbound systems result in better adherence to the planned contact schedule [39]. Our experience tells us that outbound systems are superior. Additionally, they have the capability to accept inbound calls from the user which preserves some user control. We have also given users the option to switch from outbound to inbound calls.

4.5.2. Automated interactions and response options

A design choice that has become available recently is whether to use automated speech recognition (ASR) technology. In ASR systems, the computer is "listening" for a finite set of responses after each question. Users are provided with this set of responses and can speak their answer into the telephone. Although many times, a user may reply with an answer that is different than the set of responses provided. For example, after a Yes/No question, the system says, "*Please say 'Yes' or 'No'*," but the user might answer "*yep*" or "*yeah*." ASR technology can be programmed to accept common synonyms (e.g., yea, yup, nah, nope, etc.). If the system does not understand a reply, there is a protocol for querying the user. ASR can also understand multiple word answers. In one of our TLC diet programs, we provide users with a set of five benefits of a healthy diet and ask the user to indicate their choice, e.g., the system says, "*If you are interested in hearing about how diet can improve your skin, say 'skin.' If you are interested in how a healthy diet can lower your blood pressure, say 'blood pressure'*." If no answer has been received by the end of the list, the system will say, "*Please say skin, blood pressure, …, or none now*." The ASR technology we use allows the caller to 'barge in,' meaning the system is listening for the user's answer while the question is being asked or before the set of possible responses are being listed.

There are definite advantages and disadvantage of using ASR technology. One advantage of ASR is that the call functions more like a human conversation. Given that many phones have keypads on the handset, ASR also eliminates the need to move the phone away from the user's ear. The disadvantages of ASR are that it has some problems in recognizing certain spoken responses such as large numbers (e.g., blood pressure readings). We have found that there is a small subset of users who get frustrated enough with the system not understanding their answers that they stop using the system. Another small subset of users has trouble keeping the system's technological limitations in mind—a phenomenon that might be exacerbated with the use of ASR systems that function more like a human conversation by "listening" and "understanding" responses. This more human-like conversation may increase the user's expectations of the system and frustration with its limitations. Although our recent systems have incorporated ASR technology, we are uncertain whether ASR represents a clear improvement over touch-tone technology given its current limitations.

5. Steps to writing a comprehensive dialogue

We have described the background and parameters of how to structure a dialogue system, as well as important specific design considerations. The next section will describe how we approach the writing of a TLC dialogue system as well as the overall qualities we feel are important for a high quality script. The script for the dialogue is generally written by either an expert in the content area or an expert assisted by a professional writer. The author (script writer) needs to have a good understanding of the design specifications (e.g., the goals of the system, the theory to be applied, duration and schedule of contacts) that have been discussed in the previous sections of this paper.

5.1. Characteristics of the optimal dialogue systems

It is important that the user experiences the dialogue system as an engaging and valuable conversation. This enhances the user's attentiveness and increases the likelihood that the person will respond positively to the behavioral counseling provided. It also contributes to the user's adherence to the recommended call schedule, which is important for programs that entail multiple contacts over time. Creating natural sounding and engaging conversations using automated systems, given the present technological limitations (e.g., unable to comprehend free speech), takes a concerted effort.

Based on our extensive experience in writing and evaluating automated dialogue intervention systems, we believe effective systems are based on dialogue that has several important qualities including that the dialogue: (1) is optimized for spoken communication with lay people, (2)

endows the systems with human-like characteristics to resemble real conversations, (3) is personalized to the individual user, (4) maximizes interactivity, (5) balances repetition and novelty of content, (6) mixes system and user control of the conversation, and (7) maximizes pseudo-intelligence.

Writing messages that are suited for communications that are *heard* by a lay population is often a surprisingly difficult task for health professionals without previous experience in dialogue construction. Most health professionals, especially those who are researchers and academics, principally write research grant proposals and journal articles that are addressed to a professional audience who will read rather than hear what is written. It is a twofold adjustment to both write for a much lower educational level and to write conversation instead of text to be read. What often works best is to pair a content expert with a professional writer who is known to be good at writing this type of dialogue. We also recommend that writers read out loud the dialogue they have written, even if only to themselves. One important point to remember is that only a limited amount of material can be kept in mind by the listener at any one time. Complex lines of reasoning or long lists might work in a text based intervention, but are generally not appropriate for spoken communications.

Our research suggests that our systems are quite successful in emulating real conversations as users clearly anthropomorphize the TLC voice [40]. We deliberately invest the voice with human qualities to better engage the user. First, we use actors and actresses to record the scripts. This ensures that the voice is easy to understand and is rich with variation in inflection and tone (although computer generated voices are improving in these qualities). Second, conversation structure and content is designed to emulate the characteristics of a human conversation. For our programs which are focused on care of a medical condition, the TLC dialogues are usually emulating a health professional speaking with a patient or consumer over the telephone. We have conducted a number of observational studies in which we record the interaction between training health professionals, most often registered nurses and nurse practitioners, with patients on the phone in order to identify the characteristics of human health professional delivery over the telephone. From what we have learned, we have designed the content of the TLC conversations, including the questions and declarative statements, the order of presentation of content, how the system responds to questions and the words, sentence structure and tone used, to closely match the user's expectations of what a health professional (e.g., a nurse) might ask, respond and sound like. In-depth interview studies have demonstrated that users believe that the TLC scripts successfully emulate a human health professional [17].

For programs that are less clinical in nature, such as diet improvement programs, the script is less formal, and we consider the voice to be characterized more as a friendly but expert advisor or coach. As if one was writing the dialogue for a character in a play, it is best to have a clear sense of the role of the dialogue voice of the intervention. To have a consistent presentation it is also best to only have one or two 'authors' of the script (although it can have many designers) with one doing the final editing. In programs that target multiple behaviors and are written by separate teams of content experts and writers, we use a different voice, which we give a first name, for each behavior. This allows each of the voices to take on the personality conveyed by the individual script writer and for the user to identify it with one behavior.

Other ways we 'humanize' the automated dialogue is to include some occasional humor into the script, have the system check-in with users about their status within the conversation, (e.g., "*Ready to go on?*"), or have it occasionally be self-deprecating ("*I know I am only a computer, but* ..."). The use of humor has to be judicious as to not undercut the seriousness of the topic or the expertise of the TLC 'advisor' and is not appropriate when dealing with sensitive health topics.

Although users might anthropomorphize the voice, most understand that the system is a computer but "go along" with the fiction that they are interacting with a human health professional [10]. In a similar way that movie goers 'suspend disbelief,' users, in order to maximize their participation and enjoyment, set aside their understanding of what is going on and interact with TLC as if it were a person. In qualitative interviews, we usually hear that users 'feel' that they are interacting with a person, even though they 'know' it is a computer system.

Further anthropomorphizing of TLC is engendered by maximizing the system's pseudo-intelligence. Of course, the system is not truly intelligent; however, we wish it to function in a manner that as closely as possible emulates how an excellent clinician would function. Evaluations of TLC systems indicate that the users perceive it in that manner. However, they also identify "lapses in intelligence." For example, several users who used a TLC physical activity application complained that sometimes they were not able to exercise due to the inclement weather but the closed–ended nature of the conversation made it impossible to communicate that fact to TLC. For this reason, it is important that users of dialogue systems be educated about the limits of its "intelligence" such as its inability to understand very particular circumstances the person might be in. Likewise, it is important for the designers to identify particular instances in which a section of script does not allow an option that some subset of users might expect. This recognition can only occur if systems are thoroughly tested before release, including testing by some members of the target population, and their performance closely monitored after release. In the meantime, to reduce user frustration, in some TLC applications we allow users to leave a message for the study staff at the end of each conversation in which they can tell us what they could not tell the system.

The hallmark of advanced dialogue systems for health behavior change is the ability to have the communications

be very specific to the user. Using terminology codified by Krueter et al. [41,42], this is done in three ways: personalizing, targeting, and tailoring, each described below.

First, we use information about the user to 'personalize' the conversation. The data used for this are not necessarily relevant to the user's health status or to specific behavior change strategies, but makes it evident that the system knows who it is talking to. This includes referencing the user's name, knowledge of the user's family members, names of caregivers or the neighborhood they live in. Generally, this helps build a sense of a relationship with the user. It is worth noting that one can overuse this feature. In early TLC programs, we used the person's preferred name very frequently, such as "*as you know, John, your blood pressure of 160/100 is higher than what your doctor would like it to be.*" Overuse of the person's name can be grating.

Second we 'target' to the subgroup or subgroups the user belongs to. This might be the subgroup that the whole user population belongs to, as when we have targeted an intervention to a racial or ethnic group, or we might target to a number of subgroups within a population, such as different age groups, genders, or groups defined by the TTM stages of change, described above.

Most importantly, we 'tailor' (or 'individualize') the content of the messages based on specific information we collect about the users. This information can either be collected before the beginning of the intervention or during the TLC calls. Although it is often more efficient to collect data before the intervention either from surveys or medical records, information that is collected during the call and used immediately promotes a more open and conversational experience. It is also important to collect current information on variables that might change over time. The TLC system stores this data and it can be used over time to communicate patterns of change. Studies of TLC users have indicated very positive responses when TLC "remembers" something about the person [43]. Tailoring on variables important to either the user's clinical status or to the process of behavior change allows conversations to be relevant, efficient, theory-based, and effective.

There are, however, clearly limits to the scalability of this process when using finite state machine architecture. If one is using multiple variables to tailor messages, the number of messages that have to be written, recorded, and programmed can quickly become very large. For example, if different versions of a message are being written based on the 5 stages of change, and three levels of self-efficacy (e.g., low, medium, and high), then 15 messages would have to be written. If, subsequently, it is decided to add in whether the user's self-efficacy has increased, decreased, or stayed the same since the last call, there are now 45 messages to be created, yet, because the user will only hear one of these during any one call, this might only account for 15 s of dialogue. For this reason it is important to be clear on what are the most important variables. The theoretical basis of the system should provide a guide to this prioritizing of dimensions of tailoring.

In TLC conversations we carefully mix repetition of call flow and content with new and even unexpected material. Repetition serves several purposes. For example, since in nearly all TLC programs the users are being assessed regularly, often about the same phenomena (e.g., amount of physical activity, etc), it makes the conversation go faster and more smoothly if the assessment appears in the same part of the conversation and is done in the same way each call. We believe that such repetition of sections provide almost a comforting familiarity for some users. Novelty is also necessary and comes from introducing new topics or new approaches to the same issue and in reviewing changes in user's behaviors and/or attitudes over time.

As mentioned, TLC systems are designed as finite state machines in which the flow of the conversation, though complex is completely pre-determined. This ensures that the objectives of the program during each contact are met and that the interview is complete and appropriate and the counseling is consistent with the intervention plan. At the same time, we attempt to introduce multiple opportunities throughout the TLC conversations for the user to direct the conversation in ways that make sense to that person. This includes asking the user to choose between topics to 'talk' about during a call or to decide whether to hear more detailed information about a topic or to have content they have already heard repeated. In other situations, a decision might involve input from both the user and the system. For example, in a diet program for hypertensives that focuses on consumption across four food groups, an initial decision has to be made about which food group to start with. In this program, the system allows the user to know which food group it would recommend based on the user's present consumption. Then, the user is asked to decide what food group to begin with. A similar interaction is often used in setting weekly goals, with the system either making recommendations first or having the user lead with a new goal and the system suggesting adjustment if it seems either too low or too high to be optimal. Whenever it is reasonable we design systems so the user's choice prevails.

5.2. Flow diagrams

One common challenge in writing dialogue systems for behavioral change interventions is how to manage the complexity of the potential conversations. This can be managed to some degree when flow diagrams are used. Our laboratory uses a commercially available software package designed for creating flow diagrams. A series of flows are typically written at progressing levels of detail from general to specific. We begin with what is called a "top-line" flow which outlines the more general objectives of the overall script and the basic topics addressed in each call. The theory chosen to guide the intervention should be evident in the flow's top-line. Fig. 1 displays a top-line flow for the fruit and vegetable module of a diet intervention. Notice

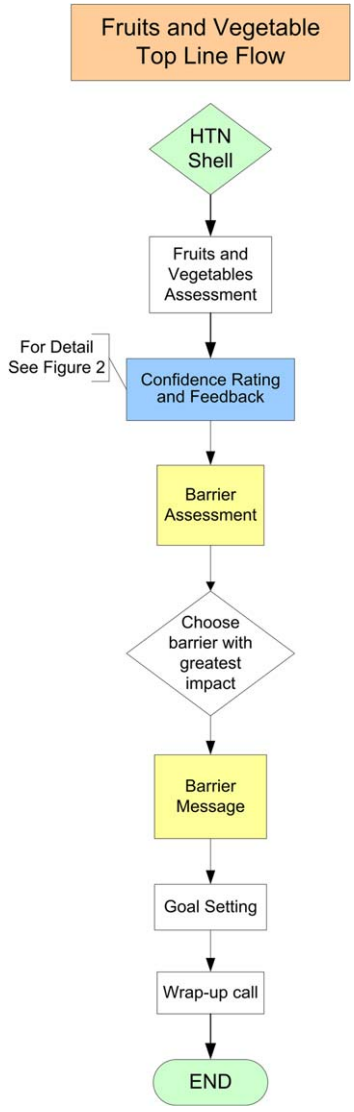


Fig. 1. Top line flow of a fruit and vegetable intervention.

that it is general and incorporates the theoretical constructs to be targeted.

As the script is outlined, the top-line flow is fleshed out with more specifics for each call, e.g., introduction, assessment, feedback, education, goal setting, and closing statement. This process of diagramming the calls continues until there is enough detail to suit the writer's needs. Fig. 2 shows the more specific flow of the call that was represented in the box labeled 'Confidence Rating and Feedback' in Fig. 1. We have found that more detailed flows can make it easier to write the script. As illustrated in Fig. 2, the flow has a branching structure to describe the various paths that a conversation can take and there is a phrase in each box conveying the message content to be conveyed to the user.

5.3. Collaborating with the technical staff

The system design team includes more than the experts in the content area, it also includes the programming staff, system testers, and often a liaison between the writers and the programmers. Our laboratory uses a liaison person to review and critique dialogue scripts, clarify logic, optimize conversational dialogue, and put the script into a format that enables it to be computer programmed e.g., includes 'go to' and 'variable' statements. Technical comment statements can be added by the authors, the liaison or the programmers to document the logic of a section of script. To this end, the flow diagram developed by the authors is used by the programmers to understand how the script flows.

There are many aspects of the script that are repetitive such as verifying what a caller has vocalized. More specifically, a typical script contains many "Yes/No questions" with the question followed by the statement "*say Yes or No now*" If yes, go to line 4, If no, go to line 5. (Note that while the early textual forms of the script use branch logic for simplicity, this is converted into finite state machine logic and runtime software implemented using state-of-the-art structured programming techniques.) The liaison can complete these dialogue components for the author. Another role of the liaison is determining whether all branching logic possibilities have been accounted for. Often there are responses to a question, which since they are atypical, are not considered by the script writer, but the control system must account for. Related to this, the caller may provide improbable responses, i.e., fall outside of the range of logically possible or acceptable answers. TLC may ask, "*Please tell me the number of days that you exercised during the past week?*" If the caller answers 9, then TLC will say, "*I am sorry, I heard 9. I was asking about the number of days you did moderate activities in the last week. The number cannot be higher than 7. Please say the number of days again. Say zero if you do not do any moderate physical activity. You may also enter the number of days using your key pad, entering a number from zero to seven.*" The liaison will add these types of responses throughout the script.

5.4. Suggestions for dealing with common dialogue situations

From our extensive experience writing dialogue for TLC systems we have developed approaches to some common situations that arise when trying to translate a behavioral change intervention into an automated dialogue system. These approaches range from simple to more complex. One situation is when it is necessary to have the user leave the phone temporarily either to get something (pen and paper) or to perform a task (weigh themselves). In these instances, we instruct the user to either press one or say "*ready*" when they return to the phone. We have learned that sometimes users forget this instruction when they return to the phone so we have the system repeat this instruction every 5 s so the user is likely to hear it when they pick up the phone. Another aspect of this situation is how long to have the line stay open because occasionally the user does not return to the phone. Since this uses some system resources (mainly a telephone line), the level of sys-

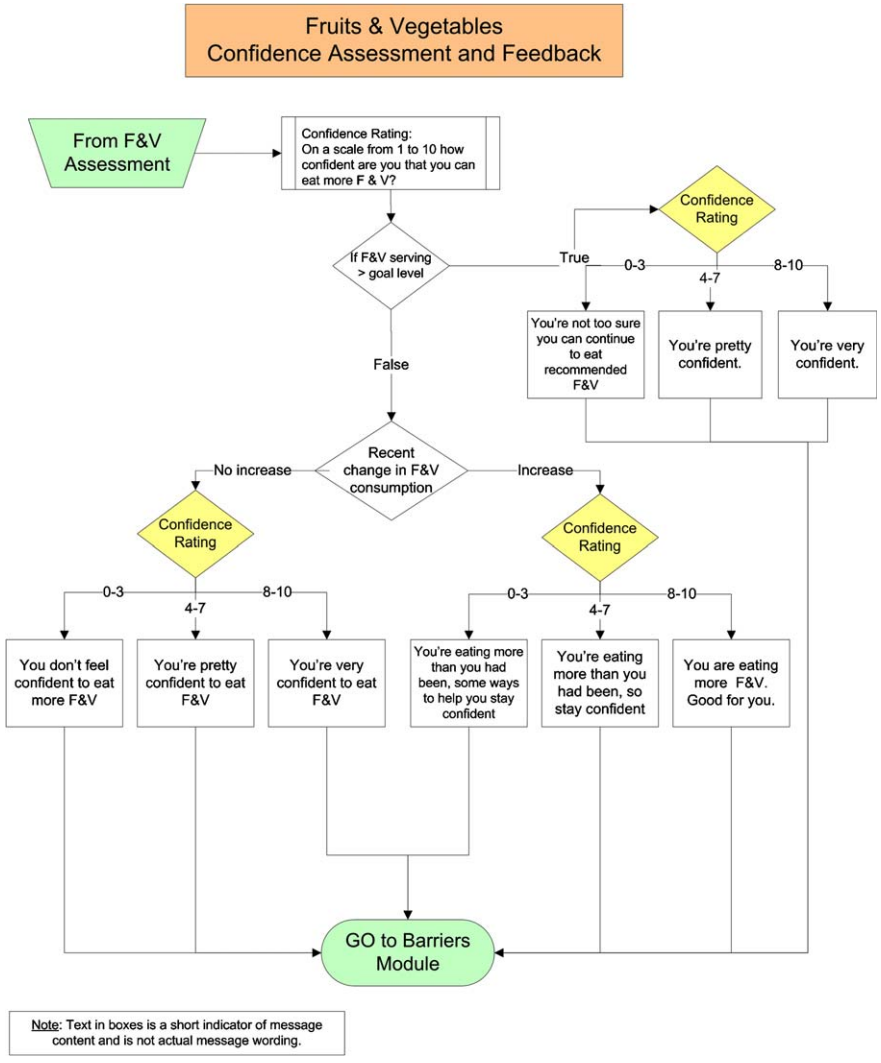


Fig. 2. Detailed flow of confidence for fruit and vegetable consumption.

tem resources somewhat determines how long the line can be left open. We usually use a 5 min timeout for simple tasks such as getting paper and pencil to write down something (e.g., a phone number, homework assignment), and have found that a longer time such as 10 min is needed if subjects have to accomplish a task (e.g., taking their blood pressure or finding something that might be misplaced, like a report we have sent them).

Another common situation is when there is extensive information to deliver to the user. We have found that lengthy monologues go against the nature of a good conversation and make the experience less engaging. The amount of information one might easily put into a paragraph in a written communication is often too long for a phone system if not broken into smaller segments. In instances in which there is extensive material to present on a single topic, we attempt to break the material into smaller segments of only a few sentences each, usually with questions between the segments. These questions might merely serve to keep connected with the users, such as "*Did you get that?*" or a rhetorical question to break up the content. Utilizing quiz type questions is another way to keep content conversational. Then you can provide detail on the topic regardless if the user answers correctly or incorrectly. For example, if the system is trying to present benefits of exercise, the system could say: "*And did you know that regular exercise will help you sleep better and think clearer? Say 'Yes' or 'No' now.*" The next message provides a bit of detail on this, but begins with: "*That's right*!" if they responded yes, and "*Well, it does*" if they responded no. Another way to increase interaction and decrease length of utterances is to present information at multiple levels of detail. For example, in a TLC health information program [7], we developed a program of over 50 health information topics derived from the Harvard health letter, a health newsletter published by Harvard University Medical School publications. To deal with the large amount of factual information, we broke the content up into sections and in each section provide a brief overview of the content in that section. We then asked if the user wished to learn more about this topic, and if they did, provided more detail before moving on to the next

sub-topic. When a lot of information has been communicated, it is important to include brief reviews of information along the way.

Asking certain types of questions can be more difficult in dialogue systems than text based systems. For individuals of average or higher literacy, a printed question can be read and responded to faster than listened to and answered. Scale instrument with many items can be too burdensome for a dialogue system, as well as being distinctly non-conversational. The other type of question that is difficult to construct is when the purpose is to pick between many different options. Since the user cannot scan the list, they are limited by how many choices they can keep in mind at any one time. We found that if there are four to five choices, they can all be presented and the user can be asked to pick the one that they want to hear about most or first. If the list is a set of topics and we want the user to hear detail on all or most of them during the program, a common approach is to present the same list during each subsequent call, minus the options already taken. When we have a question with many answers, which is often the case in these systems, a multi-step approach has to be taken, and these steps can be designed in a number of ways. For example, a common section of our scripts is to give feedback on how to overcome barriers to a behavior, where there are often 10–20 possible barriers. Optimally the user would be able to choose the barrier that they would gain the most from hearing, but this is difficult to do with this many choices. Barriers can be presented one at a time, and users are asked to state whether the barriers are important to them or not. Then only the endorsed items are repeated and they choose from the smaller set of items. This can be time consuming, and depending on the behavior, users can endorse most of the original set, inadvertently perpetuating the problem. A second option is to divide the list into subgroups and ask them to pick one barrier of the first group (or say '*none*') and then give feedback on that one. Then the system can move onto the next subgroup either during the same call or in subsequent calls. Combinations of these approaches can be designed. For instance, one could take the first approach for a subset of items so the most important barrier of the first subgroup is addressed, and then do the same in the second subgroup. Exactly which method is optimal depends on how much contact time is available both during a call and across calls for the topic, how many items are likely to be relevant to the user, and how important it is to give them feedback on only the most relevant items.

6. Conclusion

In this paper, we have endeavored to provide a road map to the essential design considerations, component tasks, and content considerations necessary to create effective, telephone-based, automated dialogue interventions. These recommendations are based on our experience developing and evaluating these systems over the past two decades. The power of these systems is likely due to the combination of three components: new and improving computer technologies, better understanding of how to assist individuals to improve health-related lifestyle behaviors, and the use of the telephone, the oldest, widely used telecommunication technology, to engage in 'human' conversations. This combination allows for a multitude of potential applications that have only started to be explored.

Interest in the use of automated dialogue systems to improve health-related outcomes both within and outside of the health care delivery system is increasing rapidly as its potential to improve both individual and public health outcomes at a relatively low cost is being recognized. As this paper was focused on the writing of dialogue, we did not present the use of telephony for disease management as those systems involve a number of other functions such as monitoring of patient health conditions and alert generation, but as the cost of health care continues to rise the use of telephony in disease management is likely to also increase.

There are other potentials that are only beginning to be explored. There are possibilities to integrate such systems with other forms of automated health interventions (e.g., web-based interventions) as well as with the increasingly computer-based health care delivery system (e.g., EMR's). An area of research we are investigating is whether different people prefer and get more benefits from interventions depending on the type of communication delivery channels used (e.g. phone, web, or print) which could inform the matching of communication channel at the individual level. Improved computer generated voices and text to speech functionality will allow a more streamlined development process—dialogue will be heard and tested virtually as it is created. Improvements in the quality of computer generated speech could reduce the cost of system development and increase system flexibility.

As mentioned, there are practical limits to the scalability of finite state machines on the number of factors that can be simultaneously considered in tailoring content. More advanced computational approaches [44] such as dialogue planning techniques have the potential to overcome these limitations. As they are still in the early stages of development, it is thus far unclear whether use of these approaches will achieve a level of effective tailoring that can be obtained with meticulously constructed dialogues delivered using finite state machine architecture. This is an important researchable question that investigators need to address.

The proven effectiveness of TLC systems notwithstanding, there are also a number of challenges ahead. There is much to understand about how best to design and implement these systems. Despite overall effectiveness, there are still many users who do not benefit from these systems or only benefit minimally. And lastly, these systems have largely been developed and tested in the context of federally funded research projects, and it is time to move to the phase of dissemination and evaluation of these low cost interventions in real world situations.

References

[1] 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 2001;15(4):215–24.

[2] Glanz K, Shigaki D, Farzanfar R, Pinto B, Kaplan B, Friedman R. Participant reactions to a computerized telephone system for nutrition and exercise counseling. Patient Educ Couns 2003;49(2):157–63.

[3] Migneault JP, Friedman RH, Resnicow K, Campbell MK, Dunbar-Jacob J, Hill M, et al. Design of an automated multi-behavioral intervention for hypertensive African-Americans. Ann Behav Med 2003:S161.

[4] Jarvis KL, Friedman RH, Heeren T, Cullinane PM. Older women and physical activity: using the telephone to walk. Womens Health Issues 1997;7(1):24–9.

[5] 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(2):113–20.

[6] King A, Friedman R, Marcus B, Napolitano M, Castro C, Forsyth L. Increasing regular physical activity via humans or automated technology: the CHAT trial. Ann Behav Med 2003(Suppl.):S111.

[7] Ramelson H, Bassey B, Friedman R. The use of computer telephony to provide interactive health information. In: American Medical Informatics Association; 2003. p. 539–43.

[8] Friedman RH, Kazis LE, Jette A, Smith MB, Stollerman J, Torgerson J, et al. A telecommunications system for monitoring and counseling patients with hypertension: impact on medication adherence and blood pressure control. Am J Hypertens 1996;9(4 Pt.):285–92.

[9] Young M, Sparrow D, Gottlieb D, Selim A, Friedman R. A telephone-linked computer system for COPD care. Chest 2001;119(5):1565–75.

[10] Farzanfar R, Finkelstein J, Friedman R. Testing the usability of two automated home-based patient-management systems. J Med Syst 2004;28:143–53.

[11] Migneault JP, Rakowski W, Friedman RH, Marks L, Clark M. User engagement with an automated telephone intervention to improve on-time repeat mammography. Ann Behav Med 2005;D33(Suppl.):S85.

[12] Morris M, Sparrow D, Gottlieb D, Selim A. A telephone linked computer system for COPD care. Chest 2001.

[13] Young DR, Gittelsohn J, Charleston J, Felix-Aaron K, Appel LJ. Motivations for exercise and weight loss among African-American women: focus group results and their contribution towards program development. Ethn Health 2001;6(3-4):227–45.

[14] Adams W, Fuhlbrigge A, Miller C, Panek C, Gi Y, Loane K, et al. TLC-asthma: an integrated information system for patient-centered monitoring, case management, and point-of-care decision support. Proc AMIA Symp 2003:1–5.

[15] Cohen J. Statistical power for the behavioral sciences. second ed. Hilldale, NJ: Erlbaum; 1988.

[16] King A, Friedman R, Marcus B, Napolitano M, Castro C, Forsyth L. Increasing regular physical activity via humans or automated technology: 12-month results of the CHAT trial. Ann Behav Med 2004;27(Suppl.):S044.

[17] Kaplan B, Farzanfar R, Friedman R. Personal relationship with an intelligent interactive telephone health behavior advisor system: a multi-method study using surveys and ethnographic interviews. Int J Med Inform 2003;71:33–41.

[18] Farzanfar R, Frishkopf S, Friedman R, Ludena K, Ward A. Humanizing the machine: refinement of an automated patient-management system. Ann Behav Med 2004;27(Suppl.):S154.

[19] Paperny DM. Computerized health assessment and education for adolescent HIV and STD prevention in health care settings and schools. Health Educ Behav 1997;24(1):54–70.

[20] Mears M, Coonrod DV, Bay RC, Mills TE, Watkins MC. Routine history as compared to audio computer-assisted self-interview for prenatal care history taking. J Reprod Med 2005;50(9):701–6.

[21] Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA 1995;273(5):402–7.

[22] Resnicow K, Braithwaite R, Kuo J. Interpersonal intervention for African American adolescents. In: Wilson D, Rodrigue J, Taylor W, editors. Adolescent health promotion in minority populations. Washington, DC: American Psychological Association; 1997. p. 201–28.

[23] Hardy ME. Perspectives on science. In: Role theory: perspectives for health professionals. second ed. Norwalk, CT: Appleton & Lange; 1988.

[24] Kerlinger FN, Lee HB. Foundations of behavioral research. fourth ed. Orlando, Fl: Harcourt College Publishers; 2000.

[25] Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.

[26] Bandura A. Self-efficacy: the exercise of control. New York: Newman; 1997.

[27] Becker MH. The health belief model and personal health behavior. Health Educ Monogr 1974;2:324–508.

[28] Fishbein M, Ajzen I. Belief, attitude, intention and behavior: an introduction to theory and research. Boston, MA: Addison-Wesley; 1975.

[29] Kanfer FH. Self-regulation and behavior. In: Neuringir C, Michael JL, editors. Behavior modification in clinical psychology. New York: Appleton-Century Crofts; 1970.

[30] Triandis HC. The analysis of subjective culture. New York: Wiley; 1972.

[31] Bandura A. Self-efficacy: toward a unifying theory of behavior change. Psychol Rev 1977;84:191–215.

[32] Prochaska J, Velicer W, Fava J, Rossi J, Tsoh J. Evaluating population-based recruitment approach and a stage-based expert system intervention for smoking cessation. Addict Behav 2001;26:583–602.

[33] Prochaska J, Velicer W, Rossi J, Redding C, Greene G, Rossi S, et al. Multiple risk expert systems interventions: impact of simultaneous stage-matched expert system interventions for smoking, high-fat diet, and sun exposure in a population of parents. Health Psychol 2004;23(5):503–16.

[34] Velicer W, Prochaska J, Fava J, Laforge R, Rossi J. Interactive versus noninteractive interventions and dose-response relationships for stage-matched smoking cessation programs in a managed care setting. Health Psychol 1999;18(1):21–8.

[35] Rogers R, Prentice-Dunn S. Protection motivation theory. In: Determinants of health behavior: personal and social. In: Gochman D, editor. Handbook of health behavior research, vol. 1. New York: Plenum Press; 1997. p. 113–32.

[36] Weinstein ND. The precaution adoption process. Health Psychology 1988;7:355–86.

[37] Prochaska J, Velicer W. The transtheoretical model of behavior change. American Journal of Health Promotion 1997;12:38–48.

[38] Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. Psychotherapy theory, research, and practice 1982;19:276–88.

[39] Wright J. Utilization of automated telephone systems for diet and physical activity: a cross study analysis. Annals of Behavioral Medicine 2005;D33(Suppl.):S84.

[40] Kaplan B, Farzanfar R, Friedman R. Ethnographic interviews to elicit patients' reactions to an intelligent interactive telephone health behavior advisor system. In: Lorenzi NM, editor. American Medical Informatics Association, Annual Symposium, 1999; Bethesda, MD; 1999. p. 555–9.

[41] Kreuter M, Lezin N, Yung L. Evaluating community-based collaborative mechanisms: implications for practitioners. Health Promotion Practice 2000;1:49–63.

[42] Kreuter M, Farrell D, Olevitch L, Brennan L. Tailoring health messages: customizing communication with computer technology. Mahwah, NJ: Lawrence Erlbaum; 2000.
[43] Farzanfar R, Frishkopf S, Migneault J, Friedman R. Telephone-Linked Care for Physical Activity (TLC-PA): a qualitative evaluation of the use patterns of an information technology program for patients. J Biomed Inform 2005;38:220–8.
[44] Bickmore T, Giorgino T. Methodological Review: Health Dialog Systems for Patients and Consumers. J Biomed Inform 2006. in this issue.

Barriers to Treatment of Hepatitis C in HIV/HCV-Coinfected Adults with Alcohol Problems

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

Background: Alcohol use and human immune deficiency virus (HIV) infection are both associated with accelerated progression of hepatitis C virus (HCV) disease and reduced response rates to interferon therapy. In this study, we assessed the prevalence of barriers to interferon treatment in a population of HIV/HCV-coinfected patients with current or past alcohol problems and the extent to which they received treatment to address the barriers.

Methods: This is a cross-sectional, descriptive analysis of baseline data from a prospective study assessing the impact of HCV and alcohol use on HIV disease progression. Using consensus guidelines, subjects were categorized as having absolute, relative, or no contraindications to interferon therapy for HCV. Absolute contraindications to treatment included heavy alcohol use, decompensated liver disease, CD4 cell count <100 cells/μL, recent needle sharing, and suicidal ideation. Relative contraindications included moderate alcohol use, recent injection drug use, depressive symptoms, and CD4 cell count from 100 to 199 cells/μL.

Results: Of 401 HIV-infected subjects, 200 were HCV RNA-positive. Fifty-three percent had an absolute contraindication to interferon therapy, 35% a relative but no absolute contraindication, and only 12% had no contraindication. Of those with an absolute contraindication, 61% reported heavy drinking and the majority (88%) had multiple contraindications. These contraindications were present despite the fact that over 50% were in receipt of substance abuse and mental health treatment.

Conclusions: Continued alcohol and drug use as well as depressive symptoms are the major barriers to interferon therapy in HCV/HIV-coinfected subjects and these barriers persist despite high treatment rates for these problems. Therefore, more intensive treatments of alcohol, drug, and mental health issues are needed to improve HCV treatment eligibility in HCV/HIV-coinfected persons.

Key Words: Alcohol, Hepatitis C, Interferon, HIV.

APPROXIMATELY 30% OF human immune deficiency virus (HIV)-infected persons are coinfected with hepatisis C virus (HCV), and a substantial proportion also have alcohol problems (Campbell et al., 2006; Galvan et al., 2002; Samet et al., 2004; Sherman et al., 2002). Both HIV coinfection and alcohol use are associated with more rapidly progressive liver disease and reduced response to interferon-based treatment (Bhattacharya and Shuhart, 2003; Di Martino et al., 2001; Poynard et al., 1997; Torriani et al., 2004). Impaired response to interferon in those who use alcohol limits opportunities to control liver-related morbidity and mortality (Blackard et al., 2004; Carrat et al., 2004; Edlin, 2002; Mochida et al., 1996; Ohnishi et al., 1996; Okazaki et al., 1994; Rockstroh et al., 2005; Torriani et al., 2004). Furthermore, only about 30% of HIV-coinfected populations are eligible for interferon therapy because of ongoing drug and alcohol use, psychiatric disease, poor adherence to medical therapy, or underlying medical contraindications (Adeyemi et al., 2004; Falck-Ytter et al., 2002; Fleming et al., 2003). However, the barriers to interferon treatment, including the associated psychological and medical issues in individuals with alcohol problems, have not been well delineated.

Strategies aimed at improving the effectiveness of interferon-based treatment will require expansion of the population eligible for treatment and maximizing

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

Received for publication March 31, 2006; accepted April 26, 2006.

This work was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) of the NIH: R01-AA13766; R01-AA11785; R01-AA10870.

Reprint requests: David Nunes, MD, Section of Gastroenterology, Boston Medical Center, 85E Concord Street, 7th Floor, Boston, MA 02118; Fax: 617-638-6529; E-mail: david.nunes@bmc.org

DOI: 10.1111/j.1530-0277.2006.00183.x

treatment responses in those who receive therapy. Aggressive management of drug dependence and psychiatric disease has been shown to increase the proportion of individuals eligible for interferon therapy (Backmund et al., 2001; Sylvestre, 2002; Taylor, 2005) but there have been no studies assessing the impact of treatment of alcohol problems on treatment eligibility. A better understanding of barriers in people with alcohol and other drug problems and associated mental health disorders may help to target interventions aimed at increasing receipt of pharmacotherapy for hepatitis C. The aim of this study was to assess the barriers to receipt of interferon therapy for HCV infection in an HIV-infected population with alcohol problems, many of whom have access to substance use and mental health treatment.

MATERIALS AND METHODS

Study Design

This is a cross-sectional analysis of baseline data from a prospective observational cohort study [HIV-Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE)]. The study enrolled 401 HIV-infected patients with current or past alcohol problems. The major recruitment sites were the Diagnostic Evaluation Unit at Boston Medical Center (BMC; Samet et al., 1995) and the HIV Primary Care and Consultation Clinics at Beth Israel Deaconess Medical Center (BIDMC). Other subjects were recruited from BMC's primary care practices, referrals by other study subjects, and flyers posted at homeless shelters and HIV/AIDS social service agencies in the Boston area. Persons responding to the flyers were assessed by telephone for alcohol problems using the CAGE questionnaire and then invited for an in-person interview to complete the screening process. Recruitment began in August 2001 and ended in July 2003.

Eligibility criteria for the HIV-LIVE cohort included the following: (1) documented HIV antibody test by ELISA and confirmed by Western blot (medical record or tested at enrollment); (2) two or more positive responses to the CAGE questionnaire (Buchsbaum et al., 1991; Mayfield et al., 1974; Samet et al., 2004) or physician-investigator diagnosis of lifetime alcohol abuse or dependence; (3) ability to speak English or Spanish; and (4) at least 1 contact person who knew the subject's whereabouts. Exclusion criteria were a score of <21 on the 30-item Folstein Mini-Mental State Examination (Folstein et al., 1975; Smith et al., 2006) or a trained interviewer assessment that the patient was incapable of comprehending informed consent or answering the interview questions. Eligible subjects were asked to provide written informed consent before enrollment. The Institutional Review Boards of BMC and BIDMC approved this study. HCV RNA-positive subjects from the HIV-LIVE cohort were deemed eligible for the current analysis.

Subject Assessments

A trained research associate interviewed all subjects. A standardized interview included questions on demographics, HIV risk behaviors, alcohol consumption, and ART use in the past 30 days. Current alcohol consumption (past 30 days) was assessed using a validated calendar method (Sobell et al., 1982, 1992). The Composite International Diagnostic Interview (CIDI) Alcohol Module (Robins et al., 1988) was administered to determine current (past 6 months) and lifetime diagnoses of alcohol abuse and dependence. Recent (past 6 months) drug use and current (past 12 months) diagnosis of drug dependence was assessed at enrollment using the CIDI. Laboratory data, including liver-related blood tests, CD4 cell counts, HCV antibody, HCV RNA, and HIV viral load, were obtained by review of medical records. If complete data were not available, testing was performed at the time of the study visit. Attendance, including number and type of visit, at drug and alcohol rehabilitation programs and for routine medical care, HIV care, and psychiatric care was recorded.

Definitions of Contraindications to Interferon Therapy

Barriers to interferon therapy were defined based upon current clinical practice guidelines and the recent recommendations of the International Panel on the Treatment of HCV/HIV coinfected Persons (Carlos Martin et al., 2004). Absolute contraindications to interferon therapy were defined as current (past 30 days) heavy alcohol use (>14 drinks per week on average or >4 drinks in 1 day for men, >7 drinks per week, or >3 drinks in a day for women), recent (past 6 months) injection drug use with needle sharing, CD4 cell count of less than 100 cells/μL, decompensated liver disease defined as a Child–Pugh score of ≥ 7, and suicidal ideation or attempt in the past 6 months. Relative contraindications were defined as current moderate alcohol use (less than heavy use but not abstinent), current injection drug use without needle sharing, CD4 cell count from 100 to 199 cells/μL, and significant depressive symptoms (CES-D ≥ 16; Burack et al., 1993).

Analyses

For analytic purposes, all HCV RNA-positive subjects were considered potentially eligible for interferon therapy. Only subjects with complete data for all of the contraindications to therapy were included in the analyses describing the proportion of the study group with contraindications to treatment. Descriptive statistics were used to characterize the sample. Chi-square tests were used to assess bivariate associations between subject characteristics and the presence of at least 1 absolute contraindication.

RESULTS

Demographic Data

Of 401 HIV-infected subjects with current or past alcohol problems, 236 were HCV antibody-positive, of whom 232 had HCV RNA test results; 86% (200/236) of these had detectable HCV RNA and were therefore candidates for interferon therapy. This sample ($n = 200$, Table 1) was predominantly male (73%) and of diverse racial/ethnic background: blacks 38%, whites 34%, Hispanics 22%, and others 6%. The mean age was 44 years (range 20–69). The vast majority reported having injected drugs in the past. About one-quarter were homeless and a similar proportion had been incarcerated. Almost all had been seen recently for medical care; 20% had had a liver biopsy, and 8% had received interferon therapy.

Prevalence of Contraindications to Interferon Therapy

The number of subjects with each contraindication to interferon is shown in Table 2. Depressive symptoms, a relative contraindication to therapy, had a prevalence of 70%, while any current alcohol use (37%) and any recent injection drug use (23%) were the other major barriers to interferon. HIV-related immune suppression (CD4 <200)

Table 1. Characteristics of HIV/HCV-Infected Persons with Current or Past Alcohol Problems ($N = 200$)

Characteristic	*N* (%)
Male	146 (73)
Race	
Black	76 (38)
White	68 (34)
Hispanic	43 (22)
Other	13 (6)
Employed	36 (18)
Homeless, past 6 mo[a]	57 (28)
Incarcerated, past 6 mo[b]	46 (23)
Injection drug use, ever	178 (89)
Injection drug use, current	47 (23)
CD4 cell count—median (IQR) ($n = 187$)	357 (106–546)
HIV RNA median (IQR)	804 (0–12914)
Undetectable	53 (29)
Antiretroviral therapy	118 (59)
HCV RNA (copies/mL) $< 1{,}000{,}000$	83 (41)
HCV Genotype ($n = 94$)	
Genotype 1	79 (84)
Genotypes 2 or 3	13 (14)
Genotype 4	2 (2)
Child–Pugh score—median (IQR) ($n = 175$)	5 (5–6)
Suicide attempt, ever	71 (35)
Suicide attempt, past 6 mo	4 (2)
Suicidal ideation, past 6 mo ($n = 199$)	19 (9)
Mental health treatment, past 6 mo[c]	110 (55)
Prescription for psychiatric medication, past 6 mo	104 (52)
AA attendance, past 6 mo[d]	132 (66)
Substance abuse treatment, past 6 mo[e]	101 (50)
Methadone treatment, past 6 mo	41 (20)

[a] $\geq$ 1 night on the street or in a shelter.
[b] 1 night or more in jail.
[c] Seen by a mental health professional.
[d] Attended Alcoholic Anonymous meeting at least once in the past 6 months.
[e] At least 6 weeks in a residential facility or half-way house, and/or at least 12 visits to a substance abuse or mental health professional and/or 30 days of day treatment or participation in a methadone program.
IQR, interquartile range.

Table 2. Prevalence of Contraindications to Hepatitis C Pharmacotherapy Among HIV-Coinfected Persons With Current or Past Alcohol Problems ($N = 200$)

Characteristic	Number (%, 95% CI)
Absolute contraindication	
Current heavy alcohol use[a]	59 (29, 23–36%)
Injection drug use (past 6 mo) with needle sharing	21 (10, 6–15%)
Suicidal ideation*	19 (9, 5–13%)
CD4 < 100*	18 (9, 5–13%)
Child–Pugh > 7*	14 (7, 3–11%)
Relative contraindication	
Moderate alcohol use[b]	16 (8, 4–12%)
Injection drug use (past 6 mo) with no needle sharing	26 (13, 8–18%)
Depressive symptoms (CES-D score > 16)	140 (70, 64–76%)
CD4 cell count 100–199*	24 (12, 7–16%)

[a] $>$ 14 standard drinks per week on average or $>$ 4 drinks in 1 day for men; $>$ 7 drinks per week or $>$ 3 drinks in a day for women.
[b] Any alcohol use not meeting the criteria for heavy use.
*32 subjects with missing data: 25 missing Child–Pugh score, 13 missing CD4 cell count, and 1 missing suicidal ideation.

and advanced liver disease were barriers in 21 and 7%, respectively. Thirty-five percent of the study population reported a prior suicide attempt and 9% recent suicidal ideation (Table 1). Rates of treatment for mental health and substance abuse in the previous 6 months were substantial; over 50% had received mental health treatment and/or a prescription for a psychiatric medicine. Fifty percent had received substance abuse treatment and two-thirds had attended AA meetings.

Of the 200 coinfected study participants, complete data were obtained in 168 (84%). Fifty-three percent of subjects had at least 1 absolute contraindication to interferon therapy, 35% had at least 1 relative but no absolute contraindication, and only 12% identified no contraindication (Fig. 1). Heavy drinking (61%), recent injection drug use with needle sharing (19%), advanced immunosuppression (19%), suicidal ideation (17%), and advanced liver disease (16%) were all important absolute contraindications (Fig. 1). Of note, only 19% of subjects with 1 absolute contraindication had no other contraindication to interferon therapy, 30% had at least 2 absolute contraindications, and 51% had 1 absolute contraindication with 1 or more relative contraindications.

Bivariate Associations with the Presence of an Absolute Contraindication

In bivariate analyses, race, particularly white race, not using antiretroviral therapy, failure to seek routine medical care, and less than weekly attendance at AA meetings were all associated with the presence of an absolute contraindication to interferon (Table 3). The association between homelessness and absolute contraindication was of borderline statistical significance. Age, gender, employment, incarceration, and receipt of substance abuse and mental health treatment were not associated with an absolute contraindication to HCV therapy.

DISCUSSION

Using current published guidelines for treatment eligibility, over half of HCV/HIV-coinfected subjects with current or past alcohol problems had absolute contraindications to interferon therapy. Over one-third had relative contraindications, and only 12% had no contraindication. Heavy alcohol use was the single most important contraindication to interferon therapy. However, it was frequently associated with multiple other barriers to treatment, including depressive symptoms, and recent drug injection use as well as poorly controlled HIV infection and decompensated liver disease. The observation that a high proportion of those with an absolute contraindication to therapy also had multiple other barriers to treatment has important implications for the management of these individuals and limits opportunities to increase treatment eligibility.

A second important observation of the current study was the finding that many of the substance abuse and

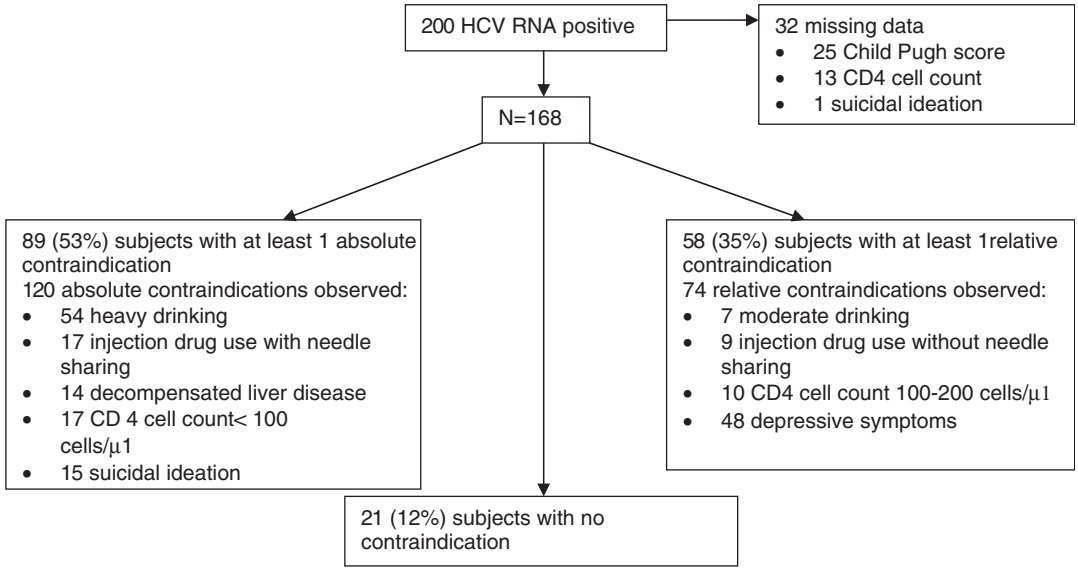


Fig. 1. Sample assessed for treatment eligibility and the reasons for treatment ineligibility.

mental health barriers to treatment were present despite a high proportion of subjects having received substance abuse and mental health treatment. Furthermore, when we assessed for factors that were predictive of treatment ineligibility, we found that white race, homelessness, nonreceipt of HAART, noncompliance with routine medical care, and attendance at few AA meetings were all associated with treatment ineligibility. These data suggest that attempts to improve interferon treatment rates in a population with alcohol problems will likely require a multidisciplinary approach that includes substance abuse, mental health, and medical care including receipt of effective HIV pharmacotherapy. Furthermore, increasing interferon treatment eligibility will require more intensive and more efficacious management of alcohol and mental health care than was available to this study population.

The findings of this study reveal an even more disappointing state of the provision of effective HCV treatment for patients with HIV coinfection and alcohol problems than were previously reported, where it was estimated that 25% to 30% of HCV-positive individuals are eligible for interferon treatment (Adeyemi et al., 2004; Falck-Ytter et al., 2002; Fleming et al., 2003; Restrepo et al., 2005). To add to this challenge, the impact of HCV treatment is further limited by poor patient acceptance due to concerns about limited treatment efficacy and adverse medication

Table 3. Characteristics of Subjects With HIV/HCV Coinfection and Current or Past Alcohol Problems Based on Eligibility for Interferon Therapy ($N = 168$[a])

Characteristic	Absolute contraindication to interferon ($n = 89$) No. (%)	Relative or no contraindication to interferon ($n = 79$) No. (%)	*p* Value
Male	70 (79%)	57 (72%)	0.37
Age (mean years)	42.9	45.0	0.63
Race			
Black	29 (33%)	36 (46%)	0.021
White	37 (41%)	19 (24%)	
Hispanic	20 (22%)	15 (19%)	
Other	3 (3%)	9 (11%)	
Employed	14 (16%)	12 (15%)	1.00
Homeless, past 6 mo	33 (37%)	18 (23%)	0.06
Incarcerated, past 6 mo	23 (26%)	19 (24%)	0.86
On HAART[b]	48 (54%)	57 (72%)	0.02
Routine Medical Visit, past 6 mo	81 (91%)	78 (99%)	0.04
AA, past 6 mo (> weekly attendance)	38 (43%)	46 (58%)	0.04
Received SAT, past 6 mo[c]	44 (49%)	38 (48%)	0.43
Received MHT, past 6 mo[d]	48 (54%)	41 (52%)	0.76

[a]168 of the 200 HIV/HCV-infected subjects had complete data on the above characteristics.
[b]HAART = On at least 3 potent anti-retroviral medications at the time of subject interview.
[c]Substance abuse treatment (SAT): At least 6 weeks in a residential facility or halfway house, and/or at least 12 visits to a substance abuse or mental health professional and/or 30 days of day treatment or participation in a methadone program.
[d]Mental health treatment (MHT): Seen by a mental health professional.

effects, particularly in coinfected patients (Blackard et al., 2004; Carrat et al., 2004; Fried et al., 2002; Hadziyannis et al., 2004; Manns et al., 2001; Rockstroh et al., 2005; Torriani et al., 2004).

There have been no prior studies evaluating HCV treatment eligibility in patients with alcohol problems despite the fact that more than 50% of people coinfected with HCV and HIV have alcohol problems (Samet et al., 2004). The importance of continued alcohol use is not only related to its adverse effect on the progression of liver disease but also to the observation that alcohol use is associated with reduced response rates to interferon, an effect that can persist even after abstinence (Edlin, 2002; Harris et al., 2001; Hezode et al., 2003; Mochida et al., 1996; Ohnishi et al., 1996; Okazaki et al., 1994; Poynard et al., 1997; Safdar and Schiff, 2004; Tabone et al., 2002). For these reasons, current guidelines recommend that abstinence be achieved before beginning interferon therapy, especially with HIV/HCV-coinfected patients who also have diminished interferon response rates and more rapidly progressive liver disease (Peters and Terrault, 2002).

Recent guidelines have included current substance abuse and psychiatric disease as relative contraindications to interferon therapy or have recommended that treatment be administered on a case-by-case basis. Early studies in the management of this population had shown that the incidence of psychiatric problems, drug relapse, and early discontinuation of therapy was higher in those with addiction (Edlin, 2002; Fleming et al., 2003; Kraus et al., 2001; Schaefer et al., 2004). It is clear, however, that individuals stable on opioid replacement therapy can receive therapy safely and effectively (Sylvestre, 2005b). For instance, Sylvestre (2005b) reported that 78% of patients in methadone maintenance treatment that included counseling completed their planned 6 to 12 months course of interferon with an end-of-treatment response rate of 54%. However, responses are lower in those with recent substance abuse and psychiatric disorders (Sylvestre, 2005a). Backmund et al. (2001) have also reported successful treatment in injection drug users using a multidisciplinary approach including specialists in HCV therapy, medical, mental health, and addiction care (Taylor, 2005). Successful treatments in highly coordinated programs like these are a model for larger centers with the necessary resources. Comparable studies evaluating the effectiveness of alcohol treatment programs on interferon treatment eligibility and outcomes are not available. The observation in this study that high attendance rates in AA meetings were associated with treatment eligibility suggests that adherence to alcohol treatment programs may be associated with improved interferon treatment eligibility.

As in previous studies of substance abusing populations, we found a very high prevalence of substantial depressive and other psychiatric symptoms (Johnson et al., 1998; Loftis and Hauser, 2004). Psychiatric disease is associated with a high incidence of interferon-related depression and drug relapse (Edlin, 2002; Fleming et al., 2003; Kraus et al., 2001). In this study, depressive symptoms were present despite high rates of psychiatric care, suggesting that the availability of these services was not sufficient to mitigate HCV therapy ineligibility related to depressive symptoms.

White race and nonreceipt of HAART were associated with treatment ineligibility, and homelessness was of borderline significance. While the impact of HAART and homelessness might be expected, it is unclear why those of white race may be less eligible for interferon therapy. However, at least 2 recent studies have assessed treatment eligibility according to race or ethnicity and both, as in this study, have suggested that minorities had higher treatment eligibility rates (Bini et al., 2005; Cheung et al., 2005).

This study has several limitations. The proportion of subjects eligible for interferon therapy may have been overestimated as a number of factors that might affect treatment eligibility were not assessed. These include patient acceptance of interferon, as well as a variety of medical issues including diabetes mellitus, other autoimmune diseases, and significant cardiac, pulmonary, and renal disease (Fleming et al., 2003; Restrepo et al., 2005). Nonetheless, even if these contraindications led to reduction of our total number of eligible patients for interferon therapy, the potential impact of substance use and mental health would remain large. As this study is largely derived from a single urban center, it may be less generalizable to other populations. Both a limitation and a strength of this study is that we used strict eligibility criteria in the assessment of contraindications for HCV therapy. This assessment was based upon recent guidelines, rather than physician opinion, about individual patients. Thus, we produced 3 categories of treatment eligibility, recognizing that many persons deemed to have a relative contraindication to therapy could receive it in the appropriate clinical context. It is also true that treatment eligibility is likely to vary over time such that treatment-ineligible persons may become eligible as their clinical and social circumstances change.

In summary, almost half of all HCV/HIV-infected individuals with a history of alcohol problems have absolute contraindications that make them ineligible for interferon-based therapies; a substantial proportion have relative contraindications based on other social and psychological barriers. The single largest barrier to interferon therapy was alcohol use, while illicit drug use and mental health problems were also substantial issues. Furthermore, these barriers were present despite the high proportion of patients utilizing psychiatric and substance abuse treatment. Improvements in interferon treatment eligibility will likely require increased coordination of substance abuse and psychiatric treatment with those providing HCV and HIV clinical care, a general recommendation endorsed by a recent IOM report (committee on Crossing the Quality Chasm, 2006). In those patients who

remain ineligible for interferon therapy, continued education and support to avoid alcohol use, maintain immune function, and select appropriate antiretroviral agents with reduced hepatotoxicity should be emphasized to limit liver injury. Interferon treatment eligibility should be reviewed regularly as treatment eligibility is likely to change over time. Addressing the important comorbidities, in particular alcohol, drug, and mental health problems of patients coinfected with HIV and HCV, presents an opportunity to improve the delivery of effective HCV pharmacotherapy to this particular at-risk population.

REFERENCES

Adeyemi OM, Jensen D, Attar B, Ghaoui R, Gallagher M, Wolen D, Cotler SJ (2004) Hepatitis C treatment eligibility in an urban population with and without HIV coinfection. AIDS Patient Care STDS 18:239–245.

Backmund M, Meyer K, Von Zielonka M, Eichenlaub D (2001) Treatment of hepatitis C infection in injection drug users. Hepatology 34:188–193.

Bhattacharya R, Shuhart MC (2003) Hepatitis C and alcohol: interactions, outcomes, and implications. J Clin Gastroenterol 36:242–252.

Bini EJ, Brau N, Currie S, Shen H, Anand BS, Hu KQ, Jeffers L, Ho SB, Johnson D, Schmidt WN, King P, Cheung R, Morgan TR, Awad J, Pedrosa M, Chang KM, Aytaman A, Simon F, Hagedorn C, Moseley R, Ahmad J, Mendenhall C, Waters B, Strader D, Sasaki AW, Rossi S, Wright TL (2005) Prospective multicenter study of eligibility for antiviral therapy among 4084 U. S. veterans with chronic hepatitis C virus infection. Am J Gastroenterol 100:1772–1779.

Blackard JT, Yang Y, Bordoni P, Sherman KE, Chung RT (2004) Hepatitis C virus (HCV) diversity in HIV-HCV-coinfected subjects initiating highly active antiretroviral therapy. J Infect Dis 189:1472–1481.

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.

Burack JH, Barrett DC, Stall RD, Chesney MA, Ekstrand ML, Coates TJ (1993) Depressive symptoms and CD4 lymphocyte decline among HIV-infected men. JAMA 270:2568–2573.

Campbell JV, Hagan H, Latka MH, Garfein RS, Golub ET, Coady MH, Thomas DL, Strathdee SA (2006) High prevalence of alcohol use among hepatitis C virus antibody positive injection drug users in three US cities. Drug Alcohol Depend 81:259–265.

Carlos Martin J, Castilla J, Lopez M, Arranz R, Gonzalez-Lahoz J, Soriano V (2004) Impact of chronic hepatitis C on HIV-1 disease progression. HIV Clin Trials 5:125–131.

Carrat F, Bani-Sadr F, Pol S, Rosenthal E, Lunel-Fabiani F, Benzekri A, Morand P, Goujard C, Pialoux G, Piroth L, Salmon-Ceron D, Degott C, Cacoub P, Perronne C (2004) Pegylated interferon alfa-2b vs standard interferon alfa-2b, plus ribavirin, for chronic hepatitis C in HIV-infected patients: a randomized controlled trial. JAMA 292: 2839–2848.

Cheung RC, Currie S, Shen H, Ho SB, Bini EJ, Anand BS, Brau N, Wright TL (2005) Chronic hepatitis C in Latinos: natural history, treatment eligibility, acceptance, and outcomes. Am J Gastroenterol 100:2186–2193.

Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders (2006) A framework for improving quality, in *Improving the Quality of Health Care for Mental and Substance-Use Conditions*, pp 56–76. Institute of Medicine, National Academy Press, Washington, DC.

Di Martino V, Rufat P, Boyer N, Renard P, Degos F, Martinot-Peignoux M, Matheron S, Le Moing V, Vachon F, Degott C, Valla D, Marcellin P (2001) The influence of human immunodeficiency virus coinfection on chronic hepatitis C in injection drug users: a long-term retrospective cohort study. Hepatology 34:1193–1199.

Edlin BR (2002) Prevention and treatment of hepatitis C in injection drug users. Hepatology 36 (Suppl 1): S210–S219.

Falck-Ytter Y, Kale H, Mullen KD, Sarbah SA, Sorescu L, McCullough AJ (2002) Surprisingly small effect of antiviral treatment in patients with hepatitis C. Ann Intern Med 136:288–292.

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

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

Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL Jr., Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 347:975–982.

Galvan FH, Bing EG, Fleishman JA, London AS, Caetano R, Burnam MA, Longshore D, Morton SC, Orlando M, Shapiro M (2002) The prevalence of alcohol consumption and heavy drinking among people with HIV in the United States: results from the HIV Cost and Services Utilization Study. J Stud Alcohol 63:179–186.

Hadziyannis SJ, Sette H Jr., Morgan TR, Balan V, Diago M, Marcellin P, Ramadori G, Bodenheimer H Jr., Bernstein D, Rizzetto M, Zeuzem S, Pockros PJ, Lin A, Ackrill AM (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med 140: 346–355.

Harris DR, Gonin R, Alter HJ, Wright EC, Buskell ZJ, Hollinger FB, Seeff LB (2001) The relationship of acute transfusion-associated hepatitis to the development of cirrhosis in the presence of alcohol abuse. Ann Intern Med 134:120–124.

Hezode C, Lonjon I, Roudot-Thoraval F, Pawlotsky JM, Zafrani ES, Dhumeaux D (2003) Impact of moderate alcohol consumption on histological activity and fibrosis in patients with chronic hepatitis C, and specific influence of steatosis: a prospective study. Aliment Pharmacol Ther 17:1031–1037.

Johnson ME, Fisher DG, Fenaughty A, Theno SA (1998) Hepatitis C virus and depression in drug users. Am J Gastroenterol 93: 785–789.

Kraus MR, Schafer A, Csef H, Faller H, Mork H, Scheurlen M (2001) Compliance with therapy in patients with chronic hepatitis C: associations with psychiatric symptoms, interpersonal problems, and mode of acquisition. Dig Dis Sci 46:2060–2065.

Loftis JM, Hauser P (2004) The phenomenology and treatment of interferon-induced depression. J Affect Disord 82:175–190.

Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 358:958–965.

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

Mochida S, Ohnishi K, Matsuo S, Kakihara K, Fujiwara K (1996) Effect of alcohol intake on the efficacy of interferon therapy in patients with chronic hepatitis C as evaluated by multivariate logistic regression analysis. Alcohol Clin Exp Res 20 (suppl): 371A–377A.

National Academy of Sciences (2006) Improving the quality of health care for mental and substance-use conditions, in *Quality Chasm Series*. National Academies of Science.

Ohnishi K, Matsuo S, Matsutani K, Itahashi M, Kakihara K, Suzuki K, Ito S, Fujiwara K (1996) Interferon therapy for chronic hepatitis C in

habitual drinkers: comparison with chronic hepatitis C in infrequent drinkers. Am J Gastroenterol 91:1374–1379.

Okazaki T, Yoshihara H, Suzuki K, Yamada Y, Tsujimura T, Kawano K, Abe H (1994) Efficacy of interferon therapy in patients with chronic hepatitis C. Comparison between non-drinkers and drinkers. Scand J Gastroenterol 29:1039–1043.

Peters MG, Terrault NA (2002) Alcohol use and hepatitis C. Hepatology 36 (suppl 1): S220–S225.

Poynard T, Bedossa P, Opolon P (1997) Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 349:825–832.

Restrepo A, Johnson TC, Widjaja D, Yarmus L, Meyer K, Clain DJ, Bodenheimer HC Jr, Min AD (2005) The rate of treatment of chronic hepatitis C in patients co-infected with HIV in an urban medical centre. J Viral Hepat 12:86–90.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH (1988) The composite international diagnostic interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 45: 1069–1077.

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

Safdar K, Schiff ER (2004) Alcohol and hepatitis C. Semin Liver Dis 24:305–315.

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.

Schaefer M, Heinz A, Backmund M (2004) Treatment of chronic hepatitis C in patients with drug dependence: time to change the rules? Addiction 99:1167–1175.

Sherman KE, Rouster SD, Chung RT, Rajicic N (2002) Hepatitis C virus prevalence among patients infected with human immunodeficiency virus: a cross-sectional analysis of the US adult AIDS clinical trials group. Clin Infect Dis 34:831–837.

Smith K, Horton N, Saitz R, Samet J (2006) Use of mini-mental state examination in clinical substance use research studies. Drug Alcohol Depend 82:231–237.

Sobell LC, Cellucci T, Nirenberg TD, Sobell MB (1982) Do quantity-frequency data underestimate drinking-related health risks? Am J Public Health 72:823–828.

Sobell LC, Toneatto T, Sobell MB, Leo GI, Johnson L (1992) Alcohol abusers' perceptions of the accuracy of their self-reports of drinking: implications for treatment. Addict Behav 17:507–511.

Sylvestre DL (2002) Treating hepatitis C in methadone maintenance patients: an interim analysis. Drug Alcohol Depend 67:117–123.

Sylvestre DL (2005a) Approaching treatment for hepatitis C virus infection in substance users. Clin Infect Dis 41 (suppl 1): S79–S82.

Sylvestre DL (2005b) Treating hepatitis C virus infection in active substance users. Clin Infect Dis 40 (suppl 5): S321–S324.

Tabone M, Sidoli L, Laudi C, Pellegrino S, Rocca G, Della Monica P, Fracchia M, Galatola G, Molinaro GC, Arico S, Pera A (2002) Alcohol abstinence does not offset the strong negative effect of lifetime alcohol consumption on the outcome of interferon therapy. J Viral Hepat 9:288–294.

Taylor LE (2005) Delivering care to injection drug users coinfected with HIV and hepatitis C virus. Clin Infect Dis 40 (suppl 5): S355–S361.

Torriani FJ, Rodriguez-Torres M, Rockstroh JK, Lissen E, Gonzalez-Garcia J, Lazzarin A, Carosi G, Sasadeusz J, Katlama C, Montaner J, Sette H Jr, Passe S, De Pamphilis J, Duff F, Schrenk UM, Dieterich DT (2004) Peginterferon Alfa2a plus ribavirin for chronic hepatitis C virus infection in HIV-infected patients. N Engl J Med 351:438–450.

Health Literacy, Antiretroviral Adherence, and HIV-RNA Suppression

A Longitudinal Perspective

Michael K. Paasche-Orlow, MD, MA, MPH,[1] Debbie M. Cheng, ScD,[1] Anita Palepu, MD, MPH,[2] Seville Meli, MPH,[1] Vincent Faber, MPH,[3] Jeffrey H. Samet, MD, MA, MPH[1]

[1]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]University of British Columbia, Vancouver, BC, Canada; [3]DM-Stat, Medford, MA, USA.

BACKGROUND: Low health literacy has been associated with worse adherence to antiretroviral therapy (ART) and higher HIV-RNA levels, but these relationships have not been evaluated in longitudinal analyses.

METHODS: We evaluated literacy using the Rapid Estimate of Adult Literacy in Medicine (REALM) (≤6th grade, 7th to 8th grade, ≥9th grade) in the HIV-Alcohol Longitudinal Cohort study of HIV-infected persons with a history of alcohol problems, conducted from 1997 to 2001. We tested HIV-RNA levels and administered a standardized questionnaire regarding demographics, substance use, receipt of ART, and adherence with ART, every 6 months for up to 7 occasions. Among the 235 subjects on ART, we investigated the relationship between literacy and 2 outcomes: 100% 3-day self-reported adherence and HIV-RNA suppression (<500 copies).

RESULTS: Subjects' literacy levels were the following: 14% ≤6th grade, 29% 7th to 8th grade, and 57% ≥9th grade. In 66% of the observations (478/725), subjects reported 100% 3-day adherence with ART. Of the 685 HIV-RNA assays from these subjects, 62% had <500 copies. In unadjusted analyses, subjects with the lowest literacy level (≤6th grade) had a higher odds of adherence (odds ratio [OR] 2.23, 95% confidence interval 1.15 to 4.30) and HIV-RNA suppression (OR 2.01, 95% confidence interval 1.03 to 3.90) compared with those with ≥9th grade literacy. This trend persisted but was no longer statistically significant in adjusted models of adherence (AOR 1.93, 95% confidence interval 0.86 to 4.31) and HIV-RNA suppression (AOR 1.70, 95% confidence interval 0.79 to 3.65).

CONCLUSION: Contrary to our hypothesis, low literacy was not associated with a lower odds of adherence or virologic suppression in this longitudinal analysis of HIV-infected patients with a history of alcohol problems. Indeed, trends in these data suggest the possibility that low literacy may be associated with a higher odds of adherence and virologic suppression. These counterintuitive findings underscore the need to pursue a fuller understanding of the mechanisms by which literacy affects health outcomes.

KEY WORDS: HIV; adherence; health literacy; literacy.
DOI: 10.1111/j.1525-1497.2006.00527.x
J GEN INTERN MED 2006; 21:835–840.

Health literacy has emerged as a cross-cutting priority to improve the quality of health and health care in America.[1–4] Medical and public health literature highlight the high reading demands made on people who are often in need of important health information.[5] In addition, there is a growing body of literature indicating that people with limited health literacy have worse health status.[6,7]

None of the authors have any conflicts of interest to declare.

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).

Health literacy has been defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" and includes aspects of basic literacy as well as vision, cognition, hearing, communication, navigation, and culture.[3] Various mechanisms linking low health literacy to worse health outcomes have been proposed.[8] However, as most reports have presented cross-sectional analyses, a mechanistic understanding of how low health literacy might actually cause worse health outcomes has remained elusive.[9,10] One prominent theory maintains that adherence to medication regimens is lower in patients with low health literacy.[8] An excellent model to examine such a mechanism is in patients with HIV as health outcomes are likely to be particularly sensitive to variation in adherence.[11]

While many studies have investigated determinants of antiretroviral therapy (ART) adherence,[12–22] few authors have examined the role of health literacy in ART adherence among patients with HIV. The goal of this paper is to evaluate the role of literacy in longitudinal analyses of adherence and viral load suppression among patients with HIV. We hypothesized that low health literacy would be associated with worse adherence and less frequent viral load suppression.

METHODS

Study Population

This is a longitudinal analysis evaluating data including literacy status, ART adherence, HIV-RNA level, and addiction severity from a prospective cohort of HIV-infected patients with a history of alcohol problems. Between July 1997 and August 2001, we recruited HIV-infected subjects with a history of alcohol problems in the following manner: Boston Medical Center HIV Diagnostic Evaluation Unit 56%; posted flyers 16%; Boston Medical Center Primary Care Clinic 13%; respite facility for homeless persons 5%; methadone clinic 4%; subject referrals 4%; and Beth Israel Deaconess Medical Center 2%. All potential subjects who gave 2 or more positive responses to the CAGE questionnaire (Cut down, Annoyed, Guilty, Eye-opener), a screening test for lifetime alcohol problems, were eligible.[23–25] In addition, potential subjects were eligible if an attending physician made a specific diagnosis of alcohol abuse or dependence.[26,27] Other entry criteria included the following: fluency in English or Spanish; Mini-Mental State Examination score ≥21; and no plans to move from the Boston area within 2 years.[28] We screened 474 subjects at these various sites. There were 422 eligible subjects, of whom 349 (82.5%) provided informed consent and agreed to participate in the original study.[29]

A nested randomized-controlled intervention trial to promote ART adherence was conducted with 150 of the 349 subjects, as depicted in Figure 1. Intervention subjects had:

(1) assessment and discussion of alcohol use based on stage of readiness for behavioral change; (2) use of a watch that served as a medication timer device, a practical aid to improving adherence; (3) counseling to promote perceived efficacy of medications; and (4) individualized HIV counseling and exploration of ways to tailor medication use to specific circumstances. Subjects randomized to the intervention group were scheduled for an initial 60-minute individual appointment (within 2 weeks after randomization), a follow-up home visit within the first 3 weeks, and 2 subsequent 15- to 30-minute appointments at 1 and 3 months with the nurse interventionist who delivered the adherence enhancement intervention.[30,31] In the analysis for this study, appropriate adjustments were made to account for trial status. The Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center approved this study.

For the current analyses, health literacy is the primary independent variable of interest. Thus, subjects were excluded from the analysis if they did not complete the literacy assessment ($n=11$) or if they conducted the research interview in Spanish ($n=23$), as the Rapid Estimate of Adult Literacy in Medicine (REALM) is an English test.[32] In addition, subjects were excluded if they were not on ART ($n=80$). Consequently, the final cohort for this analysis included 235 subjects, as depicted in Figure 1. If a subject began ART during the period of observation, only data collected after that point were included.

Data Collection

After obtaining informed consent, a research associate interviewed subjects using a standardized instrument to ascertain baseline information. We attempted to obtain CD4 cell counts and HIV-RNA 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. Research interviews and data collection were conducted every 6 months for up to 7 occasions.

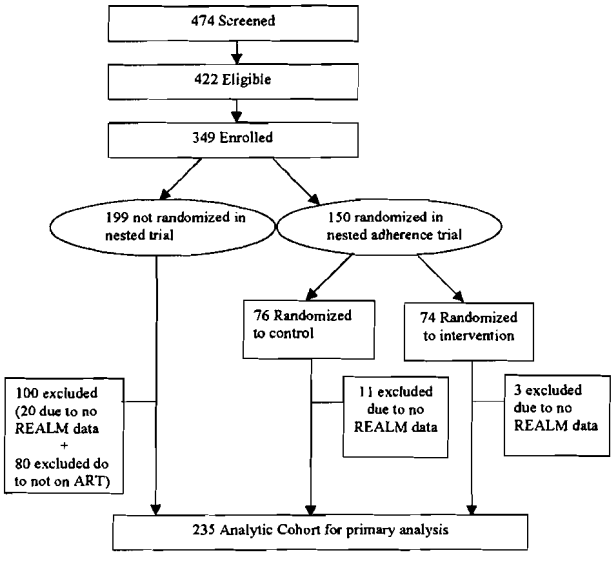


FIGURE 1. Sampling cascade for the analytic cohort.

Outcome Variables

The outcome variables of interest were (1) 3-day ART adherence (100% adherent vs <100% adherent) and (2) viral load suppression (<500 vs ≥500 copies/mL). Adherence was determined with the AIDS Clinical Trials Group Adherence Instrument, a self-reported questionnaire that was modified to evaluate 3-day, as opposed to 2-day, ART adherence.[33] Patients reported the names of the antiretroviral medications, as well as the number of doses and the total number of pills prescribed daily. We defined adherence as a dichotomous variable, in which patients who were <100% adherent during the previous 3 days were considered nonadherent.[12,30] Measurement of HIV-RNA was performed using branched-chain DNA techniques.[34] The threshold for detection at the time of the study was 500 copies/mL; viral load suppression was defined as having an undetectable HIV-RNA. In secondary analyses, we used 30-day self-reported ART adherence (≥95% adherent vs <95% adherent) as an alternate measure.

Primary Independent Variables

Our measure for health literacy was the 66-word REALM.[32] This is a 2 to 3-minute English test of medically relevant vocabulary. The REALM is a valid test of word pronunciation and has been shown to correlate well with tests that evaluate a range of literacy skills.[35] Three categories of literacy were defined based on the REALM: literacy levels of 6th grade and below (REALM score 0 to 44), 7th to 8th grades (REALM score 45 to 60), and 9th grade and above (REALM score 61 to 66).

Other Independent Variables

Other specific variables assessed included: gender, age, race (black, white, or other—the latter subjects were mostly Hispanic), severity of alcohol and drug dependence as measured by the Addiction Severity Index (ASI alcohol and ASI drug),[36] drank to intoxication in the past 30 days, injected drugs in the past 6 months, level of education, complexity of medication regimen, homelessness, and randomization group. Level of education was used as a dichotomous variable representing whether a subject had a high school diploma or general equivalency degree (GED) versus neither. The complexity of the medication regimen was categorized as low (2 to 4 doses/d), moderate (5 to 6 doses/d), or high (7 to 10 doses/d). 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.[37]

Analysis

Bivariate analyses of baseline data were performed to assess the associations between various subject characteristics and literacy. Subject characteristics were compared across literacy groups using the χ^2 test for categorical variables and the nonparametric Kruskal-Wallis test for continuous variables. Separate longitudinal logistic regression models were constructed to examine the association between literacy and each outcome (ART adherence and HIV-RNA suppression) over time. A generalized estimating equations (GEE) approach using an independence working correlation matrix was used to account for correlation due to analyzing repeated measures from the same subject over time, and empirical standard errors were used for all analyses.[38] Multivariable analyses were conducted in the

following sequence: first, gender, race, age, education, drank to intoxication in the past 30 days, injected drugs in the past 6 months, homelessness, and randomization group were entered as covariates; then, the complexity of the medication regimen was added to the model. Models examining HIV-RNA suppression also adjusted for current ART adherence. In addition, we conducted 2 sensitivity analyses to evaluate the stability of our findings for HIV RNA suppression. First, the number of observations included in the analysis was expanded to include instances of discontinuation, i.e., observations where a subject who had been on ART was no longer in receipt of ART. Then, we expanded the analytic cohort to include subjects who had a clinical indication for ART but were not in receipt of ART. All analyses were carried out using SAS (SAS Version 8.2, SAS Institute Inc., Cary, NC).

RESULTS

A total of 235 subjects contributing 725 observations (average 3.1 observations/subject) were included in the current analyses. The distribution of literacy was: 14% ≤6th grade, 29% 7th to 8th grade, and 57% ≥9th grade. In bivariate analyses, higher level of education and being white were associated with higher literacy. Baseline subject characteristics are described in Table 1.

Adherence

Among the 725 observations, 100% 3-day adherence was recorded in 478 (66%). In an unadjusted analysis, low literacy was associated with an increased odds of adherence (OR 2.23 for low vs high literacy level, 95% confidence interval [CI] 1.15 to 4.30). Although not statistically significant, in longitudinal logistic regression models (Table 2), the inverse trend remained between literacy and the odds of adherence. Factors negatively associated with adherence in both Adherence Model 1 and Adherence Model 2 (which includes the complexity of the regimen) were having drunk to intoxication in the past 30 days (Model 2: AOR 0.32, 95% CI 0.21 to 0.48) and injected drugs in the past 6 months (Model 2: AOR 0.26, 95% CI 0.15 to 0.4). Having a less complex medication regimen was associated with a higher odds of adherence (AOR 1.96 for ≤4 doses/d compared with 7 to 10 doses/d; 95% CI 1.08 to 3.58). All findings for 30-day adherence were consistent with the 3-day adherence results.

HIV-RNA Suppression

Among the 685 HIV-RNA tests, suppression was recorded in 422 (62%). In an unadjusted analysis, the odds of HIV-RNA suppression was higher for those with low literacy compared with those with high literacy (OR 2.01 for low vs high literacy level, 95% CI 1.03 to 3.90). Although not statistically significant, in longitudinal logistic regression models (Table 3), the inverse trend remained between literacy and the odds of HIV-RNA suppression. Medication adherence was the only variable that was consistently associated with HIV-RNA suppression (Model 3: AOR 1.80; 95% CI 1.20 to 2.69).

Sensitivity Analysis

In the first sensitivity analysis, we added 101 observations (subjects who had been on ART but who were off ART at a subsequent study visit). In the second sensitivity analysis, we further expanded the analytic cohort to include all observations from subjects for whom ART was clinically indicated (CD4 count ≤350). This resulted in a sample of 266 subjects with 879 observations, an increase of 31 subjects and 53 observations. The analyses conducted on these expanded

Table 1. Characteristics of Subjects in the Analytic Dataset: Members of the HIV-ALC (HIV-Infected Persons with a History of Alcohol Problems) with a Clinical Indication for Antiretroviral Therapy (*N*=266)

Characteristics, *N* (%)	≤6th grade: 32 (14)	7th to 8th grade: 69 (29)	≥9th grade: 134 (57)	Total: 235
Age in y (median [IQR])	41 [7]	42 [12]	42 [9]	42 [9]
Female	7 (22)	14 (20)	28 (21)	49 (21)
Ethnicity				
Black	**24 (75)**	**40 (58)**	**41 (32)**	105 (45)
White	**2 (6)**	**15 (22)**	**73 (54)**	90 (38)
Other	**6 (19)**	**14 (20)**	**20 (15)**	40 (17)
Homeless, y	7 (22)	24 (35)	24 (18)	55 (23)
High school graduate or equivalent degree, y	**6 (19)**	**43 (62)**	**100 (75)**	149 (63)
Nested adherence trial status				
Not in nested trial	13 (41)	23 (33)	63 (47)	99 (42)
Intervention subject in nested trial	6 (19)	24 (35)	41 (31)	71 (30)
Control subject in nested trial	13 (41)	22 (32)	30 (22)	65 (28)
Viral load suppressed at baseline visit, y	20 (63)	37 (58)	78 (61)	135 (60)
Alcohol consumption (median drinks/d [IQR])	4 [4]	6 [19]	5 [6]	6 [9]
Drank to intoxication in the past 30 d	6 (19)	24 (35)	47 (35)	77 (33)
Injected drugs in the past 6 mo	2 (6)	11 (16)	31 (23)	44 (19)
Mean ASI alcohol score (median [IQR])	0.1 [0.3]	0.1 [0.3]	0.1 [0.3]	0.1 [0.3]
Mean ASI drug score (median [IQR])	0.1 [0.2]	0.1 [0.2]	0.1 [0.2]	0.1 [0.2]
100% 3-d adherence at baseline	22 (69)	43 (63)	83 (62)	148 (64)
Complexity of ART regimen at baseline				
2 to 4 doses/d	18 (56)	39 (57)	58 (44)	115 (49)
5 to 6 doses/d	7 (22)	22 (32)	49 (37)	78 (33)
7 to 10 doses/d	7 (22)	7 (10)	26 (20)	40 (17)

P < .05 *indicated by bold text.*

ALC, alcohol longitudinal cohort; ART, antiretroviral therapy; ASI, addiction severity index; IQR, interquartile range.

Table 2. Longitudinal Relationship Between Literacy and ART Adherence

Literacy Level	Unadjusted Odds Ratios (95% CI)	Adjusted Odds Ratios (95% CI)	
		Model 1*	Model 2†
≤ 6th grade	**2.23 (1.15 to 4.30)**	1.90 (0.84 to 4.30)	1.93 (0.86 to 4.31)
7th to 8th grade	1.26 (0.76 to 2.08)	1.33 (0.79 to 2.24)	1.29 (0.77 to 2.19)
≥ 9th grade	1	1	1

Model 1 includes: gender, age, education, randomization group, ethnicity, homeless status, drank to intoxication in the past 30 days, and injected drugs in the past 6 months.

†*Model 2 includes: model 1 (gender, age, education, randomization group, ethnicity, homeless status, drank to intoxication in the past 30 days, injected drugs in the past 6 months)+complexity of regimen.*

P < .05 *indicated by bold text.*

CI, confidence interval.

samples produced results that were similar to the findings from the primary analyses.

DISCUSSION

We evaluated the role of health literacy in ART adherence and HIV-RNA suppression in a 36-month prospective cohort study of HIV-infected patients with a history of alcohol problems. Contrary to our hypotheses, in unadjusted analyses, the odds of ART adherence and HIV-RNA suppression were higher in those with lower literacy. Such trends were noted even after adjusting for gender, age, education, ethnicity, homeless status, drinking to intoxication in the past 30 days, injecting drugs in the past 6 months, participation in an adherence-promoting intervention, and the complexity of the medication regimen, although in adjusted models these findings were not statistically significant.

Our findings run counter to a common hypothesis in the health literacy literature.[39] Indeed, the Institute of Medicine report on health literacy is called "Health Literacy: A Prescription to End Confusion," evoking the importance of medication adherence as a link between low literacy and worse health outcomes.[3] However, evidence about the effects of health literacy on adherence has been mixed. The Agency for Healthcare Research and Quality evidence report on Literacy and Health Outcomes identified 3 studies evaluating the relationship between literacy and medication adherence and 2 of these did not support such an association.[7,20,40,41] Two additional studies have subsequently reported the association between low literacy and low adherence to preoperative medication instructions[42] and medications for cardiovascular diseases.[43]

The 1 prior longitudinal evaluation of health literacy in a cohort of HIV-infected subjects focused exclusively on adherence: Golin et al.[44] followed patients starting on a new ART regimen for 48 weeks. In their cohort, low health literacy was associated with poor ART knowledge 8 weeks after initiation of an ART regimen; however, at 48 weeks there was no association between literacy and adherence.[20] This suggests that low health literacy may be a barrier to adherence early in the course of treatment, but that familiarity with a stable care plan could mitigate such an influence over time.

In a series of cross-sectional reports on 3 cohorts of patients with HIV in Atlanta, Georgia, Kalichman et al.[40,45–47] reported the association between low health literacy and a lower level of knowledge about HIV, worse ART adherence, lower CD4 cell counts, and higher viral loads. However, Wolf et al. reported no association between health literacy and ART adherence in a cohort of patients in a Southern U.S. HIV clinic, and van Servellen et al. described no association between health literacy and ART adherence in the setting of an intervention program for Latino men and women.[48–50]

Why might low literate patients in our study have been more adherent to ART and have better HIV-RNA suppression than those with higher literacy? The current cohort was different from previously analyzed populations in important ways. Unlike patients in the study by Golin et al., subjects in the current study were not recruited at the time of ART initiation. Whereas Kalichman et al. recruited subjects from community outreach venues, the current cohort was recruited in medical settings among people with alcohol problems. Possibly, the resilience required for low literate subjects to access care and participate in this longitudinal study resulted in a

Table 3. Longitudinal Relationship Between Literacy and HIV-RNA Suppression

Literacy Level	Unadjusted Odds Ratios (95% CI)	Adjusted Odds Ratios (95% CI)		
		Model 1*	Model 2†	Model 3‡
6th grade	**2.01 (1.03,3.90)**	1.85 (0.87,3.94)	1.72 (0.81,3.69)	1.70 (0.79,3.65)
7th to 8th grade	1.28 (0.79,2.09)	1.30 (0.77,2.20)	1.28 (0.75,2.16)	1.29 (0.77,2.18)
≥ 9th grade	1		1	1

Model 1 includes: gender, age, education, randomization group, ethnicity, homeless status, drank to intoxication in the past 30 days, and injected drugs in the past 6 months.

†*Model 2 includes: Model 1 (gender, age, education, randomization group, ethnicity, homeless status, drank to intoxication in the past 30 days, and injected drugs in the past 6 months)+medication adherence.*

‡*Model 3 includes: Model 2 (gender, age, education, randomization group, ethnicity, homeless status, drank to intoxication in the past 30 days, and injected drugs in the past 6 months, and medication adherence)+complexity of regimen.*

P < .05 *indicated by bold text.*

CI, confidence interval.

selection bias toward low literate subjects with higher levels of adherence. While this is a possible contributing factor, it is important to note that other variables in the analyses operated as expected (e.g., a less complex drug regimen was associated with better adherence, and having drunk to intoxication in the past 30 days was associated with worse adherence).[17,19,51]

It is also possible that patients with low literacy may have difficulty acquiring the self-management skills required for a new drug regimen, but once in a therapeutic steady state may actually follow directions more readily than those with higher literacy. Low literacy has been linked with low self-efficacy, which has been identified as a risk factor for worse adherence.[52] However, it is also possible that patients with higher levels of self-management efficacy may be predisposed to worse adherence, as may have been the case for the higher literacy subjects in this study.[53,54] An example of how this could occur would be if higher literacy subjects felt they could adjust medications due to perceived side effects without conferring with a health provider. Future analyses to evaluate potential mediators such as aspects of self-efficacy, knowledge, and understanding will be important to elucidate the causal pathways for the varied findings on the association between literacy and adherence.

This study has several limitations. First, adherence was evaluated by self-report. Although self-reported adherence typically correlates with other measures of adherence, other forms of adherence evaluations were not conducted.[55,56] An additional adherence variable (30-day self-report) was evaluated in this study, but these results were not significantly different from those in our primary analyses. Also, self-reported adherence was a significant, independent predictor of viral suppression in our study, which provides a measure of validity for this assessment. Second, adherence measures have not been specifically validated for use with subjects who have low health literacy. While it is possible that systematic bias could be introduced as a result, the central findings were concordant with models of HIV-RNA suppression, an outcome that is not susceptible to such bias. Third, it is possible that the results are because of important factors or interactions not introduced into regression models. It is unlikely, however, that residual confounding plays a significant role in these analyses as the parameter estimates and trends remained remarkably stable across all models. Fourth, health literacy was defined by the 66-word REALM, which is merely a word pronunciation test. While the REALM is the most commonly used tool to measure literacy in the medical literature and correlates well with other established health literacy and basic literacy instruments (r=.84 to .97), a more comprehensive test of health literacy might have provided different results.[4,7] Fifth, HIV-RNA levels used in this analysis and assessment of adherence were not conducted concurrently. While this is not an optimal approach to evaluating the relationship between adherence and HIV-RNA suppression, adherence data likely reflect subjects' general behavior. Sixth, the generalizability of these findings may be limited owing to the fact that the cohort includes patients with a history of alcohol problems. Finally, it is possible that the subjects with low literacy may have received additional support in some fashion because of recognition of their vulnerability. While this is possible, the clinical staff was not informed of the subjects' health literacy status and providers do not readily discern which of their patients have low health literacy.[57,58]

CONCLUSION

Contrary to our hypothesis and prior observational reports, in this longitudinal cohort of HIV-infected patients with a history of alcohol problems, subjects with low health literacy had a consistent trend toward higher odds of adherence and virologic suppression. Although these trends were not statistically significant in all analyses, the counterintuitive findings presented in this paper underscore the need to pursue a fuller understanding of the mechanisms by which literacy affects health outcomes.

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.

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); and RO1-AA10870 (Enhanced Linkage of Alcohol Abusers to Primary Care).

REFERENCES

1. **Institute of Medicine.** Priority Areas for National Action: Transforming Health Care Quality. Committee on Identifying Priority Areas for Quality Improvement. Adams K, Corrigan JM, eds. Washington, DC: National Academies Press; 2003.
2. **Carmona RH.** Sugeon 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. Saturday, June 14, 2003. Available at: http://www.surgeongeneral.gov/news/speeches/ama061403.htm. Accessed February 19, 2004.
3. **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.
4. **Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR.** The prevalence of limited health literacy. J Gen Intern Med. 2005;20:175–84.
5. **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. 1st edn. New York: Jossey-Bass; 1999.
6. **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.
7. **DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP.** Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med. 2004;19:1228–39.
8. **Baker DW.** Reading between the lines: deciphering the connections between literacy and health. J Gen Intern Med. 1999;14:315–7.
9. **Baker DW, Parker RM, Williams MV, et al.** The health care experience of patients with low literacy. Arch Fam Med. 1996;5:329–34.
10. **Report of the National Work Group on Literacy and Health.** Communicating with patients who have limited literacy skills. J Fam Pract. 1998;46:168–76.
11. **Mannheimer S, Friedland G, Matts J, Child C, Chesney M.** The consistency of adherence to antiretroviral therapy predicts biologic outcomes for human immunodeficiency virus-infected persons in clinical trials. Clin Infect Dis. 2002;34:1115–21.
12. **Palepu A, Horton NJ, Tibbetts N, Meli S, Samet JH.** Uptake and adherence to highly active antiretroviral therapy among HIV-infected people with alcohol and other substance use problems: the impact of substance abuse treatment. Addiction. 2004;99:361–8.
13. **Proctor VE, Tesfa A, Tompkins DC.** Barriers to adherence to highly active antiretroviral therapy as expressed by people living with HIV/AIDS. AIDS Patient Care STDS. 1999;13:535–44.
14. **Remien RH, Hirky AE, Johnson MO, Weinhardt LS, Whittier D, Le GM.** Adherence to medication treatment: a qualitative study of facilitators and barriers among a diverse sample of HIV+ men and women in four US cities. AIDS Behav. 2003;7:61–72.

15. **Roberts KJ.** Barriers to and facilitators of HIV-positive patients' adherence to antiretroviral treatment regimens. AIDS Patient Care STDS. 2000;14:155–68.
16. **Spire B, Duran S, Souville M, Leport C, Raffi F, Moatti JP.** Adherence to highly active antiretroviral therapies (HAART) in HIV-infected patients: from a predictive to a dynamic approach. Soc Sci Med. 2002; 54:1481–96.
17. **Ammassari A, Trotta MP, Murri R, et al.** Correlates and predictors of adherence to highly active antiretroviral therapy: overview of published literature. J Acquir Immune Defic Syndr. 2002;31:S123–7.
18. **Avants SK, Margolin A, Warburton LA, Hawkins KA, Shi J.** Predictors of nonadherence to HIV-related medication regimens during methadone stabilization. Am J Addict. 2001;10:69–78.
19. **Fogarty L, Roter D, Larson S, Burke J, Gillespie J, Levy R.** Patient adherence to HIV medication regimens: a review of published and abstract reports. Patient Educ Couns. 2002;46:93–108.
20. **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.
21. **Hinkin CH, Castellon SA, Durvasula RS, et al.** Medication adherence among HIV+ adults: effects of cognitive dysfunction and regimen complexity. Neurology. 2002;59:1944–50.
22. **Kemppainen JK, Levine R, Buffum M, Holzemer W, Finley P, Jensen P.** Antiretroviral adherence in persons with HIV/AIDS and severe mental illness. J Nerv Ment Dis. 2004;192:395–404.
23. **Fiellin DA, Reid MC, O'Connor PG.** Screening for alcohol problems in primary care: a systematic review. Arch Intern Med. 2000;160: 1977–89.
24. **Ewing JA.** Detecting alcoholism. The CAGE questionnaire. JAMA. 1984;252:1905–7.
25. **Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ.** Screening for alcohol abuse using CAGE scores and likelihood ratios. Ann Intern Med. 1991;115:774–7.
26. **Samet JH, Libman H, LaBelle C, et al.** A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. Arch Intern Med. 1995;155: 1629–33.
27. **Samet JH, Phillips SJ, Horton NJ, Traphagen ET, Freedberg KA.** Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. AIDS Res Hum Retroviruses. 2004;20:151–5.
28. **Folstein MF, Folstein SE, McHugh PR.** "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
29. **Samet JH, Horton NJ, Traphagen ET, Lyon SM, Freedberg KA.** Alcohol consumption and HIV disease progression: are they related? Alcohol Clin Exp Res. 2003;27:862–7.
30. **Samet JH, Horton NJ, Meli S, Freedberg KA, Palepu A.** Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. Alcohol Clin Exp Res. 2004;28:572–7.
31. **Samet JH, Horton NJ, Meli S, et al.** A randomized controlled trial to enhance antiretroviral therapy adherence in patients with a history of alcohol problems. Antivir Ther. 2005;10:83–93.
32. **Doak CC, Doak LG, Root JH.** Teaching Patients with Low Literacy Skills. 2nd edn. Philadelphia, PA: J. B. Lippincott Company; 1996.
33. **Chesney MA, Ickovics JR, Chambers DB, et al.** Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments. Patient Care Committee & Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). AIDS Care. 2000;12:255–66.
34. **Pachl C, Todd JA, Kern DG, et al.** Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay. J Acquir Immune Defic Syndr Hum Retrovirol. 1995;8:446–54.
35. **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.
36. **McLellan AT, Luborsky L, Cacciola J, et al.** New data from the Addiction Severity Index. Reliability and validity in three centers. J Nerv Ment Dis. 1985;173:412–23.
37. **Kertesz SG, Horton NJ, Friedmann PD, Saitz R, Samet JH.** Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. J Subst Abuse Treat. 2003;24:197–207.
38. **Zeger SL, Liang KY.** Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986;42:121–30.
39. **Baker DW.** Reading between the lines: deciphering the connections between literacy and health. J Gen Intern Med. 1999;14:315–7.
40. **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.
41. **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.
42. **Chew LD, Bradley KA, Flum DR, Cornia PB, Koepsell TD.** The impact of low health literacy on surgical practice. Am J Surg. 2004;188:250–3.
43. **Murray MD, Wu J, Tu W, et al.** Health literacy predicts medication adherence. Clin Pharmacol Ther. 2004;75:76.
44. **Miller LG, Liu H, Hays RD, et al.** Knowledge of antiretroviral regimen dosing and adherence: a longitudinal study. Clin. infect Dis. 2003;36:514–8.
45. **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.
46. **Kalichman SC, Rompa D, Cage M.** Reliability and validity of self-reported CD4 lymphocyte count and viral load test results in people living with HIV/AIDS. Int J STD AIDS. 2000;11:579–85.
47. **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.
48. **van Servellen G, Carpio F, Lopez M, et al.** Program to enhance health literacy and treatment adherence in low-income HIV-infected Latino men and women. AIDS Patient Care STDS. 2003;17:581–94.
49. **van Servellen G, Brown JS, Lombardi E, Herrera G.** Health literacy in low-income Latino men and women receiving antiretroviral therapy in community-based treatment centers. AIDS Patient Care STDS. 2003;17:283–98.
50. **Wolf MS, Davis TC, Cross JT, Marin E, Green K, Bennett CL.** Health literacy and patient knowledge in a Southern US HIV clinic. Int J STD AIDS. 2004;15:747–52.
51. **Stone VE, Hogan JW, Schuman P, et al.** Antiretroviral regimen complexity, self-reported adherence, and HIV patients' understanding of their regimens: survey of women in the her study. J Acquir Immune Defic Syndr. 2001;28:124–31.
52. **Roter DL, Stashefsky-Margalit R, Rudd R.** Current perspectives on patient education in the US. Patient Educ Couns. 2001;44:79–86.
53. **Kerr T, Palepu A, Barness G, et al.** Psychosocial determinants of adherence to highly active antiretroviral therapy among injection drug users in Vancouver. Antivir Ther. 2004;9:407–14.
54. **Kerr T, Marshall A, Walsh J, et al.** Determinants of HAART discontinuation among injection drug users. AIDS Care. 2005;17:539–49.
55. **Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB.** The concordance of self-report with other measures of medication adherence: a summary of the literature. Med Care. 2004;42:649–52.
56. **Arnsten JH, Demas PA, Grant RW, et al.** Impact of active drug use on antiretroviral therapy adherence and viral suppression in HIV-infected drug users. J Gen Intern Med. 2002;17:377–81.
57. **Bass PF III, Wilson JF, Griffith CH, Barnett DR.** Residents' ability to identify patients with poor literacy skills. Acad Med. 2002;77:1039–41.
58. **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.

COMMENTARIES

How Health Care Systems Can Begin to Address the Challenge of Limited Literacy

Michael K. Paasche-Orlow, MD, MA, MPH,[1] Dean Schillinger, MD,[2] Sarah M. Greene, MPH,[3] Edward H. Wagner, MD, MPH[3]

[1]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA;
[2]UCSF Division of General Internal Medicine, San Francisco General Hospital, University of California, San Francisco, CA, USA;
[3]MacColl Institute for Healthcare Innovation, Center for Health Studies, Group Health Cooperative, Seattle, WA, USA.

KEY WORDS: health care systems; self-management; health literacy; literacy.
DOI: 10.1111/j.1525-1497.2006.00544.x
J GEN INTERN MED 2006; 21:884–887.

Limited literacy has been shown to be associated with poor health in a wide variety of settings, and is particularly prevalent among the elderly, minorities, those with lower levels of educational attainment, and those with chronic disease.[1] The literacy and health literature calls attention to the ways in which the current health care system is inadequate, not only for the estimated 90 million U.S. adults with limited literacy, but for most users of the system. The implications of limited literacy should be understood as a challenge to the basic justice of a health care system organized for the most highly educated and powerful members of our society.

The National Institutes of Health have 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."[2,3] According to this definition, health literacy relates to both the cognitive and functional skills a person has to make health-related decisions. This definition is problematic from a number of perspectives. While an individual's health literacy is likely to be associated with their literacy level, as suggested by the article by Fang et al.[4] in this issue, we believe that an individual's level of health literacy is not a fixed characteristic, and that it should not be defined only via an evaluation of an individual's skills. Rather, health literacy reflects the contextual demands placed on the individual by (a) their specific clinical condition and associated health care decisions, (b) the communication characteristics of the dominant medical culture, (c) the structure and function of clinical services that assume limitless health literacy and require self-advocacy and vigilance, and (d) the emphasis that society (fueled by a health consumer-oriented marketplace) places on individual, rather than ecological, determinants of health. As such, at a minimum, when we conceptualize health literacy, we consider not only a patient's literacy and numeracy skills but also the complexity of the tasks required, the accessibility of the health care workforce for the target populations, the preparedness of this health care workforce to engage productively with the patient, and the features of the health care system and communities in which care-giving and self-management support take place.

The Institute of Medicine has identified health literacy as a national priority area for transforming health care quality.[5] How will this occur? A fundamental reevaluation of health care in America is warranted. The goal of this paper is to shift the focus of inquiry and analysis from the patient to the system. We offer 3 overarching principles to guide needed adaptations to health care. Our suggestions reflect changes to the organization and delivery of health care based on an integration of emerging research findings related to literacy and the Care Model.[6] Such changes could ameliorate not only the health effects of limited literacy, but improve the overall quality of U.S. health care and engender a more "health literate" society.[7,8]

FIRST PRINCIPLE: PROMOTE PRODUCTIVE INTERACTIONS

Exercise Universal Precautions to Assure Comprehension

Over 300 studies have demonstrated that most patient education materials, explanations of health services and benefits, and documents that purport to advance patients' rights are incomprehensible to a significant portion of the patients we serve. However, in the current paradigm, limited literacy is considered to be the exception to the rule. We advocate that a process of confirming comprehension should be the standard in clinical care, and a basic universal precaution embedded into practice at multiple levels, using multiple methods. For example, to confirm that patients understand their medication regimens, clinicians should ask patients how they plan to take their medications. If a patient does not understand, the clinician can tailor teaching and reassess comprehension until the patient has exhibited mastery (teach-to-goal). This iterative "teach-back" and "teach-to-goal" approach attends to a wide range of factors (e.g., literacy, anxiety, culture, distracting symptoms) that can influence a patient's understanding, and has been endorsed as a patient safety standard to improve informed consent by the National Quality Forum.[9] While the article by Fang et al.[4] in this issue raises serious questions as to the extent of informed decision making for those with limited literacy undergoing long-term anticoagulation with warfarin, the article by Sudore et al.[10] illustrates the potential benefits of embedding an iterative educational practice into decision-making processes in health care.

Improve Providers' Communication Capacities

Patients with limited literacy, when compared with those with adequate literacy, more often report that their doctors use words they do not understand, speak too fast, do not provide

None of the authors have any conflict of interest to declare.

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).

enough information about medical conditions, and fail to make certain that they understand their health problems.[11] Similarly, limited literacy has been associated with more distrust of providers, pessimism about treatment, lower satisfaction, and a worse assessment of the quality of care.[12,13] Providers tend to be unaware of their patients' limited literacy,[14,15] but screening for limited literacy does not appear, in and of itself, to facilitate successful communication.[16] To do so, providers need to (1) learn a set of communication skills, including how to convey empathy, promote trust, and encourage dialogue, how to elicit patient questions, and how to confirm comprehension and tailor education; (2) be imbued with a set of attitudes that can foster productive relationships and therapeutic alliances; and (3) be provided with system-level supports, including time, tools, and incentives, that enable them to utilize these skills. Given the growing racial, ethnic, and linguistic diversity of patient populations, there is also an urgent need to increase the diversity of the health care workforce and to expand the responsibilities of mid-level practitioners and community health workers.

Develop Communication Technology Platforms and Implement Models to Promote Meaningful Communication

Communication about complex ideas can be facilitated by pictures, video, multimedia, and other decision aids.[17-21] Technological support for patient education and collaborative goal setting has begun to proliferate. However, to capitalize on the potential role that such media have in engaging patients with limited literacy, prototypes need to be developed and evaluated. Several promising interactive education technology platforms that customize content according to the patient's responses and provide information to patients and providers (e.g., automated phone systems, touch screens, and embodied conversational agents) are emerging and are being tested in clinical studies.[22,23] How to integrate these technologies into clinical care to ensure broad reach remains a critical question.

SECOND PRINCIPLE: ADDRESS THE ORGANIZATION OF HEALTH CARE

Make Patient-Centered Care a System Property

High-quality medical care integrates evidence-based clinical care with a patient-centered orientation. A patient-centered orientation for health care is one that: (1) includes preactivation to prepare patients and tailor appropriate messages; (2) prioritizes collaborative goal-setting and relationship-centered care during the visit; (3) delivers postvisit reinforcements and follow-up services for both cognitive and behavioral outcomes; (4) offers proactive surveillance during the intervisit period to identify unanticipated changes in health trajectory or access difficulties; and (5) broadens the array of available self-management support strategies. The systematic delivery of these steps requires more than a motivated clinician; it requires a redesigned care system as described by the Care Model.[24] A growing body of literature suggests that tailored implementation of elements of the Care Model can disproportionately benefit those with limited literacy[25,26]; however, only a very small proportion of patients with limited literacy have access to such programs.

Additional technological opportunities in areas such as electronic messaging, internet-based personal health records, and biometric sensing may be able to further advance patient-centered care by providing opportunities for portable records and bidirectional data. However, such efforts will remain tools for the *digerati* unless the interface systems are simplified and proven to be useful for both patients with limited literacy and their providers, and usable in the actual care setting. Ultimately, there should be many options for self-management support, so that we move from a one-size-fits-all approach to one that enables patients and providers to select which is best for them. As described above, expanding the roles and responsibilities of mid-level and community health workers can also reap health dividends.

Streamline, Simplify, and Standardize

There is a tremendous need to simplify and standardize how patients access and utilize the U.S. health system. This is relevant across a range of processes including applications for publicly financed health insurance, understanding of patients' rights and end-of-life decision making, health-plan benefits, pharmacy formularies, and self-management support resources. Patients at every education level will benefit from a reduction of paperwork, plain and simple communication, and standardized processes. Satisfaction, comprehension, and retention of information are enhanced for all patients when they are presented with plain language materials. The current financing structure of our health system and the increasing focus on patient "choice" in health plans and benefits as a means to control costs are likely to increase the complexity of health care. Davis et al.[27] in the current issue, present alarming data regarding comprehension of the most common prescription warning labels and, in their discussion, note the absence of a standardized system of warning labels in the United States.

Develop Structures, Incentives, and Reinforcements to Meet Quality Targets for Vulnerable Populations

In an editorial in this issue, Parker and Kindig[28] discuss the ongoing need for research to advance the cause of patients with limited health literacy. We agree. However, an important reason for why many of the ideas listed above have not become standard of care is that there are few financial incentives. Providers and systems should be rewarded for investing in technologies to support patient education and self-management, and for engaging their patients in the use of these system supports. Currently, financial and time pressures act as disincentives to some of the basic activities and structures needed for patients with limited literacy. While there has been interest among payers, purchasers, and policy makers in having quality improvement driven by financial incentives, current pay-for-performance contracts do not promote health literacy-related processes, such as reducing rates of discrepancies in medication regimens, promoting patient activation, or establishing collaborative behavioral action plans. Projects that target quality indicators such as hemoglobin A1C testing have not been shown to improve outcomes.[29] While the typical pay-for-performance arrangement is for health providers to receive bonuses for meeting or exceeding such targets, there is little

evidence that this leads to quality improvement and some evidence to suggest that it directs resources to providers with higher performance at baseline.[30,31] Given the concentration of patients with limited literacy in underresourced, public sector delivery systems, financial incentives need to be designed to "level the playing field" and promote the types of collaborative health care teams and system redesign needed to address the needs of vulnerable populations. One benefit of financial incentives is that, regardless of the effect on quality, they do seem to spur infrastructure investments.[32] However, to ensure that infrastructural improvements (e.g., registries or other information technology help, self-management resources) can benefit those with limited health literacy, resources, and incentives need to be allocated specifically to health care settings that care for a disproportionate share of vulnerable populations.

There is an urgent need to develop complementary quality metrics that can serve as markers for health literacy-related quality of care. Careful attention needs to ensure that measures of patient experience, routinely used in performance assessment, adequately capture the perspective of patients with limited literacy. For example, the Consumer Assessment of Health Providers and Systems (CAHPS) is typically administered as a mailed survey, and the contribution of limited literacy to response bias has not been evaluated.[33] Current efforts on the part of the Joint Commission and the National Quality Forum to integrate the issue of literacy in their safety and quality initiatives provide hopeful harbingers of needed attention.[9,34]

THIRD PRINCIPLE: EMBRACE A COMMUNITY-LEVEL, ECOLOGICAL PERSPECTIVE

Develop Intervention Models that Acknowledge the Multilevel Nature Of Vulnerability

Most researchers have attempted to isolate the independent effects of limited literacy on health care quality. Based on such work, several have suggested pathways by which limited health literacy may lead to worse outcomes.[35–40] While logical from a biomedical perspective, this reductionism does not acknowledge the mutuality of various biopsychosocial, economic, environmental, and cultural factors influencing health and health care for those with limited literacy. In the current issue, Sentell and Halpin[41] provide evidence that limited literacy contributes to racial and ethnic health disparities. Vulnerabilities such as limited literacy often coexist and interact with other social vulnerabilities, at both the individual and community level, and successful intervention efforts often need to attend to an array of influences on peoples' lives. In another project reported in this issue, Weiss et al.[42] referred patients with depression and limited literacy to a community-based adult literacy program and demonstrated that adult basic education is an effective adjuvant therapy to depression care. These studies provide examples of the interrelationships between social vulnerabilities, with important implications for the design of future interventions.

Advocate for, and Develop More Robust, Independent, and Trusted Public Health Communication Voice(s)

Current "independent" sources of health information include the doctor or office staff, health plan, family and friends, and government agencies. However, a broad array of "non-independent," far-reaching, and often competing health communication channels now exist, including coverage of health issues in the mass media, direct-to-consumer prescription drug advertising, health consumer industry advertising, internet sources, and entertainment television, often in the form of health-related "reality" programming. These channels are increasingly influencing the public's awareness of health issues, redefining what is health and illness, shaping consumers' expectations of health and their demands on the health care system, and narrowing public opinion regarding the attribution of and solutions to common health problems to the level of the individual. While the social marketing techniques used by the private sector are extremely effective in influencing public opinion and creating demand for services, there is evidence that such messages may have disproportionate uptake among those with lower educational attainment.[43] It is apparent that, in the midst of this cacophony of voices, there is an urgent need to develop effective, reliable, and objective voices for health communication messages that can be delivered at home, at work, at school, and in the community.

CONCLUSION

The growing literacy and health literature calls attention to the ways in which the U.S. health care system is inadequate and even unjust, not only for the estimated 90 million U.S. adults with limited literacy, but for many other users of the system. We have presented 3 overarching principles for health system transformation that focus on promoting productive interactions between patients and providers, reorganizing health care delivery, and embracing a community level and ecological perspective. We believe that instituting such changes could improve the quality of care not only for patients with limited literacy, but for all health care consumers, and could contribute to the development of a more "health literate" society.

Dr. Schillinger was supported by NIH Clinical Scientist Award K23RR16539-04.

REFERENCES

1. **DeWalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP.** Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med. 2004;19:1228–39.
2. Understanding and promoting health literacy (R01) PAR-04-116. Available at: http://grants2.nih.gov/grants/guide/pa-files/PAR-04-116.html. Accessed April 7, 2005.
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. Available at: http://www.nlm.nih.gov/pubs/cbm/hliteracy.html. Accessed May 9, 2006. Vol. NLM Pub. No. CBM 2000-1. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services; 2000.
4. **Fang MC, Machtinger EL, Wang F, Schillinger D.** Health literacy and anticoagulation-related outcomes among patients taking warfarin. J Gen Intern Med. 2006;21:841–6.
5. **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.
6. Improving chronic illness care: A National Program of the Robert Wood Johnson Foundation. Available at: http://www.improvingchroniccare.org/. Accessed May 9, 2006.
7. **Wagner EH.** Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract. 1998;1:2–4.

8. **Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A.** Improving chronic illness care: translating evidence into action. Health Aff (Millwood). 2001;20:64–78.
9. Improving patient safety through informed consent for patients with limited health literacy: an implementation report. National Quality Forum. Available at: http://www.qualityforum.org/docs/informed_consent/webinformedconsentMember+public09-13-05.pdf. Accessed May 9, 2006.
10. **Sudore RL, Landefeld CS, Williams B, Barnes D, Lindquist KS, Schillinger D.** Use of a modified informed consent process among vulnerable patients: a descriptive study. J Gen Intern Med. 2006;21:867–73.
11. **Schillinger D, Bindman A, Wang F, Stewart A, Piette J.** Functional health literacy and the quality of physician-patient communication among diabetes patients. Patient Educ Couns. 2004;52:315–23.
12. **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.
13. **Mancuso CA, Rincon M.** Asthma patients' assessments of health care and medical decision making: the role of health literacy. J Asthma. 2006;43:41–4.
14. **Bass PF, Wilson JF, Griffith CH, Barnett DR.** Residents' ability to identify patients with poor literacy skills. Acad Med. 2002;77:1039–41.
15. **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.
16. **Seligman H, Wang F, Palacios J, Wilson C, Daher C, Schillinger D.** Physician notification of their diabetes patients with limited health literacy: a randomized, controlled trial. J Gen Intern Med. 2005;20:1001–7.
17. **Houts PS, Doak CC, Doak LG, Loscalzo MJ.** The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns. 2006;61:173–90.
18. **Wofford JL, Smith ED, Miller DP.** The multimedia computer for office-based patient education: a systematic review. Patient Educ Couns. 2005;59:148–57.
19. **Krishna S, Balas EA, Boren SA, Maglaveras N.** Patient acceptance of educational voice messages: a review of controlled clinical studies. Methods Inf Med. 2002;41:360–9.
20. **O'Connor AM, Stacey D, Rovner D, et al.** Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2001; CD001431.
21. **Lewis D.** Computer-based approaches to patient education: a review of the literature. J Am Med Inform Assoc. 1999;6:272–82.
22. **Friedman RH, Stollerman J, Rozenblyum L, et al.** A telecommunications system to manage patients with chronic disease. Medinfo. 1998;9:1330–4.
23. **Bickmore T, Giorgino T.** Health dialog systems for patients and consumers. J Biomed Inform. 2006 (in press).
24. **Wagner EH, Bennett SM, Austin BT, Greene SM, Schaefer JK, Vonkorff M.** Finding common ground: patient-centeredness and evidence-based chronic illness care. J Altern Complement Med. 2005;11(suppl 1):S7–S15.
25. **DeWalt DA, Pignone M, Malone R, et al.** Development and pilot testing of a disease management program for low literacy patients with heart failure. Patient Educ Couns. 2004;55:78–86.
26. **Rothman R, Malone R, Bryant B, Horlen C, DeWalt D, Pignone M.** The relationship between literacy and glycemic control in a diabetes disease-management program. Diabetes Educ. 2004;30:263–73.
27. **Davis TC, Wolf MS, Bass, PF III, et al.** Low literacy impairs comprehension of prescription drug warning labels. J Gen Intern Med. 2006;21:847–51.
28. **Parker RM, Kindig DA.** Beyond the IOM health literacy report: are the recommendations being taken seriously? J Gen Intern Med. 2006;21:891–2.
29. **O'Connor PJ, Desai J, Solberg LI, et al.** Randomized trial of quality improvement intervention to improve diabetes care in primary care settings. Diabetes Care. 2005;28:1890–7.
30. **Rosenthal MB, Frank RG, Li Z, Epstein AM.** Early experience with pay-for-performance: from concept to practice. JAMA. 2005;294:1788–93.
31. **Rosenthal MB, Frank RG.** What is the empirical basis for paying for quality in health care? Med Care Res Rev. 2006;63:135–57.
32. **Casalino L, Gillies RR, Shortell SM, et al.** External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. JAMA. 2003;289:434–41.
33. **Castle NG, Brown J, Hepner KA, Hays RD.** Review of the literature on survey instruments used to collect data on hospital patients' perceptions of care. Health Serv Res. 2005;40:1996–2017.
34. **Health Literacy and Patient Safety.** Joint Commission on Accreditation of Healthcare Organizations. Available at: http://www.jointcommission.org/PublicPolicy/health_literacy.htm. Accessed May 9, 2006.
35. **Baker DW.** Reading between the lines: deciphering the connections between literacy and health. J Gen Intern Med. 1999;14:315–7.
36. **Pignone M, DeWalt D, Sheridan S, Berkman ND, Lohr KN.** Interventions to improve health outcomes for patients with low literacy: a systematic review. J Gen Intern Med. 2005;20.
37. **Roter D.** Health literacy and the patient-provider relationship. In: Schwartzberg JG, VanGeest JB, Wang CC, eds. Understanding Health Literacy: Implications for Medicine and Public Health. American Medical Association Press; 2005:87–100.
38. **Rudd RR, Renzulli D, Pereira A, Daltroy L.** Literacy demands in health care settings: the patient perspective. In: Schwartzberg JG, VanGeest JB, Wang CC, eds. Understanding Health Literacy: Implications for Medicine and Public Health. American Medical Association Press; 2005:69–84.
39. **Schillinger D, Chen AH.** Literacy and language: disentangling measures of access, utilization, and quality. J Gen Intern Med. 2004;19:288–90.
40. **Rootman I, Ronson B.** Literacy and health research in Canada: where have we been and where should we go? Can J Public Health. 2005;96(suppl 2):S62–S77.
41. **Sentell TL, Halpin HA.** Importance of adult literacy in understanding health disparities. J Gen Intern Med. 2006;21:862–6.
42. **Weiss BD, Francis L, Senf JH, Heist K, Hargraves R.** Literacy education as treatment for depression in patients with limited literacy and depression—a randomized controlled trial. J Gen Intern Med. 2006;21:823–8.
43. **Nutbeam D.** Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int. 2000;15:259–67.

ELSEVIER

Journal of Substance Abuse Treatment 31 (2006) 411–417

Journal of Substance Abuse Treatment

Regular article

Substance abuse treatment and receipt of liver specialty care among persons coinfected with HIV/HCV who have alcohol problems

Anita Palepu, (M.D., M.P.H.)[a,*], Debbie M. Cheng, (Sc.D.)[b], Theresa Kim, (M.D.)[c], David Nunes, (M.D.)[d], John Vidaver, (M.A.)[e], Julie Alperen, (Dr.P.H.)[b], Richard Saitz, (M.D., M.P.H.)[b,f], Jeffrey H. Samet, (M.D., M.A., M.P.H.)[b,g]

[a]*Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Department of Medicine, University of British Columbia, Vancouver, BC, Canada V6Z 1Y6*

[b]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*

[c]*Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA*

[d]*Section of Gastroenterology, Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA*

[e]*DM-STAT Inc., Malden, MA, USA*

[f]*Youth Alcohol Prevention Center, and the Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA*

[g]*Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA*

Received 25 October 2005; received in revised form 4 May 2006; accepted 16 May 2006

Abstract

We examined the association of substance abuse treatment with access to liver specialty care among 231 persons coinfected with HIV and hepatitis C virus (HCV) with a history of alcohol problems who were recruited and followed up in the HIV-Longitudinal Interrelationships of Viruses and Ethanol cohort study from 2001 to 2004. Variables regarding demographics, substance use, health service use, clinical variables, and substance abuse treatment were from a standardized research questionnaire administered biannually. We defined substance abuse treatment services as any of the following in the previous 6 months: 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 any participation in a methadone maintenance program. Liver specialty care was defined as a visit to a liver doctor, a hepatologist, or a specialist in treating hepatitis C in the past 6 months. At study entry, most of the 231 subjects (89%, n = 205) had seen a primary care physician, 50% had been exposed to substance abuse treatment, and 50 subjects (22%) had received liver specialty care. An additional 33 subjects (14%) reported receiving liver specialty care during the follow-up period. In the multivariable model, we observed a clinically important although not statistically significant association between having been in substance abuse treatment and receiving liver specialty care (adjusted odds ratio = 1.38; 95% confidence interval = 0.9–2.11). Substance abuse treatment systems should give attention to the need of patients to receive care for prevalent treatable diseases such as HIV/HCV coinfection and facilitate its medical care to improve the quality of care for individuals with substance use disorders. The data illustrate the need for clinical care models that give explicit attention to the coordination of primary health care with addiction and hepatitis C specialty care while providing ongoing support to engage and retain these patients with complex health needs.

Keywords: Substance abuse; Hepatitis C virus; Liver specialty care; Substance abuse treatment

1. Introduction

The prevalence of hepatitis C virus (HCV) among injection drug users with HIV infection is close to 90% (Sulkowski & Thomas, 2003). It is well known that HIV coinfection accelerates the progression of HCV liver disease

* Corresponding author. 620B-1081 Burrard Street, Vancouver, BC, Canada V6Z 1Y6. Tel.: +1 604 682 2344x63194; fax: +1 604 806 8005.
E-mail address: anita@hivnet.ubc.ca (A. Palepu).

0740-5472/06/$ – see front matter
doi:10.1016/j.jsat.2006.05.015

(Sulkowski, Mast, Seeff, & Thomas, 2000), and a meta-analysis estimated that the risk of cirrhosis due to HIV coinfection is twofold (Graham et al., 2001). As HIV becomes a chronic disease due to the success of highly active antiretroviral therapy (Hogg et al., 1999; Palella et al., 1998), HCV liver disease is increasingly becoming a significant cause of morbidity and death among these individuals (Bica et al., 2001; Clanon, Mueller, & Harank, 2005; Monga et al., 2001). The current standard of care for chronic HCV infection is pegylated interferon and ribavirin, which can achieve sustained viral response of approximately 42–82%, depending on genotype. This is a substantial improvement over previous interferon monotherapy response rates of 6–20% (Desmond et al., 2006; National Institutes of Health [NIH], 2002). Unfortunately, the treatment given to persons coinfected with HIV/HCV is less effective (17–83%) at this time, but optimism about therapeutic options is warranted (Chung et al., 2004; Perez-Olmeda et al., 2003; Sulkowski, 2006). Because HCV treatment requires extensive evaluation and monitoring (Sulkowski, 2006), it is typically provided by specialists after primary care physician referral.

Active drug and alcohol use may pose challenges regarding the access to and the provision of HCV specialty care for individuals coinfected with HIV/HCV (Edlin et al., 2001). Heavy alcohol use is particularly harmful in the setting of chronic HCV infection, and it has been shown to reduce the success of HCV treatment (Corrao & Arico, 1998; Pessione et al., 1998; Thomas et al., 2000). In one study of patients infected with HCV in an opiate dependence treatment program, only 30% had been evaluated for HCV treatment and 34% were aware of HCV treatment underscoring the limited understanding of their disease and HCV treatment options (Walley, White, Kushel, Song, & Tulsky, 2005).

The 2002 NIH Consensus Statement on HCV infection recommended increased availability of HCV treatment to patients with high likelihood of acquiring or spreading this disease, such as injection drug users (NIH, 2002), but translating this to improved delivery of treatment is challenging. There has been increasing attention to HCV prevention and treatment among drug treatment programs, given the high prevalence among their clientele (Astone, Strauss, Hagan, & Jarlais, 2004; Munoz-Plaza, Strauss, Astone, Jarlais, & Hagan, 2004; Vassilev, Strauss, Astone, Friedmann, & Jarlais, 2004). From an overall health care management perspective, primary care providers may view patients coinfected with HIV/HCV with alcohol problems who are actively participating in substance abuse treatment as having their addiction issues addressed and, thus, more likely to be able to tolerate the HCV therapy and adhere to the close monitoring. This, in turn, may result in primary care physicians being more likely to refer such individuals to liver specialty care.

However, it is not clear whether HCV infection, a consequence of injection drug use, is effectively integrated into substance abuse treatment programs. To explore this issue, we tested the following hypothesis in an observational cohort of persons coinfected with HIV/HCV who have alcohol problems, most of whom with existing primary care: Participation in substance abuse treatment improves the likelihood of being evaluated by liver specialty care providers. In assessing this association of substance abuse treatment with one measure of desired quality health care (i.e., attention to HCV infection), we sought evidence of effective current collaboration of addiction treatment and general medical care.

2. Materials and methods

2.1. Study design and population

We analyzed data from participants of the HIV-Longitudinal Interrelationships of Viruses and Ethanol (LIVE) prospective cohort, which is an observational study of persons with HIV infection who have alcohol problems. Patients who were infected with HIV and had a history of alcohol problems were identified by explicit eligibility criteria: a documented HIV antibody by ELISA confirmed by Western blot; two or more affirmative responses to the Cut Down, Annoyed, Guilty, and Eye Opener (CAGE) alcohol screening questionnaire for lifetime alcohol abuse or dependence (Buchsbaum, Buchanan, Centor, Schnoll, & Lawton, 1991) or coinvestigator physician diagnosis of an alcohol disorder; ability to speak English or Spanish; and at least one contact person who was likely to know their whereabouts. The only exclusion criterion was a score of <21 on the 30-item Folstein Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) or trained interviewer assessment that the patient was incapable of comprehending informed consent or answering the interview questions.

From August 2001 to July 2003, subjects were recruited using multiple methods and from several sources. These included medical clinics responsible for the evaluation of patients with HIV infection presenting for medical care: the HIV Diagnostic Evaluation Unit at Boston Medical Center (BMC; Samet et al., 1995) and the HIV Consult Clinic and the Primary Care Clinic at Beth Israel Deaconess Medical Center (BIDMC). Other subjects were recruited from the following locations: BMC primary care practices, referrals by friends, and posted flyers at homeless shelters and HIV/AIDS social service agencies in the Boston area. Persons responding to the flyers were administered a preliminary screening over the telephone (CAGE questionnaire) and, if eligible, were invited for an interview to complete the screening process.

If a patient at these clinical sites or from other referral sources agreed to participate, a study research associate scheduled an appointment for the first interview at BMC's General Clinical Research Center or BIDMC's Clinical Research Center. All subjects provided written informed

consent in the Clinical Research Centers prior to enrollment. The Institutional Review Boards of BMC and BIDMC approved this study. A subject's privacy was made more secure through the issuance of a Certificate of Confidentiality by the Department of Health and Human Services; that is, it protects the release of a subject's research data even if a court order or subpoena is issued.

The HIV-LIVE study recruited 401 subjects. However, the current analyses focus on receipt of liver specialty care in just 231 (58%) subjects coinfected with HIV and HCV.

2.2. Data collection

After enrollment, all subjects received an interviewer-administered baseline assessment. The baseline instrument included questions on the following: demographics, Short Form Health Survey (Ware, Kosinski, & Keller, 1996), depressive symptoms (Center for Epidemiologic Studies Depression Scale [CES-D]; Andresen, Malmgren, Carter, & Patrick, 1994), psychological status (questions from the Addiction Severity Index; McLellan et al., 1985), health status, medication use, health care and addiction services use, alcohol and drug use quantity, current and lifetime alcohol abuse and dependence (Composite International Diagnostic Interview [CIDI]; Robins et al., 1988), HIV risk behaviors (modified version of the Risk Assessment Battery; Navaline et al., 1994), trauma history, social support, and social networks. All subjects in this cohort were tested for HCV infection through measurement of HCV antibody; HCV RNA testing was sought for all HCV-antibody-positive persons. Follow-up was conducted over 3 years at 6-month intervals and included a reassessment of the domains covered at baseline. Subjects received a cash compensation of US$20 at baseline and US$25–30 at the follow-up time points. Results from follow-up assessments as of February 25, 2004, were utilized in these analyses.

2.3. Outcome variable

Receipt of liver specialty care in the previous 6 months was modeled as a dichotomous response. Typically, primary care physicians provide referrals for liver specialty care. Subjects were asked "How many times did you see each of the following health care professionals during the past six months?" Receipt of liver specialty care was defined as having an affirmative response to being seen by a "liver doctor, hepatologist, or specialist in treating Hepatitis C."

2.4. Primary independent variable

Substance abuse treatment services was a dichotomous variable indicating whether the subject received any of the following services in the past 6 months: at least 12 weeks in a halfway 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 any participation in a methadone maintenance program (Palepu, Horton, Tibbetts, Meli, & Samet, 2005; Palepu, Raj, et al., 2005; Palepu et al., 2004). Information on substance abuse treatment services was obtained from patient self-report at each study interview.

2.5. Other independent variables

Other specific variables assessed included age; gender; race (Black, White, or other—the latter subjects were mostly Hispanic); an indicator of whether the subject had participated in a previous cohort study of subjects with HIV infection who have alcohol problems; depressive symptoms (yes vs. no) as measured by the 20-item CES-D (Andresen et al., 1994), where a higher cutoff score of ≥ 21 was used to denote depressive symptoms in persons with chronic diseases rather than the standard score of 16 that has been used for the general population; any liver complications (i.e., jaundice, ascites, esophageal varices, hepatic encephalopathy, or gastrointestinal bleeding) in the previous 6 months; current receipt of highly active antiretroviral therapy; ever been treated for psychiatric disorders; drug injection in the past 6 months; abstinent from alcohol in the past 30 days; alcohol dependence in the past 6 months as defined by the CIDI short form (Robbins et al., 2003); average alanine aminotransferase (ALT); the number of months since baseline time point; and CD4 cell count.

2.6. Analysis

Two-sample t tests and chi-square tests were used to assess bivariable relationships between demographics, behavioral and clinical data, and the outcome accessing liver specialty care at baseline. We used generalized estimating equations (GEE) logistic regression models to examine the association between substance abuse treatment and accessing liver specialty care, adjusting for potential confounding factors: gender, age, race, liver complications, current receipt of antiretroviral therapy, depressive symptoms, ever been treated for psychiatric disorders, drug injection in the past 6 months, 30-day alcohol abstinence, recent alcohol dependence and ALT level, CD4 cell count, rollover, and the number of months since baseline time point. The GEE approach was used to adjust for the correlation due to analyzing repeated measures from the same subject over time (Liang & Zeger, 1986; Zeger & Liang, 1986). The empirical standard errors from the GEE approach were used for all analyses. All of the predictor variables except for gender, age, race, rollover, and treatment for psychiatric disorders were allowed to vary with time. We examined the potential for collinearity in multivariate models by assessing the correlation between pairs of independent variables and verified that no pair of variables included in the same regression model was highly correlated (i.e., no correlation was greater than .40). Although the outcome and most covariates were assessed semiannually,

ALT was collected once yearly. Thus, we used the average ALT value of the most recent and subsequent measures to impute values for missing time points. All analyses were carried out using SAS version 8.2 (SAS Institute, Cary, NC).

3. Results

Of the 231 subjects coinfected with HIV/HCV in the study sample, half (116/231) were engaged in substance abuse treatment at the initial observation, of whom 47 (20%) were in a methadone treatment program. Among the 115 subjects who were not engaged in substance abuse treatment at the initial observation, 39 (34%) subjects entered substance abuse treatment during the study period.

There were 50 (22%) subjects who received liver specialty care and 205 (89%) subjects who reported having seen a primary care physician at the baseline interview. An additional 33 (14%) subjects reported receiving liver specialty care during the 3-year follow-up period. In this research study, the subjects were followed up every 6 months for up to six visits, and the median number of observations per subject was 3. For the current analysis, the proportion of subjects who completed one, two, three, four, five, and six observations during the 3-year follow-up period was 5%, 17.2%, 21.1%, 24.7%, 30.2%, and 1.7%, respectively. Overall, the 231 subjects contributed 696 observations to the longitudinal analysis; however, due to incomplete data, 76 subjects were dropped.

The baseline characteristics of the 231 subjects coinfected with HIV/HCV are presented in Table 1. In the bivariable analyses based on the baseline assessments only, subjects who were currently receiving highly active antiretroviral therapy, who had liver complications in the previous 6 months, and who were abstinent from alcohol in the past 30 days were more likely to have accessed liver specialty care.

In the longitudinal multivariable model (Table 2), we observed a clinically important although not statistically significant association between substance abuse treatment and accessing liver specialty care (adjusted odds ratio [OR] = 1.38; 95% confidence interval [CI] = 0.9–2.11). Liver complications (adjusted OR = 2.15; 95% CI = 1.02–4.54), current receipt of highly active antiretroviral therapy (adjusted OR = 2.18; 95% CI = 1.33–3.59), CD4 cell count per 100 cells/mm^3 increase (adjusted OR = 1.11; 95% CI = 1.02–1.20), and 30-day alcohol abstinence (adjusted

Table 1
Characteristics and access to liver specialty care at baseline of participants with HIV and HCV coinfection

Baseline characteristics	Access to liver specialty care: With access, n = 50	Without access, n = 181	p
Age, *M* (*SD*)	45.0 (6.3)	43.5 (7.0)	.14
Female gender, *n* (%)	14 (28)	52 (29)	.92
Race or ethnicity, *n* (%)			.33
Black	17 (34)	70 (39)	
White	21 (42)	56 (31)	
Other	12 (24)	55 (30)	
Substance abuse treatment, *n* (%)	29 (58)	87 (48)	.21
Current ART receipt[a], *n* (%)	40 (80)	102 (56)	.002
Baseline CD4 cell count (cells/mm^3), *M* (*SD*)	423 (253)	408 (260)	.86
Liver disease complications[b,c], *n* (%)	9 (18)	6 (3)	.0002
Elevated ALT (>40 IU/L), *n* (%)	36 (75)	120 (68)	.39
Depressive symptoms[d], *n* (%)	29 (58)	110 (61)	.72
Ever had psychiatric treatment, *n* (%)	35 (70)	119 (66)	.57
Injection drug use[b], *n* (%)	7 (14)	45 (25)	.10
30-day alcohol abstinence, *n* (%)	38 (76)	107 (59)	.03
Alcohol dependence[b,e], *n* (%)	2 (4)	21 (12)	.11

[a] Highly active antiretroviral therapy.
[b] In the past 6 months.
[c] Any liver complications refer to jaundice, ascites, hepatic encephalopathy, esophageal varices, or gastrointestinal bleeding.
[d] Using the CES-D, where ≥21 denotes depressive symptoms in chronic diseases.
[e] Alcohol dependence as defined by the CIDI short form.

Table 2
Multivariable logistic regression model for factors associated with accessing liver specialty care (using generalized estimating equations)

Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Substance abuse treatment	1.42 (0.99–2.03)	1.38 (0.90–2.11)
Female gender	0.95 (0.53–1.7)	0.87 (0.44–1.70)
Age (per 10-year increase)	1.37 (0.95–1.97)	1.22 (0.80–1.86)
Race or ethnicity[a]		
Black vs. White	0.77 (0.43–1.39)	0.70 (0.36–1.36)
Other vs. White	0.82 (0.44–1.52)	0.73 (0.38–1.41)
Current ART receipt[b]	2.50 (1.58–3.93)	2.18 (1.33–3.59)
CD4 (per 100 cells/mm^3 increase)	1.12 (1.03–1.21)	1.11 (1.02–1.20)
Liver disease complications[c,d]	2.09 (1.18–3.72)	2.15 (1.02–4.54)
ALT (per 40 IU/L)	1.15 (0.98–1.35)	1.10 (0.90–1.33)
Depressive symptoms[e]	0.87 (0.59–1.28)	0.81 (0.51–1.27)
Ever had psychiatric treatment	1.42 (0.82–2.45)	1.21 (0.65–2.25)
Injection drug use[c]	0.78 (0.49–1.24)	0.84 (0.48–1.45)
30-day alcohol abstinence	1.67 (1.17–2.38)	1.55 (1.04–2.32)
Alcohol dependence[c,f]	0.62 (0.35–1.08)	0.72 (0.35–1.49)

Note. The number of observations is 620.
[a] The reference group is White.
[b] Highly active antiretroviral therapy.
[c] In the past 6 months.
[d] Any liver complications refer to jaundice, ascites, hepatic encephalopathy, esophageal varices, or gastrointestinal bleeding.
[e] Using the CES-D, where ≥21 denotes depressive symptoms in chronic diseases.
[f] Alcohol dependence as defined by the CIDI short form.

OR = 1.55; 95% CI = 1.04–2.32) were positively associated with accessing liver specialty care.

4. Discussion

Substance abuse treatment can be an effective means of HCV education, facilitated access, or linkage and may be a motivating factor for seeking out HCV care for some patients (Strauss, Astone, Des Jarlais, & Hagan, 2005; Strauss, Astone, Hagan, & De Jarlais, 2004; Strauss, Astone, Jarlais, & Hagan, 2004). However, in our longitudinal study of persons coinfected with HIV/HCV who have alcohol problems, of whom some were receiving community-based substance abuse treatment, we were not able to detect a significant association between substance abuse treatment and receipt of specialty care for hepatitis C. We found that the presence of significant liver disease or factors associated with HCV treatment eligibility, such as tolerating antiretroviral therapy, having a higher CD4 cell count, and being recently abstinent from alcohol, were the factors associated with receipt of liver specialty care. In our system, referrals were likely through the primary care physician where initiation and maintenance of ART for HIV infection may have been the first step in addressing the complex health needs of these patients. Among HCV monoinfected injection drug users, Strathdee et al. (2005) found that having a usual source of primary care, a high perceived threat of progressive liver disease, no evidence of alcohol dependence, and higher readiness scores for quitting drug use were all factors associated with being interested in HCV treatment. Assessing the specific impact of substance abuse treatment on receipt of liver specialty care may be difficult, given that the receipt of addiction treatment could be interpreted as a barrier to HCV therapy. Substance abuse treatment may be indicative of recent drug and alcohol use, and yet, it may also be a motivator or facilitator for patients and their primary care provider to seek further care for the patients' comorbid medical conditions. Substance abuse treatment, as a pathway to remission of substance abuse, could set the stage for addressing previously unattended chronic medical conditions. Our findings that receipt of ART and 30-day alcohol abstinence are associated with accessing liver specialty care likely reflect the hierarchy of priorities primary care providers face in managing these patients with complicated health needs. Such priorities are supported by current NIH and international guidelines (Soriano et al., 2002), where it is recommended that HIV and substance abuse are addressed prior to initiating HCV therapy.

The lack of coordination of medical care and substance abuse treatment has been cited as the most significant barrier to HCV treatment for persons with substance use disorders, rather than the more typical barriers to medical care such as health insurance or transportation (Litwin, Soloway, & Gourevitch, 2005). Litwin et al. recently described a multidisciplinary model of care that addressed substance abuse and psychiatric conditions, as well as HCV screening and treatment. This approach resulted in substantial rates of initiation of antiviral therapy, and the colocation of these services significantly improved HCV treatment access (Litwin et al., 2005). Clearly, methadone maintenance and other substance abuse treatment programs can be gateways to enhancing access to HCV treatment through screening and education (Walley et al., 2005). A few integrated models of care that address HIV, HCV, and substance abuse have been recently reported (Clanon et al., 2005; Flanigan, Taylor, & Mitty, 2005; Fleming, Tumility, Murray, & Nues, 2005; Sylvestre, 2005; Taylor, 2005). These studies highlight the importance of a multidisciplinary team approach to such patients with complex health needs. In one clinic setting, one third of the patients coinfected with HIV/HCV were eligible for treatment (Fleming et al., 2005). Reasons for ineligibility include nonadherence with clinic visits, active psychiatric disease and ongoing drug and alcohol use, advanced HIV disease, decompensated liver disease, and significant comorbid illness. Furthermore, two thirds of those who were eligible declined HCV treatment with interferon and ribivirin. In sum, only 8% (21/260) were treated for their HCV and two patients achieved sustained virological response. These studies highlight the importance of addressing the modifiable barriers to HCV treatment eligibility and adherence. Given the significant overlap of HCV and addictions, more coordination and integration of these treatment services would be desirable to address the burden of HCV–liver disease in this vulnerable population.

Our study has several limitations. We assessed HCV infection solely based on HCV antibody test and not HCV RNA (viral load). There were 10% to 15% of the HCV-antibody-positive patients who did not have detectable HCV RNA, and when we fitted the models with those who were positive for HCV RNA, the findings were unchanged. However, our study sample included all HCV-antibody-positive patients, as substance abuse treatment providers would not be expected to ask such a medical question and, in fact, some patients might not be aware of the HCV antigen status. Although our measure of substance abuse treatment may not be as stringent as that used by Laine et al. (2001), we think that it has face validity. Approximately half of our cohort was receiving substance abuse treatment services at a reasonable level of exposure. Our main explanatory and outcome measures were obtained through self-report, although validated instruments were used where possible. We were unable to explore the precise reasons for why subjects with HIV/HCV coinfection did or did not receive liver specialty care in terms of source of referral, patient refusal, or specialist refusal. The study was observational and we cannot infer causation. This analysis was also potentially underpowered to detect a statistically significant effect of substance abuse treatment. For post hoc power calculations, assuming 18% of the subjects not receiving substance abuse treatment accessed liver specialty

care (based on data from study entry), our study would have approximately 80% power to detect an OR as small as 2.0. Thus, it is likely that the study was not adequately powered to detect a small association of the observed magnitude. Finally, these findings may not be generalizable to health care systems that have differing availability of liver specialty care.

In summary, we did not observe a clear association between substance abuse treatment and receipt of liver specialty care among persons coinfected with HIV/HCV who have a history of alcohol problems. The current system of addiction treatment in our study appears to be isolated from primary care as well as liver specialty care and does not utilize the substance abuse treatment opportunity to motivate the patient to address other chronic medical problems. The patients who reported receiving liver specialty care were persons who were having their HIV disease managed, had liver disease complications, and were not actively using alcohol. These are the types of patients primary care providers would likely deem as having a higher likelihood of HCV treatment eligibility, need, or success. Substance abuse treatment systems should prioritize prevalent treatable diseases such as HIV/HCV coinfection and facilitate medical care to improve the quality of care for individuals with substance use disorders. Shared care models that give explicit attention to the coordination of addictions, primary care, and specialty care, while providing ongoing support to effectively engage these patients with complex health needs in appropriate care, need to be implemented and evaluated.

Acknowledgments

The authors appreciate the contributions of the research associates on the project (Rachel Levison, Elizabeth Curley, and Eileen O'Connor) and data management assistance (Vincent Faber). Support for this study came from the following grants from the National Institute of Alcohol Abuse and Alcoholism of the NIH: RO1-AA13216 (Clinical Impact of HCV and Alcohol in HIV-Infected Persons), 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 MO1 RR00533) and in the General Clinical Research Center at BIDMC (USPHS Grant MO1 RR01032). Anita Palepu is supported by a Senior Scholar award from the Michael Smith Foundation for Health Research.

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.

Astone, J. M., Strauss, S. M., Hagan, H., & Jarlais, D. C. D. (2004). Outpatient drug treatment program directors' hepatitis C-related beliefs and their relationship to the provision of HCV services. *American Journal of Drug and Alcohol Abuse, 30*, 783–797.

Bica, I., McGovern, B., Dhar, R., Stone, D., McGowan, K., Scheib, R., et al. (2001). Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clinical Infectious Diseases, 32*, 492–497.

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.

Chung, R. T., Andersen, J., Volberding, P., Robbins, G. K., Liu, T., Sherman, K. E., et al. (2004). Peginterferon alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. *New England Journal of Medicine, 351*, 451–459.

Clanon, K. A., Mueller, J. J., & Harank, M. (2005). Integrating treatment for hepatitis C virus infection into an HIV clinic. *Clinical Infectious Diseases, 40*, S362–S366.

Corrao, G., & Arico, S. (1998). Independent and combined action of hepatitis C virus infection and alcohol consumption on the risk of symptomatic liver cirrhosis. *Hepatology, 27*, 914–919.

Desmond, C. P., Roberts, S. K., Dudley, F., Mitchell, J., Day, C., Nguyen, S., et al. (2006). Sustained virological response rates and durability of the response to interferon-based therapies in hepatitis C patients treated in the clinical setting. *Journal of Viral Hepatitis, 13*, 311–315.

Edlin, B. R., Seal, K. H., Lorvick, J., Kral, A. H., Ciccarone, D. H., Moore, L. D., et al. (2001). Is it justifiable to withhold treatment for hepatitis C from illicit-drug users? *New England Journal of Medicine, 345*, 211–214.

Flanigan, T. P., Taylor, L. E., & Mitty, J. A. (2005). Use of community-based, directly observed therapy for HIV infection: Lessons learned for treatment of hepatitis C virus infection. *Clinical Infectious Diseases, 40*, S346–S348.

Fleming, C. A., Tumilty, S., Murray, J. E., & Nunes, D. (2005). Challenges in the treatment of patients co-infected with HIV and hepatitis C virus: Need for team care. *Clinical Infectious Diseases, 40*, S349–S354.

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.

Graham, C. S., Baden, L. R., Yu, E., Mrus, J. M., Carnie, J., Heeren, T., et al. (2001). Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: A meta-analysis. *Clinical Infectious Diseases, 33*, 562–569.

Hogg, R. S., Yip, B., Kully, C., Craib, K. J., O'Shaughnessy, M. V., Schechter, M. T., et al. (1999). Improved survival among HIV-infected patients after initiation of triple-drug antiretroviral regimens. *Canadian Medical Association Journal, 160*, 659–665.

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.

Liang, K. Y., & Zeger, S. Y. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

Litwin, A. H., Soloway, I., & Gourevitch, M. N. (2005). Integrating services for injection drug users infected with hepatitis C virus with methadone maintenance treatment: Challenges and opportunities. *Clinical Infectious Diseases, 40*, S339–S345.

McLellan, A. T., Cassiola, J., Griffith, J., Evans, F., Bass, H. L., & O'Brien, C. P. (1985). New data from the addiction severity index: Reliability and validity in three centers. *Journal of Nervous and Mental Disease, 173*, 412–423.

Monga, H. K., Rodriguez-Barradas, M. C., Breaux, K., Khattak, K., Troisi, C. L., Velez, M., et al. (2001). Hepatitis C virus infection—Related morbidity and mortality among patients with human immunodeficiency virus infection. *Clinical Infectious Diseases, 33*, 240–247.

Munoz-Plaza, C. E., Strauss, S. M., Astone, J. M., Jarlais, D. C., & Hagan, H. (2004). Drug treatment programs as sites of opportunity for the delivery of hepatitis C prevention education: Client and staff perspectives. *Journal of Drug Issues, 34*, 861–878.

National Institutes of Health. (2002). NIH consensus development conference statement: Management of hepatitis C. *Gastroenterology, 123*, 2082–2099.

Navaline, H. A., Snider, E. C., Petro, C. J., Tobin, D., Metzger, D., Alterman, A. I., et al. (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.

Palella, F. J., Jr., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., et al. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV outpatient study investigators. *New England Journal of Medicine, 338*, 853–860.

Palepu, A., Horton, N. J., Tibbetts, N., Meli, S., & Samet, J. H. (2005). Substance abuse treatment and hospitalization among a cohort of HIV-infected individuals with alcohol problems. *Alcoholism: Clinical and Experimental Research, 29*, 389–394.

Palepu, A., Raj, A., Horton, N. J., Tibbetts, N., Meli, S., & Samet, J. H. (2005). Substance abuse treatment and risk behaviors among HIV-infected persons with alcohol problems. *Journal of Substance Abuse Treatment, 28*, 3–9.

Palepu, A., Tyndall, M. W., Chan, K., Wood, E., Montaner, J. S., & Hogg, R. S. (2004). Initiating highly active antiretroviral therapy and continuity of HIV care: The impact of incarceration and prison release on adherence and HIV treatment outcomes. *Antiviral Therapy, 9*, 713–719.

Perez-Olmeda, M., Nunez, M., Romero, M., Gonzalez, J., Castro, A., Arribas, J. R., et al. (2003). Pegylated ifn-alpha 2b plus ribavirin as therapy for chronic hepatitis C in HIV-infected patients. *AIDS, 17*, 1023–1028.

Pessione, F., Degos, F., Marcellin, P., Duchatelle, V., Njapoum, C., Martinot-Peignoux, M., et al. (1998). Effect of alcohol consumption on serum hepatitis C virus RNA and histological lesions in chronic hepatitis C. *Hepatology, 27*, 1717–1722.

Robbins, G. K., De Gruttola, V., Shafer, R. W., Smeaton, L. M., Snyder, S. W., Pettinelli, C., et al. (2003). Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. *New England Journal of Medicine, 349*, 2293–2303.

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

Samet, J. H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D. E., et al. (1995). A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Archives of Internal Medicine, 155*, 1629–1633.

Soriano, V., Sulkowski, M., Bergin, C., Hatzakis, A., Cacoub, P., Katlama, C., et al. (2002). Care of patients with chronic hepatitis C and HIV co-infection: Recommendations from the HIV–HCV international panel. *AIDS, 16*, 813–828.

Strathdee, S. A., Latka, M., Campbell, J., O'Driscoll, P. T., Golub, E. T., Kapadia, F., et al. (2005). Factors associated with interest in initiating treatment for hepatitis C virus (HCV) infection among young HCV-infected injection drug users. *Clinical Infectious Diseases, 40*, S304–S312.

Strauss, S. M., Astone, J. M., Des Jarlais, D. C., & Hagan, H. (2005). Integrating hepatitis C services into existing HIV services: The experiences of a sample of US drug treatment units. *AIDS Patient Care and STDs, 19*, 78–88.

Strauss, S. M., Astone, J. M., Hagan, H., & Des Jarlais, D. C. (2004). The content and comprehensiveness of hepatitis C education in methadone maintenance and drug-free treatment units. *Journal of Urban Health, 81*, 38–47.

Strauss, S. M., Astone, J. M., Jarlais, D. C., & Hagan, H. (2004). A comparison of HCV antibody testing in drug-free and methadone maintenance treatment programs in the United States. *Drug and Alcohol Dependence, 73*, 227–236.

Sulkowski, M. S. (2006). Treatment algorithm for the management of hepatitis C in HIV-coinfected persons. *Journal of Hepatology, 44* (1 Suppl), S49–S55.

Sulkowski, M. S., Mast, E. E., Seeff, L. B., & Thomas, D. L. (2000). Hepatitis C virus infection as an opportunistic disease in persons infected with human immunodeficiency virus. *Clinical Infectious Diseases, 30*, S77–S84.

Sulkowski, M. S., & Thomas, D. L. (2003). Hepatitis C in the HIV-infected person. *Annals of Internal Medicine, 138*, 197–207.

Sylvestre, D. L. (2005). Treating hepatitis C virus infection in active substance users. *Clinical Infectious Diseases, 40*, S321–S324.

Taylor, L. E. (2005). Delivering care to injection drug users co-infected with HIV and hepatitis C virus. *Clinical Infectious Diseases, 40*, S355–S361.

Thomas, D. L., Astemborski, J., Rai, R. M., Anania, F. A., Schaeffer, M., Galai, N., et al. (2000). The natural history of hepatitis C virus infection: Host, viral, and environmental factors. *Journal of the American Medical Association, 284*, 450–456.

Vassilev, Z. P., Strauss, S. M., Astone, J. M., Friedmann, P. D., & Jarlais, D. C. D. (2004). Provision of on-site medical care to patients with hepatitis C in drug treatment units. *Journal of Health Care for the Poor and Underserved, 15*, 663–671.

Walley, A. Y., White, M. C., Kushel, M. B., Song, Y. S., & Tulsky, J. P. (2005). Knowledge of and interest in hepatitis C treatment at a methadone clinic. *Journal of Substance Abuse Treatment, 28*, 181–187.

Ware, J., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*, 220–233.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121–130.

AIDS and Behavior, Vol. 10, No. 2, March 2006 (© 2006)
DOI: 10.1007/s10461-005-9050-x

Sex Trade, Sexual Risk, and Nondisclosure of HIV Serostatus: Findings from HIV-Infected Persons with a History of Alcohol Problems

Anita Raj,[1,4] Debbie M. Cheng,[2] Rachel Levison,[3] Seville Meli,[3] and Jeffrey H. Samet[1,3]

Published online: Feb. 16, 2006

The purpose of this study was to assess the relationships between disclosure of HIV serostatus to sex partners and recent sexual risk behavior, substance abuse, and violence among sexually active HIV-infected patients with a history of alcohol problems. Participants ($n = 124$) were 79% males; age 25–61 years; 49% Black; and 35% with less than a high school education. Separate logistic regression models were used to assess relationships between each independent variable of interest and nondisclosure. Results demonstrate that buying sex and having more than one sex partner in the past 6 months were significantly associated with nondisclosure of HIV serostatus to a sex partner. Findings from this study underscore the ongoing need for behavioral interventions with HIV-infected individuals concerning disclosure. Programs that emphasize serostatus disclosure and/or consistent condom use in the context of sex trade and with multiple sexual partners will be particularly important.

KEY WORDS: HIV; serostatus disclosure; sex risk; sex trade; drug risk behaviors; violence.

INTRODUCTION

Prevention targeting people living with HIV/AIDS is important in reducing the spread of HIV in the United States (Jaffe and Janssen, 2003). Prevention efforts include promotion of sexual risk reduction behaviors such as condom use and serostatus disclosure to sex partners (Centers for Disease Control and Prevention [CDC], 2003). The need for such efforts is supported by research documenting that the majority of people living with HIV are sexually active (Stein *et al.*, 1998) and many of these individuals maintain high risk sexual behaviors including unprotected sex and multiple sex partners (Kalichman and Nachimson, 1999; Kline and VanLandingham, 1994; Marks and Crepaz, 2001; McGowan *et al.*, 2004; Stein *et al.*, 1998). These high-risk sexual behaviors are particularly common among seropositive individuals who are substance users (Kline and VanLandingham, 1994; Marks and Crepaz, 2001; Vanable *et al.*, 2001) and/or involved in sex trade (McGowan *et al.*, 2004).

Studies have shown HIV serostatus disclosure to sexual partners is inconsistent across population groups (Hays *et al.*, 1993; Mansergh *et al.*, 1995; Marks *et al.*, 1991, 1994; Niccolai *et al.*, 1999; Perry *et al.*, 1994; Schnell *et al.*, 1992; Simoni *et al.*, 1995a,b; Sowell *et al.*, 1997; Stein *et al.*, 1998; Stempel *et al.*, 1995; Wenger *et al.*, 1994). U.S. laws often make nondisclosure of HIV serostatus to sex partners a felony (Wolf and Vezina, 2004), but these laws do not appear to be associated with greater disclosure (Duru *et al.*, 2003). More work is needed to identify factors associated with nondisclosure to better support growing prevention efforts.

[1]Department of Social Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts.
[2]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts.
[3]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston, Massachusetts.
[4]Correspondence should be directed to Anita Raj, Ph.D., Boston University School of Public Health, 715 Albany St, T2W, Boston, Massachusetts 02118; e-mail: anitaraj@bu.edu.

Consistent condom use among serodiscordant couples has been found to be associated with lower rates of seroconversion for the HIV-negative partner (Weller and Davis, 2002). However, unprotected sex combined with nondisclosure of serostatus to sex partners can result in HIV transmission to those denied the opportunity to make an informed decision regarding use of protection. Nondisclosure and unprotected sex may be particularly likely in the context of a more casual sexual relationship, substance abuse, and/or violence, where one might feel less obligated, able and/or safe to disclose to sex partners (Bedimo *et al.*, 1998; Klitzman and Bayer, 2003). Research has documented increased sexual risk for HIV among those at the intersection of sex trade, substance abuse and violence, where lack of control of one's behavior or sexual encounter can increase the likelihood of unprotected sex and multiple sex partners (e.g., Brown, 1998; Pisani *et al.*, 2003; Robertson *et al.*, 2004; Romero-Daza *et al.*, 2003; Surratt and Inciardi, 2004; Vanwesenbeeck, 2001; Weber *et al.*, 2001). Research has also shown that sex trade, substance use, violence, unprotected sex, and STD continue subsequent to infection for many people living with HIV (Cohen *et al.*, 2000; McGowan *et al.*, 2004; Turner *et al.*, 2001).

The literature shows mixed results with regard to whether serostatus disclosure is associated with condom use. Niccolai *et al.*'s (1999) study showed a significant relationship between condom use and disclosure for both men and women. Kalichman and Nachimson (1999) found similar findings for men but not for women, and other studies have shown no relationship between disclosure and condom use (Marks and Crepaz, 2001; Stein *et al.*, 1998). In contrast, research has consistently demonstrated that HIV-infected people are more likely to disclose their serostatus to longer term/committed partners as compared with casual partners (Duru *et al.*, 2003; Perry *et al.*, 1994), and disclosure to all sex partners is significantly more common among those with fewer sex partners (Niccolai *et al.*, 1999; Stein *et al.*, 1998). Although sex trade is associated with more and casual sexual partners (Vanwesenbeeck, 2001) as well as less consistent condom use among people living with HIV (McGowan *et al.*, 2004), it has not been examined in terms of its association with disclosure. Given the gender and power differences between those buying as compared with selling sex (Pitts *et al.*, 2004; Vanwesenbeeck, 2001), buying and selling sex must be assessed separately in terms of its relation to sexual risk and disclosure.

Substance abuse is a major issue for HIV-positive individuals; a population-based probability sample of HIV-positive individuals in care found that 46% reported drug abuse in the past year and 36% reported binge alcohol use in the past 4 weeks (Turner *et al.*, 2001). Among HIV-positive samples, those with substance abuse problems are more likely to report unsafe sex practices, such as multiple sex partners, a characteristic related to nondisclosure (Duru *et al.*, 2003; Marks and Crepaz, 2001; Kalichman and Nachimson, 1999; Kline and Van-Landingham, 1994; Stein *et al.*, 1998). Additionally, cognitive impairment due to substance abuse could impede ability to disclose. However, prior research has consistently demonstrated no relationship between current substance use and disclosure (Duru *et al.*, 2003; Kalichman and Nachimson, 1999; Perry *et al.*, 1994; Stein *et al.*, 1998). Nonetheless, as previous research has not been restricted to a population with known substance use problems, there remains a need to ascertain whether this relationship exists within an HIV-infected cohort with a history of alcohol problems.

Studies demonstrate that HIV-infected individuals, particularly women, are significantly more likely than those uninfected to have experienced violence and related trauma in their lives, including physical and sexual victimization (Bedimo *et al.*, 1997; Cohen *et al.*, 2000; Department of Justice [DOJ], 2004; Gielen *et al.*, 2000; Liebschutz *et al.*, 2000; Sowell *et al.*, 1999; Zierler *et al.*, 2000). Some researchers have suggested that disclosure may result in violence from the partner or others (Rothenberg and Paskey, 1995), inhibiting individuals' willingness to disclose. Although Stein *et al.* (1998) found that among the HIV-infected women, those disclosing to all sex partners were significantly more likely to have had a history of victimization, this violence was not assessed for its temporal relationship to disclosure. Duru *et al.* (2003) found no relationship between disclosure and partner violence.

Understanding why disclosure to sex partners does or does not occur is an important step in promoting disclosure, as well as sexual risk reduction, among people living with HIV. Overall, our review of the literature points to the need for examination of correlates of nondisclosure for HIV-infected individuals with substance abuse problems. Thus, the purpose of this study was to assess the relationships between disclosure of HIV serostatus to sex partners and behavioral risk (i.e., unsafe sex behavior, substance abuse, and recent exposure to violence) in a

cohort of HIV-infected patients with a history of alcohol problems.

METHODS

Study participants were from the HIV-Alcohol Longitudinal Cohort (HIV-ALC), a prospective, observational study of HIV-infected patients with past or current history of alcohol problems. HIV-ALC participants ($N = 349$) were recruited and followed-up between July 1997 and July 2001; participants were followed every 6 months for up to seven observations (median number of observations in sample = 2; Ehrenstein *et al.*, 2004). The eligible HIV-ALC participants were also participants of the ADHERE randomized controlled trial (ADHERE RCT), an evaluation of an intervention to increase medication adherence among HIV-infected people with a history of alcohol problems. Data for the current analyses were from the third observation point of HIV-ALC participants and included ADHERE RCT study participants; this observation was the first time point in which disclosure of serostatus to sex partners was assessed. Only participants reporting sexual activity within the past 6 months and responding to the disclosure question at this time point were included in the current cross-sectional analyses ($n = 124$). Of these, 32% were not in the ADHERE RCT, 31% were in the treatment group of the ADHERE RCT, and 37% were in the ADHERE RCT control group.

Recruitment and Eligibility

Recruitment of HIV-ALC participants occurred by multiple methods and from several sites: Boston Medical Center HIV Diagnostic Evaluation Unit 56%; posted flyers 16%; Boston Medical Center Primary Care Clinic 13%; respite facility for homeless persons 5%; methadone clinic 4%; participant referrals 4%; and Beth Israel Deaconess Medical Center 2%. The majority were recruited from medical settings that addressed HIV-related issues.

All HIV-ALC participants were required to be HIV-infected, age 18 years or older, and with a history of alcohol problems. A history of alcohol problems was defined as having two or more positive responses to the CAGE questionnaire (Ewing, 1984), or, if the patient was recruited from the Boston Medical Center HIV Diagnostic Evaluation Unit and did not meet CAGE criteria, having a diagnosis of alcohol abuse or dependence from one of two attending physicians. Ninety percent of participants (313/349) met CAGE eligibility, and 10% (36/349) met eligibility based on clinical assessment. Diagnostic interviews for alcohol problems in a sample of participants meeting CAGE criteria ($N = 141$) confirmed that alcohol was a problem for these individuals in that it revealed a lifetime history of alcohol dependence in 80% (113/141) or abuse in 15% (21/141) of cases (Samet *et al.*, 2004).

Other entry criteria for the HIV-ALC study included the following: fluency in English or Spanish; Mini-Mental State Examination (MMSE) score greater or equal to 21 (Folstein *et al.*, 1975); and no plans to move from the Boston area in the subsequent two years. As chronic alcohol use is associated with cognitive impairment, we used the MMSE cut-off of 21 to exclude participants in whom such impairment may preclude obtaining informed consent, an accurate and complete interviewer-administered questionnaire, or a follow-up interview (Smith *et al.*, 2003).

Procedure

After screening for eligibility and obtaining informed consent, a research associate interviewed participants in either English or Spanish using a standardized instrument to ascertain baseline information on demographics, health and medical care, HIV risk behaviors, and drug and alcohol use. 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. Participants were then followed-up with surveys every 6 months for up to 36 months. Although biological data were not included in current analyses; they were collected for use in this study. Attempts were made to obtain CD4 cell counts and HIV RNA (viral load) levels from all participants. These data were obtained via laboratory tests performed within 6 months of the interview, as part of patients' clinical care; if these tests were not available, blood samples were obtained from participants and tested for CD4 cell count and HIV RNA using the Boston Medical Center Clinical Laboratory.

Measures

Demographic characteristics, including age, gender, race, and sexual orientation (dichotomized as

gay/lesbian/bisexual vs. heterosexual) were measured via single items.

Nondisclosure of Serostatus to sex partners was measured via a single item created for use in this survey, "Have you told any of the following people that you are HIV-infected?" For the response related to sexual partners of the past 6 months, data were dichotomized as nondisclosure to one or more sex partners of the past 6 months versus disclosure to all sex partners of the past 6 months.

Sexual Risk Variables were measured via items from the Risk Assessment Battery (RAB), which is designed to measure sexual and drug use risk for HIV (Navaline *et al.*, 1994) and has been used with substance using and HIV-infected populations in previous studies (Ehrenstein *et al.*, 2004; Rees *et al.*, 2001). Only sexual risk items from the RAB were included in analyses. All items assessed risk in the past 6 months and were dichotomized. These items included inconsistent/no condom use (vs. consistent condom use), number of sexual partners (one vs. two or more), selling sex for money or drugs, and buying sex with money or drugs. Additionally, STD in the past 6 months was measured via self-report.

Substance Abuse in this study was defined as binge alcohol consumption or illicit drug use in the past 30 days. Both of these variables were calculated using the alcohol and drug use frequency questions from the Addiction Severity Index (ASI; McLellan *et al.*, 1992). Illicit drug use was defined as having used heroin, other opiates/analgesics, barbiturates, cocaine, or amphetamines in the past 30 days. Binge alcohol consumption in the past 30 days was defined as having five or more drinks on an occasion for males or four or more drinks on an occasion for females.

Victimization from interpersonal violence within the past 6 months was measured via two yes/no items, one assessing physical abuse and the other assessing sexual abuse. Physical abuse or assault was defined as having been "kicked, hit, choked, shot, stabbed, burned, or held at gunpoint by a stranger, a family member or someone you know." And sexual assault was defined as "unwanted sexual touching anywhere on your body, touching of genitals and/or breast, or made to have oral sex or vaginal or anal intercourse against your will by force or the threat of force by a stranger, a family member or someone you know" (Liebschutz *et al.*, 2000).

Data Analyses

Two sample t tests and chi-square tests were used to assess bivariate relationships between demographics and disclosure. Separate logistic regression analyses were performed to evaluate the associations between each independent variable of interest (sexual risk behaviors, STD, victimization, and substance use) and disclosure. Unadjusted models and adjusted models controlling for age, gender, race, sexual orientation, and RCT involvement and group assignment (as some of this cohort participated in the ADHERE RCT study) were fit to the data. The ADHERE RCT study involved evaluation of a medication adherence intervention that was not designed to affect HIV disclosure to sex partners or high risk sexual behaviors. The intervention did not show increased medication adherence among treatment as compared with control participants (Samet *et al.*, in press), but there was the possibility that the intervention or participation in the RCT could affect nondisclosure. A chi-square analysis was used to assess the relationship between RCT involvement/group assignment and nondisclosure; a significant association was not observed. However, in an additional effort to avoid potential confounding, we controlled for RCT involvement and group assignment in all adjusted analyses. All initial analyses were conducted with two-sided tests of hypotheses at the .05 significance level. In addition, the Bonferroni method was used to adjust for multiple comparisons.

RESULTS

Cohort Demographics

Participants (n = 124) were aged 25–61 years (mean 42 years), predominantly male (79%), racially mixed (49% Black), one-third without high school graduation (35%), and a minority homeless (14%). Almost one-third of participants (n = 38) identified as gay, lesbian, or bisexual (n = 28 gay or lesbian, 10 bisexual), only five of these (13%) were females.

Cohort Characteristics: Nondisclosure, Sexual Risk Behaviors, Victimization, and Substance Use

One-third of participants (32%) reported nondisclosure of HIV serostatus to a sex partner in

the past 6 months. In terms of sexual risk behaviors, over one-third of the sample (38%) reported multiple sex partners, and 42% reported inconsistent/no condom use. Ten percent of participants had purchased sex with drugs or money, and nine percent had sold sex for drugs or money. (Because of the gendered aspects of sex trade, we assessed whether the prevalence of sex bought and sex sold differed by gender.) Among those buying sex, 100% (n = 13) were males. Among those selling sex, 55% [n = 6] were males, and 45% were females (n = 5). Only 4% of the total sample reported an STD in the past 6 months.

Victimization was less common than sexual risk behaviors. Recent sexual victimization was reported by 3% of the cohort, and recent victimization from physical abuse by 8%. Over one-third of the sample reported binge drinking (39%) and illicit drug use (40%) in the past 30 days.

Age, gender, race, and education were not associated with disclosure of HIV serostatus (Table I). However, participants identifying as gay, lesbian, or bisexual were significantly more likely not to disclose to all sex partners than those identifying as heterosexual (53% and 22%, respectively; $p = .002$). Further analyses were conducted to assess whether effects were attributable to bisexual ($n = 10$) as compared with gay/lesbian ($n = 28$) participants by conducting a Fisher's exact test to assess differences; these groups did not significantly differ in terms of probability of nondisclosure. Additional analyses stratifying prevalence of nondisclosure among gay, lesbian, and bisexual participants by gender demonstrate that two of the five lesbian or bisexual women reported nondisclosure as compared to 16 of the 33 gay or bisexual men.

Nondisclosure of Serostatus—Multivariate Associations

Participants reporting multiple sex partners in the past 6 months, $OR_{adj} = 8.9$, 95% CI = 3.4–23.4, and sex bought with drugs or money, $OR_{adj} = 1.5$, 95% CI = 2.6–50.9, were significantly more likely than those not reporting these behaviors to not disclose to a sex partner (Table II). The association between sex bought and nondisclosure pertains only to

Table I. Demographics for Total Sample of HIV-Infected Persons with a History of Alcohol Problems (N=124) and Frequency of Reporting Nondisclosure vs. Disclosure by Demographics

	Nondisclosers ($n = 40$)	Disclosers ($n = 84$)	χ^2 or t-test value (df)[a]
Gender (%)			0.2 (1, 123)
Female	36	64	
Male	31	69	
Race (%)			1.6 (1, 123)
Black	38	62	
Not Black	27	73	
Education (%)			0.003 (1, 123)
Non-high school graduate	33	67	
HS graduate	32	68	
Sexual orientation (%)			11.1 (1, 123)**
Gay or bisexual	53	47	
Heterosexual	22	78	
Age, years, M (SD)[b]	42.2 (6.3)	41.3 (8.1)	0.8 (1, 122)

[a]Chi-square analyses were used to assess significant differences in dichotomized demographic variables (i.e., gender, race, education and sexual orientation) by disclosure; t tests were used to assess significant differences in continuous demographic variables (i.e., age) by disclosure.

[b]Age was a continuous variable, requiring a t test to assess differences between means; thus means and standard deviations (SD) rather than percentages are presented for this variable.

**$p < .01$.

Table II. Prevalence and Crude and Adjusted Odds Ratios Assessing Associations Between Nondisclosure of HIV Serostatus by Sexual Risk Behaviors and STD, Substance Abuse, and Victimization (Independent Variables)

	Nondisclosers ($n=40$)	Disclosers ($n=84$)	Crude OR (95% CI)	Adjusted OR (95% CI)[a]
Sexual risk (%)				
Multiple sex partners	60	40	9.0 (3.8–21.4)	8.9 (3.4–23.4)
One sex partner	14	86		
Inconsistent/no condom use	29	71	0.8 (0.4–1.7)	0.5 (0.2–1.3)
Consistent condom use	34	66		
Sex sold	64	36	4.2 (1.2–15.5)	2.4 (0.6–10.5)
No sex sold	29	71		
Sex bought	77	23	9.0 (2.3–34.9)	11.5 (2.6–50.9)
No sex bought	27	73		
Recent STD	80	20	9.2 (1.0–85.4)	5.8 (0.5–66.0)
No STD	30	70		
Substance abuse (%)				
Binge drinking	40	60	1.6 (0.8–3.5)	1.1 (0.5–2.7)
No binge drinking	29	71		
Illicit drug use	31	69	0.9 (0.4–1.9)	0.7 (0.3–1.6)
No illicit drug use	33	67		
Victimization (%)				
Sexual abuse	75	25	6.7 (0.7–66.9)	6.0 (0.4–82.1)
No sexual abuse	31	69		
Physical abuse	40	60	1.4 (0.4–5.4)	1.9 (0.4–9.0)
No physical abuse	32	68		

[a]All analyses control for treatment group, sexual orientation, race, age (continuous), and gender.

men as sex purchase was only reported by men in our sample. Adjusting for multiple comparisons using the Bonferroni method did not change the results of the logistic regression analyses as multiple sex partners and buying sex remained significant predictors of nondisclosure.

To assess whether sex trade was significantly associated with nondisclosure after controlling for multiple sex partners, a logistic regression model was created that included both of these variables as well as demographics (age, gender, race, sexual orientation) and RCT involvement and group assignment. Multiple sex partners ($OR_{adj}=6.7$, 95% CI = 2.4–18.4) remained significantly associated with nondisclosure and sex bought ($OR_{adj}=5.0$, 95% CI = 1.0–23.5) became marginally significant in this multivariate model.

DISCUSSION

Approximately one-third (32%) of HIV-infected persons with a history of alcohol problems did not disclose their HIV serostatus to one or more sex partners. The percentage of nondisclosure is comparable to that previously reported by HIV clinic samples using similar questions and timeframes (Duru *et al.*, 2003; Stein *et al.*, 1998). The finding that one-third of HIV-infected patients did not disclose their HIV serostatus to all sex partners is problematic, particularly given the fact that 39% of nondisclosing participants did not use condoms consistently. Similar to previous studies (Duru *et al.*, 2003; Stein *et al.*, 1998), condom use was not related to disclosure. Given that most of our sample was tied into HIV clinical care for 6 months or more at assessment, results from this study support the need for better sexual risk reduction programs promoting condom use and disclosure of serostatus to sex partners within the clinical care setting.

Findings from this study additionally demonstrate that the odds of nondisclosure of HIV serostatus were significantly higher among participants who had multiple sex partners in the past 6 months and participants who had purchased sex in the past 6 months. Although previous studies have demonstrated that nondisclosure is more common among those reporting multiple sex partners and within the context of more casual sexual relationships (Duru *et al.*, 2003; Niccolai *et al.*, 1999; Perry *et al.*, 1994;

Stein *et al.*, 1998), this study offers the first demonstration that those buying sex with drugs or money are significantly more likely to report nondisclosure. Further, an association between buying sex and nondisclosure is observed even after controlling for multiple partners. Notably, this association is only indicative of men, as only men in our sample reported buying sex. Nonetheless, these results, as well as the trend suggesting an association between nondisclosure and selling sex, support the idea that disclosure may be less likely in the context of more casual sexual encounters in which one may feel less obligated or safe disclosing to their sex partner (Klitzman and Bayer, 2003).

Findings from this study further suggest that less disclosure among those reporting sex trade and multiple sex partners is not counteracted by increased condom use. As mentioned above, disclosure is not related to condom use; additionally, study findings suggest higher odds of nondisclosure among those reporting recent STD. Those reporting a recent STD had almost 6 times the odds of nondisclosure as compared with those not reporting a recent STD. The lack of statistical significance of this finding despite the large effect size is likely due to the small number of participants reporting recent STD. Overall, these results speak to the need for prevention programs that focus on promotion of condom use during sex trade and within casual sexual relationships, regardless of willingness or comfort disclosing to these partners.

Consistent with previous research (Duru *et al.*, 2003; Stein *et al.* 1998,), recent drug and alcohol use were not related to disclosure. Victimization from physical abuse was also not related to disclosure, but victimization from sexual abuse was. Participants reporting sexual abuse had about six times the odds of nondisclosure of HIV serostatus. Again, these findings were not statistically significant likely due to small numbers of participants reporting victimization from sexual violence. Although previous studies have identified high rates of sexual abuse among people living with HIV/AIDS (Gielen *et al.*, 2000; Sowell *et al.*, 1999; Zierler *et al.*, 2000), research has not specifically assessed sexual abuse and nondisclosure. Given the lack of choice given to people being victimized in a sexual attack, opportunity for disclosure may not occur or may place the individual at greater risk.

Overall, these results support the need for interventions to facilitate disclosure and condom use among HIV-infected persons with a history of alcohol problems, particularly those reporting sex trade and multiple sex partners. Given recent study demonstrating that a second infection with a different HIV-1 strain can occur and accelerate disease progression among people living with HIV (Allen and Altfeld, 2003), increasing consistent condom use with this population will reduce risk for both the infected individual and their partner.

There are certain limitations to the current study. These include reliance on self-report, potential social desirability and recall biases, and use of a modest sample size; cross-sectional data analyses also preclude assumptions of causality. Additionally, our sample cannot be generalized to HIV-positive individuals as a whole; however, it does appear to be generalizable to sexually active HIV-infected persons with alcohol problems (Samet *et al.*, 2004). These limitations speak to the need for further longitudinal studies of nondisclosure to sex partners among HIV-infected individuals with and without a history of alcohol problems. Additionally, use of larger samples in future research would allow further exploration of the effects of both gender and sexual orientation.

An additional limitation of this study was the use of a dichotomous measure of nondisclosure. Participants reporting nondisclosure to one sex partner is categorized the same as those reporting nondisclosure to all sex partners, but associations between disclosure and sex sold may differ between these groups; our current measure of disclosure precludes our ability to assess such associations. Our measures also did not assess whether nondisclosure was specifically occurring in the context of non-condom use, sex trade, substance abuse or violence; further research is needed to assess the contexts of nondisclosure. Recent qualitative research with gay men in San Francisco indicates that HIV-infected men disclose their HIV serostatus in adherence with community norms, but unprotected sex was more rather than less likely in the context of their disclosure (Sheon and Crosby, 2004). Further research is needed to assess the context of nondisclosure in more detail.

ACKNOWLEDGMENTS

Data management was provided by DM-STAT, Inc., Medford, MA. The authors appreciate the contributions of the clinical staff of the HIV Diagnostic Evaluation Unit at Boston Medical Center, including

Colleen LaBelle, RN, and Jennifer Doyle. We also acknowledge Dr. Anita Palepu and Dr. Theresa Kim for their feedback on this work. Finally, the authors appreciate the contributions of Howard Libman, MD, at the Beth Israel Deaconess Medical Center. This research was conducted in part in the General Clinical Research Center at Boston University School of Medicine, USPHS grant M01 RR00533. Development of this work was supported by 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).

REFERENCES

Allen, T. M., and Altfeld, M. (2003). HIV-1 superinfection. *Journal of Allergy and Clinical Immunology, 112*(5), 829–835.

Bedimo, A. L., Bennett, M., Kissinger, P., and Clark, R. A. (1998). Understanding barriers to condom usage among HIV-infected African American women. *Journal of the Association of Nurses in AIDS Care, 9*(3), 48–58.

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

Brown, B. S. (1998, September). Health Services Resource: HIV/AIDS and drug abuse treatment services—Parts A and B. Retrieved July 18, 2004, from http://165.112.78.61/about/organization/DESPR/HSR/datre/BrownHIV.html

Centers for Disease Control and Prevention [CDC] (2003). Incorporating HIV prevention into the medical care of persons living with HIV: Recommendations from the Centers for Disease Control and Prevention, the Health Resources and Services Administration, the National Institutes of Health, the HIV Medicine Association of the Infectious Diseases Society of America. *MMWR, 52,* RR12 (inclusive page numbers). Retrieved December 13, 2004, from http://www.cdc.gov/mmwr/PDF/rr/rr5212.pdf

Cohen, M., Deamant, C., Barkan, S., Richardson, J., Young, M., Holman, S., Anastos, K., Cohen, J., and 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,* 560–565.

Department of Justice (2004). United States. Bureau of Justice Statistics. *Summary findings.* Retrieved February 27, 2004, from http://www .ojp.usdoj.gov/bjs/cvictgen.htm/

Duru, O., Collins, R., Beckman, R., and Kanouse, D. (2003). Correlates of sex without serostatus disclosure among a probability sample of HIV patients. *Journal of General Internal Medicine, 18*(S1), 265.

Ehrenstein, V., Horton, N. J., and 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. *JAMA, 252,* 1905–1907.

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

Fortenberry, J. D., Brizendine, E. J., Katz, B. P., and Orr, D. P. (2002). The role of self-efficacy and relationship quality in partner notification by adolescents with sexually transmitted infections. *Archives of Pediatric Medicine, 156,* 1133–1137.

Gielen, A. C., McDonnell, K. A., Burke, J. G., and O'Campo, P. (2000). Women's lives after an HIV-positive diagnosis: Disclosure and violence. *Maternal and Child Health Journal, 4*(2), 111–120.

Hays, R. B., McKusick, L., Pollack, L., Hilliard, R., Hoff, C., and Coates, T. J. (1993). Disclosing HIV seropositivity to significant others. *AIDS, 7,* 425–431.

Jaffe, H. W., and Janssen, R. S. (2003). Incorporating HIV prevention into the medical care of persons living with HIV. *Morbidity and Mortality Weekly Report, 52,* 1–24.

Kalichman, S. C., and Nachimson, D. (1999). Self-efficacy and disclosure of HIV-positive serostatus to sex partners. *Health Psychology, 18*(3), 281–287.

Kline, A., and VanLandingham, M. (1994). HIV-infected women and sexual risk reduction: The relevance of existing models of behavior change. *AIDS Education and Prevention, 6*(5), 390–402.

Klitzman, R., and Bayer, R. (2003). *Mortal secrets: Truth and lies in the age of AIDS.* Baltimore, MD: Johns Hopkins University Press.

Liebschutz, J. M., Feinman, G., Stein, M. D., and Samet, J. H. (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.

Mansergh, G., Marks, G., and Simoni, J. M. (1995). Self-disclosure of HIV infection among men who vary in time since seropositive diagnosis and symptomatic status. *AIDS, 9,* 639–644.

Marks, G., and Crepaz, N. (2001). HIV-positive men's sexual practices in the context of self-disclosure of HIV status. *Journal of Acquired Immunodeficiency Syndromes, 27*(1), 79–84.

Marks, G., Richardson, J. L., and Maldonado, N. (1991). Self-disclosure of HIV infection to sexual partners. *American Journal of Public Health, 81,* 1321–1323.

Marks, G., Ruiz, M. S., Richardson, J. L., Reed, D., Mason, H. R., Sotelo, M., and Turner, P. A. (1994). Anal intercourse and disclosure of HIV infection among seropositive gay and bisexual men. *Journal of Acquired Immune Deficiency Syndrome, 7*(8), 866–869.

McGowan, J. P., Shah, S. S., Ganea, C. E., Blum, S., Ernst, J. A., Irwin, K. L., Olivo, N., and Weidle, P. J. (2004). Risk behavior for transmission of human immunodeficiency virus (HIV) among HIV-seropositive individuals in an urban setting. *Clinical Infectious Diseases, 38*(1), 122–127.

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grisson, G., Pettinati, H., and Argeriou, M. (1992). The fifth addition of the Addiction Severity Index. *Journal of Substance Abuse Treatment, 9,* 199–213.

Navaline, H. A., Snider, E. C., Petro, C. J., Tobin, D., Metzger, D., Alterman, A. I., and 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*(S2), 281–283.

Niccolai, L. M., Dorst, D., Myers, L., and Kissinger, P. J. (1999). Disclosure of HIV status to sexual partners: Predictors and temporal patterns. *Sexually Transmitted Diseases, 26,* 281–285.

Perry, S. W., Card, C. A. L., Moffatt, M., Ashman, T., Fishman, B., and Jacobsberg, L. (1994). Self-disclosure of HIV infection to sexual partners after repeated counseling. *AIDS Education and Prevention, 6*(5), 403–411.

Pisani, E., Garnett, G. P., Grassly, N. C., Brown, T., Stover, J., Hankins, C., Walker, N., and Ghys, P. D. (2003). Back to

basics in HIV prevention: Focus on exposure. *BMJ, 326*(7403), 1384–1387.

Pitts, M. K., Smith, A. M., Grierson, J., O'Brien, M., and Misson, S. (2004). Who pays for sex and why? An analysis of social and motivational factors associated with male clients of sex workers. *Archives of Sexual Behavior, 33*(4), 353–358.

Rees, V., Saitz, R., Horton, N. J., and Samet, J. H. (2001). Association of alcohol consumption with HIV sex—and drug-risk behaviors among drug users. *Journal of Substance Abuse Treatment, 21,* 129–134.

Robertson, M. J., Clark, R. A., Charlebois, E. D., Tulsky, J., Long, H. L., Bangsberg, D. R., and Moss, A. R. (2004). HIV seroprevalence among homeless and marginally housed adults in San Francisco. *American Journal of Public Health, 94*(7), 1207–1217.

Romero-Daza, N., Weeks, M., and Singer, M. (2003). "Nobody gives a damn if I live or die": Violence, drugs, and street-level prostitution in inner-city Hartford, Connecticut. *Medical Anthropology, 22*(3), 233–259.

Rothenberg, K. H., and Paskey, S. (1995). The risk of domestic violence and women with HIV infection: Implications for partner notification, public policy, and the law. *American Journal of Public Health, 85,* 1569–1576.

Samet, J. H., Horton, N. J., Meli, S., Dukes, K., Tripp, T., Sullivan, L., and Freedberg, K. A. (2005). A randomized controlled trial (RCT) of a multidimensional intervention to enhance adherence to anti-retroviral therapy in HIV-infected patients with a history of alcohol problems. *Antiviral Therapy, 10*(1), 83–93.

Samet, J. H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D. E., and 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., and Freedberg, K. A. (2004). Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research on Human Retroviruses, 20,* 151–155.

Schnell, D. J., Higgins, D. L., Wilson, R. M., Goldbaum, G., Cohn, D. L., and Wolitski, R. J. (1992). Men's disclosure of HIV test results to male primary sex partners. *American Journal of Public Health, 82*(12), 1675–1676.

Sheon, N., and Crosby, G. M. (2004). Ambivalent tales of HIV disclosure in San Francisco. *Social Science and Medicine, 58*(11), 2105–2118.

Simoni, J., Mason, H., Marks, G., Ruiz, M., Reed, D., and Richardson, J. (1995a). Women's self-disclosure of HIV infection: Rates, reasons, and reactions. *Journal of Consulting Clinical Psychology, 63,* 474–478.

Simoni, J. M., Mason, H. R., Marks, G., Ruiz, M. S., and Richardson, J. L. (1995b). Women living with HIV: Sexual behaviors and counseling experiences. *Women's Health, 23,* 17–26.

Smith, K. L., Saitz, R., Horton, N. J., and Samet, J. H. (2003). Use of the Mini-Mental State Examination in clinical alcohol research studies. *Alcoholism: Clinical and Experimental Research, 27,* 28.

Sowell, R., Seals, B., Moneyham, L., Guillory, J., and Mizuno, Y. (1999). Experiences of violence in HIV-seropositive women in the south-eastern United States of America. *Journal of Advanced Nursing, 30*(3), 606–615.

Sowell, R. L., Lowenstein, A., Moneyham, L., Demi, A., Mizuno, Y., and Seals, B. (1997). Resources, stigma and patterns of disclosure in rural women with HIV infection. *Public Health Nursing, 14,* 302–312.

Stein, M. D., Freedberg, K. A., Sullivan, L. M., Savetsky, J., Levenson, S. M., Hingson, R., and Samet, J. H. (1998). Sexual ethics: Disclosure of HIV-positive status to partners. *Archives of Internal Medicine, 158,* 253–257.

Stempel, R. R., Moulton, J. M., and Ross, A. R. (1995). Self-disclosure of HIV-1 antibody test results: The San Francisco General Hospital Cohort. *AIDS Education and Prevention, 7,* 116–123.

Surratt, H. L., and Inciardi, J. A. (2004). HIV risk, seropositivity and predictors of infection among homeless and non-homeless women sex workers in Miami, Florida, USA. *AIDS Care, 16*(5), 594–604.

Turner, B. J., Fleishman, J. A., Wenger, N., London, A. S., Burnam, M. A., Shapiro, M. F., Bing, E. G., Stein, M. D., Longshore, D., and Bozzette, S. A. (2001). Effects of drug abuse and mental disorders on use and type of antiretroviral therapy in HIV-infected persons. *Journal of General Internal Medicine, 16*(9), 625–33.

Vanable, P. A., McKirnan, D. J., Kittiwut, J. T., Ostrow, D. G., and Hope, B. A. (2000). Impact of combination therapies on HIV risk perceptions and sexual risk among HIV-positive and HIV-negative gay and bisexual men. *Health Psychology, 19*(2), 134–145.

Vanwesenbeeck, I. (2001). Another decade of social scientific work on sex work: A review of research 1990–2000. *Annual Review of Sex Research, 12,* 242–289.

Weber, A. E., Craib, K. J., Chan, K., Martindale, S., Miller, M. L., Schechter, M. T., and Hogg, R. S. (2001). Sex trade involvement and rates of human immunodeficiency virus positivity among young gay and bisexual men. *International Journal of Epidemiology, 30*(6), 1449–1456.

Weller, S., and Davis, K. (2002).Condom effectiveness in reducing heterosexual HIV transmission [update of *Cochrane Database of Systematic Reviews* 2001;(3):CD003255; PMID: 11687062]. *Cochrane Database of Systematic Reviews, 1,* CD003255.

Wenger, N. S., Kusseling, F. S., Beck, K., and Shapiro, M. F. (1994). Sexual behavior of individuals infected with the Human Immunodeficiency virus. *Archives of Internal Medicine, 154*(16), 1849–1854.

Wolf, L. E., and Vezina, R. (2004). Crime and punishment: Is there a role for criminal law in HIV prevention policy. *Whittier Law Review, 25,* 821–886. Retrieved December 13, 2004, from http//:www.ari.ucsf.edu/pdf/crime_and_punishment.pdf

Zierler, S., Cunningham, W. E., Andersen, R., Shapiro, M. F., Nakazono, T., Morton, S., Crystal, S., Stein, M., Turner, B., St. Clair, P., and 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.

Violence Against Women
Volume 12 Number 10
October 2006 936-949

10.1177/1077801206292935
http://vaw.sagepub.com
hosted at
http://online.sagepub.com

Victims of Intimate Partner Violence More Likely to Report Abuse From In-Laws

Anita Raj
Kaí N. Livramento
M. Christina Santana
Boston University
Jhumka Gupta
Jay G. Silverman
Harvard University

The current study of South Asians in the United States was designed to assess quantitatively the association between intimate partner violence (IPV) and emotional abuse by in-laws (n = 169) and to qualitatively identify via in-depth interviews with battered women (n = 23) forms of abuse perpetrated by in-laws. Quantitative findings demonstrate a significant relationship between IPV and abuse from in-laws (odds ratio = 5.7, 95% confidence interval = 1.5-21.5). Qualitative data demonstrate that abuse by in-laws includes emotional abuse (e.g., isolation, social and economic control, and domestic servitude), awareness or support of IPV, and direct physical abuse. Domestic violence interventions with South Asian women must consider abuse from in-laws and IPV experiences.

Keywords: *domestic violence; immigrants; South Asians*

South Asians—people of Indian, Pakistani, Bangladeshi, Nepali, Sri Lankan, Bhutanese, or Maldivian ancestry—compose one of the largest and fastest-growing populations in the United States (Barnes & Bennett, 2002; Indian American Center for Political Awareness [IACFPA], 2004; Migration Information Source [MIS], 2005; U.S. Department of Commerce, Bureau of the Census, 2003) but remain largely ignored in the research literature. One topic that has been investigated in this population is intimate partner violence (IPV). Reported rates of IPV among South Asian women in greater Boston range from 20% to 40% (Raj & Silverman, 2002b; Raj, Silverman, McCleary-Sills, & Liu, 2005); these rates are disproportionately higher than those seen in representative samples of women from Massachusetts and the

Authors' Note: This project was funded through a grant from Boston University School of Public Health. The authors would like to thank the following agencies for advising them on this project: Saheli South Asian Women's Support Network and Asian Task Force Against Domestic Violence. The authors would also like to thank Paromita Shah, JD, for her guidance on immigration laws and Sitara Naheed and Mala Suchdeva for their assistance in recruitment.

United States (Hathaway et al., 2000; Tjaden & Thoenes, 2000) and than those of some Indian provinces (Martin, Tsui, Maitra, & Marinshaw, 1999). Given these disproportionately high rates of IPV, combined with the doubling in size of this U.S. population from 1990 to 2000 (Asian American Federation of New York, 2004; Barnes & Bennett, 2002), there will be increased need for health care providers and legal and social service providers for battered women to understand family violence against South Asian immigrant women in the United States. One recognized cultural aspect of violence against South Asian women is abuse by in-laws.

Although popular press media (Mehra, 2003; "Pregnant Woman," 2004; Singh, 2000) and internationally recognized Asian Indian films such as *Bandit Queen* (Kapur, 1995) and *Fire* (Mehta, 1997) have profiled in-law-perpetrated emotional abuse of South Asian women, there has been a paucity of research exploring both the range of experiences of abuse by in-laws and its relation to IPV. Research in India has documented both abuse from in-laws (Ramanathan, 1996) and increased risk for partner violence by Indian men whose parents are dissatisfied with their dowry (Verma & Collumbien, 2003). Smaller, primarily qualitative studies with South Asian battered women in the United States demonstrate that in-laws tolerate or support the partner violence and also may directly perpetrate emotional and physical abuse against these women (Abraham, 1999; Mehotra, 1999; Rianon & Shelton, 2003; Supriya, 1996). Notably, this work indicates that mothers-in-law and sisters-in-law are more commonly the perpetrators of these types of abuse (Mehotra, 1999; Supriya, 1996). Unfortunately, none of this research was specifically designed to assess the association between in-law abuse and IPV reported by women, nor was it developed to provide insight into the forms of abuse perpetrated by in-laws. Building on previous work, the purpose of this study is twofold: (a) to assess quantitatively the prevalence of emotional abuse by in-laws and its relation to IPV among a community-based sample of South Asian women residing in greater Boston and (b) to describe types and experiences of emotional and physical abuse perpetrated by in-laws using qualitative data from local South Asian victims of IPV.

Method

This project was conducted in two components: (a) a quantitative survey assessment with South Asian women currently in heterosexual relationships ($N = 210$) and (b) in-depth interviews with South Asian women reporting a history of victimization from a male partner ($N = 23$). Methods for both components are outlined below; the Institutional Review Board of Boston University Medical Center approved these studies.

Method for Survey Assessment With South Asian Married Women

Participants in the cross-sectional survey ($N = 210$) were recruited via community outreach (fliers, snowball sampling, referrals) to participate in a South Asian

women's health study conducted from August 2001 to January 2002; the recruitment strategies used did not allow the assessment of response rates. Survey data were collected through 15-minute, anonymous surveys offered either at locations deemed convenient by participants (e.g., participants' homes or nearby libraries) or over the phone for women who preferred not to complete the survey in person ($n = 9$). The survey included demographic questions and assessments of history of IPV and health outcomes. All survey data were collected in English. Data included in the current analyses excluded two participants who did not answer IPV questions and 39 participants who were unmarried at the time of the interview, yielding a sample size of 169 for the current quantitative analyses.

Written consent was obtained from all participants surveyed in person; verbal consent was obtained from all telephone participants. Participants received $15 for their participation. Subsequent to survey administration, all participants received a list of referrals for culturally tailored IPV, mental health, and sexual health services. Telephone participants received this list verbally at the time of the interview and later received a copy of the list, the consent form, and the monetary incentive by mail. Women trained in women's health and survey administration served as proctors and interviewers for this study.

Survey sample. Participants in the cross-sectional survey ($n = 169$) ranged in age from 22 to 68 years ($Mdn = 32$ years). The vast majority of participants (95.9%) were Indian; 98.2% were not U.S.-born. One third of the sample (30.2%) were U.S. citizens; 30.2% were legal permanent residents, and 27.8% were on spousal visas (i.e., their visa status was dependent on the employment of their spouse in the United States). Immigrants reported immigration from 6 months to 40 years ago ($Mdn = 6.0$ years); 26.5% of the sample had immigrated within the past 2 years. The sample earned a relatively high income and was highly educated, with 69.0% reporting an annual household income of $50,000 or greater and 47.3% reporting postgraduate training; 12.4% of the sample reported a high school education or less. Almost all participants reported a South Asian husband (98.8%), and 70.2% had children.

Measures. Single survey items assessed personal and partner demographics, including age, income, education, country of birth, years in the United States, current visa or citizenship status, and relationship status and length. Four items adapted from the Massachusetts Behavioral Risk Factor Surveillance System (Massachusetts Department of Public Health, 2000) were used to assess physical abuse (1 item), sexual abuse (2 items), and injury from abuse (1 item) by their current male partner; response options for these items were "Yes, 1-2 times in the past year," "Yes, more than 2 times in the past year," "Not in the past year but previously in our relationship," and "Never in our relationship." Emotional abuse from in-laws was measured via a single item, created for use in this survey, that asked married participants if they had ever been emotionally abused (e.g., sworn at, called stupid or crazy, kept from seeing

family or friends) by their in-laws. For those reporting emotional abuse by in-laws, they were asked an open-ended question regarding which in-laws had done this.

Data analyses. Frequency analyses were conducted with cross-sectional survey data from South Asian women to assess prevalence of emotional abuse against the participant by an in-law and the type of in-law who perpetrated the abuse. An odds ratio (OR) and a 95% confidence interval (CI) were generated via simple logistic regression analysis to assess the association between emotional abuse by in-laws and IPV.

Method for In-Depth Interviews With South Asian Victims of IPV

In-depth interview participants (N = 23) were recruited via referral from community leaders known for assisting battered women and via outreach to all participants of the cross-sectional survey. We asked women with a history of IPV to participate in a study of South Asian women's experiences of abuse and mistreatment in their relationships. This study was conducted from August 2001 to May 2002. Two of the 23 survey participants were recruited from the cross-sectional survey. Again, the recruitment strategies did not allow assessment of response rates. Interview data were collected on IPV, help seeking, the immigrant experience, and health through 60- to 90-minute, confidential, audiotaped interviews offered in convenient and secure locations of the participants' choosing. Survey data were also collected on demographics, IPV, and related help seeking. Because of resource limitations, we planned to collect all data in English; however, some interviewees lapsed into South Asian languages for parts of interviews. We attempted to elicit repetition of the information in English, but when participants were unable to translate, we translated interviews on transcription. Survey data were linked to transcribed or translated and transcribed tapes via unique identifiers to preserve confidentiality.

Written consent was obtained from all interviewees. On interview completion, all interviewees received $65 in incentive payment ($50 for the interview and $15 for the survey) and a list of referrals for culturally tailored IPV, mental health, and sexual health services. Women trained in women's health and survey and in-depth interview administration served as proctors and interviewers for the studies.

Sample. In-depth interview participants (N = 23) ranged in age from 25 to 53 years (Mdn = 37 years). This sample was less educated and earned a lower income than the survey sample. One fourth of interviewees (26.0%, n = 6) reported a high school education or less; 39.1% (n = 9) of the sample had postgraduate training. One third of the sample (34.8%, n = 8) reported an annual household income of $20,000 or less; 34.7% ($n$ = 8) reported an annual household income of $50,000 or more. The majority of the interview sample (65.2%, n = 15) was Indian, and 30.4% (n = 7) were Bangladeshi; one participant was Nepali. All were non-U.S. born; 56.5% (n = 13) were legal permanent residents, 17.4% (n = 4) were U.S. citizens, and 13.0% (n = 3)

were on spousal visas. Participants had been in the United States for 0.5 to 20 years (*Mdn* = 6 years); 56.5% came to the United States because their partner resided in the United States or was to come to the United States for his educational or employment opportunities. Approximately half of interviewees (47.8%, $n = 11$) were currently involved with their abusive partner.

Measures. Measurement for the in-depth interviews involved 12 open-ended questions regarding participants' relationship with their abusive husband, the types of abuse experienced, family involvement and awareness of abuse, perceived physical and mental health-related effects of the abuse, present physical and mental health status, and utilized or needed social, legal, and health services.

Data analyses. Qualitative data analysis of semistructured, in-depth interviews was conducted using a grounded theory approach such that codes were generated iteratively. According to Glaser and Strauss (1967), the grounded theory approach uses an emergent theme technique for iterative code generation. In addition, it provides the researcher with tools to link concepts to facilitate development of a model for understanding human experience (Glaser & Strauss, 1967; Strauss, 1987; Strauss & Corbin, 1990). Based on this approach, our qualitative research team (two trained coders and the principal investigator) read each transcript, identifying and recording (memoing) themes. Following review of the 23 transcripts, memos were reviewed for recurring themes across transcripts. These were viewed as emergent codes. Following this initial memoing, the two coders reviewed text for each code category and memoed linkages across categories. Segments of transcriptions could receive multiple codes. Intercoder reliability was assessed using the technique used by Carey, Morhan, and Oxtoby (1996). This technique involved the coders independently coding the transcriptions and then coming together to reach consensus. In cases of disagreement between coders, the principal investigator made the final decision. This process generated the following emergent themes related to abuse by in-laws: control and isolation, economic control, criticism of her family and their dowry provision, verbal abuse, domestic servitude, delaying access to food, physical abuse, and tolerance or support of IPV. Quotes from these themes were included to support and explain quantitative findings.

Results

Results From Survey Assessment With South Asian Married Women

IPV in the current marital relationship was reported by 23.1% of survey participants ($n = 169$), and emotional abuse by in-laws was reported by 5.9% of participants. Only 3 of the 10 participants reporting abuse by in-laws reported the perpetrator's relationship to them. One indicated abuse by the father-in-law, mother-in-law, and

sister-in-law, one reported abuse by the mother-in-law only, and one reported abuse by the sister-in-law only. Logistic regression analysis revealed that participants with a history of IPV were significantly more likely than those with no such history to report emotional abuse by in-laws (OR = 5.7, 95% CI = 1.5-21.5). Among participants reporting no IPV in their current relationship, 3.1% reported emotional abuse from in-laws. Among participants reporting IPV in their current relationship, 15.4% reported emotional abuse from in-laws.

Results From In-Depth Interviews With South Asian Victims of IPV

Emotional abuse. Of the 23 IPV victims participating in our in-depth interviews, 12 (54.5%) reported emotional abuse from in-laws. In their interviews, they described various forms of emotional abuse by in-laws, including isolation, economic control, verbal abuse and degradation, criticism of her family and complaints about dowry, domestic servitude, and controlling intake or access to food.

Control and isolation. Women spoke at length of the ways their in-laws would isolate them, limit their contact with family, and control the way they spent their time:

> I could not use the phone to call my family in India. Even though my sister had given me a calling card to call her with, I always had to sneak the phone calls in when my husband was at work and my mother[-in-law] and father-in-law were at church or otherwise out of the house. I was trapped in my husband's family's home without the ability to communicate freely.

> It . . . was just me being in the house all of the time with no permission to do anything by my own, no permission to call my parents . . . . [When I asked] "Why am I not allowed to go out?' [My in-laws responded] "No, because this is a rule of our family; you're not supposed to go out and step out of the house without your husband being with you."

> I was not allowed to watch TV, nothing, I couldn't do anything . . . . So [my husband] would call me [from America when I was in India with his family], [my mother-in-law] wouldn't even let me talk to my husband.

Economic control. Women described how their in-laws prevented their economic autonomy and how this affected their freedom and self-confidence, often through the abusive husband.

> We got the [wedding] gifts and everything, cash, and I don't really know because my in-laws, basically my mother-in-law, did not show me any of the gifts we received . . . . [My mother-in-law said to me] "This is a watch that I think you should have it . . . . We received a lot of cash but because we spend on the wedding, we took the cash."

> I and my husband lived together [without my in-laws], but even that didn't turn out to be okay because my husband used to call his mother . . . . If I need clothes . . . a phone card [to call India], I have to take permission from my husband. My husband has to take permission from his mother . . . . If mother agrees yes and his father agrees yes, then it's a yes.
>
> [Since my in-laws had the money,] [my father-in-law] would go shopping and he would decide for me, what colors to buy, what clothes to buy . . . . I started to lose my self-confidence.

Verbal abuse and degradation. Almost all of the women reporting abuse from in-laws described verbal abuse and degradation. This form of abuse most commonly involved questioning her character and her contributions to the family and household.

> My mother in-law . . . stayed with me for 4 months here. She always used to scold me that her son was not happy with me . . . . [She would say], "You do not have a good character . . . . It is our fate that we got married to such a worst girl in our life."
>
> She'd call me names. She'd called me names like prostitute. She'd threaten me . . . [that she would] get her son to divorce me.
>
> [My mother-in-law and father-in-law would say], "You are good for nothing. You sit and you eat and you have no shame. And you don't realize how much money and how many dollars you have wasted by not earning and just sitting at home and eating food."

Criticizing her family and complaints about dowry. They also described enduring criticisms about their family and complaints about insufficient or lack of dowry.

> [My mother-in-law] used to say to me, . . . "Your mom was not at all a nice person . . . and we are the ones who are suffering. And we did not get a proper dowry."
>
> The reason I think my mother-in-law was not happy with our wedding and stuff was because, umm, my parents didn't give enough . . . . My mother-in-law told me that she didn't get enough [from my family] and that people were laughing at her.

Domestic servitude. One of the most common forms of abuse perpetrated by in-laws was forced servitude; every mention of this type of abuse was at the hands of the mother-in-law. This was commonly described as starting early in the marriage when the daughter-in-law first came home and lived with her in-laws, and then it would continue through visits once the woman no longer lived with her in-laws. This servitude was demanded regardless of the woman's pregnancy condition.

> And the next day of my marriage it was like, "Oh well, here's the soap and here's the broom. Clean the bathroom first and then iron the clothes." And it was a heap of clothes to iron and I was really shocked.
>
> Even though we had a washing machine, my mother-in-law continued to demand that I wash the heavy clothes, curtains, and turbans by hand, which was very difficult for me, since I was already 5 months pregnant.

> My mother-in-law wants me to massage her legs every night before I sleep, when I was 9 month pregnant. At the time also, she wants me to work so much. I didn't know what I can do.

Denied or delayed access to food. In conjunction with their role as domestic servant for the household, many women described how they were not allowed to join the family at meals and were required to wait until the family had completed their meal before being allowed to eat.

> All of item, whatever I used to cook but I won't get it . . . . If she is home, the mother is always home, so I won't get food. And whenever he [my husband] will come, he will give me a little. Little. But not like if I cooked four or five things in the dinner table, I will get only rice and daal.

> Meanwhile my mother-in-law was treating me worse . . . . She then started prohibiting me from eating meals with the family. He and his parents then started forcing me to prepare and eat meals with meat in them [the participant was a vegetarian] . . . . After doing all this for my husband and his family, I still was not allowed to eat with them.

Physical abuse. Although neither in-laws' involvement with IPV nor their perpetration of physical abuse was measured in our survey with cross-sectional survey participants or in-depth interviewees, it was described by some of our interviewees, although less often than was emotional abuse. Women reported that their in-laws, particularly their mother-in-law, were often aware of and tolerated the IPV being perpetrated against them, and at times instigated the IPV or directly assaulted the women themselves.

In-laws' tolerance or support of IPV perpetration. Women described their mother-in-law's tolerance and sometimes even support of the IPV their sons perpetrated against their wives, not coming to aid the victim even when called at times of victimization.

> When his mom was there for the 4 months, he like used to make me touch her feet and apologize to her saying that like I was a bad girl who had entered into their life and I, uh, I had to pay as a slave and dog . . . . His mother never tried to stop when I was acting in this way.

> He would hit me in front of his mom. Initially, she would worry a bit because she would worry I might complain to, in that view she used to stop 2 to 3 times when she was here, but after that she would never try to intervene in between us. I mean there were moments like when he would hurt me so bad that I couldn't walk and like I would scream for help, calling for her, but she never came.

In-laws' incitement of IPV. Furthermore, some women described feeling that, at times, their mother-in-law actually incited the IPV perpetrated. Mothers-in-law were often blamed for the violence of their sons.

> [My mother-in-law] would complain it to [her] son. That "she didn't do this thing right," that "she purposely didn't do it the way I wanted her to do it." . . . And he would slap me or he would do something to me so I am not doing it again.
>
> He used to hit me a lot. If his mother said there was too little salt in the food, he would hit me . . . . When he used to come back from work . . . his mother would complain [about me]. Almost all the time he would hit me because of his mother.
>
> [My brother-in-law and sister-in-law] tell my husband that I tell bad things about my in-laws to outsiders. My husband becomes very angry and beats me up, sometimes he even throttles me.

In-laws' perpetration of physical abuse. Some interviewees described experiencing severe physical abuse from in-laws, again even during pregnancy. Although participants reported physical abuse by their brothers-in-law and sisters-in-law, physical abuse by their mothers-in-law was more often discussed.

> I had to wake up at 4 o'clock. I had to scrub all the dishes, make chappatis, cook for the whole big family . . . . Each time I made a mistake, [my mother-in-law] used to hit me. She never used to give me to eat or drink. I was starving.
>
> So that time when I came from the hospital [after being abused by her husband] also, she really hit me. She said, "You are my daughter-in-law. We got you married to my son and you have to listen to everything he says. For his happiness you have to agree for everything he wants you to do."
>
> [My mother-in-law] used to hold my hair and bang my head to the wall. Sometimes she used to beat me with firewood. I have marks wherever she used to beat me. She burnt me with firewood, too.
>
> One time [my mother-in-law] poured kerosene in my hair to burn me . . . and she closed the door. She was hinting for matches and we have gas lighter . . . and she didn't know how to use one. One time, I was pregnant with my son . . . 8 months, she pushed me into the well . . . . She physically abused me a lot.

Discussion

Findings from our quantitative study of married South Asian women in the United States indicate that emotional abuse by in-laws is not common but is significantly more likely among women experiencing IPV. Although 15% of women reporting IPV also reported emotional abuse by in-laws, only 3% of women who were not in a relationship in which IPV had occurred reported emotional abuse by in-laws. These findings demonstrate a significant link between IPV and emotional abuse by in-laws in this population.

Results from this study also document the various types of emotional abuse inflicted on South Asian female IPV victims by their in-laws. These include isolation, social and economic control, verbal abuse and degradation, dowry complaints,

and domestic servitude; these forms of abuse are notable in their overlap with forms of violence commonly perpetrated against women by their abusive partner (National Coalition Against Domestic Violence [NCADV], 2004). This overlap suggests that, similar to male-perpetrated IPV against women (NCADV, 2004), in-laws' abuse of South Asian women stems from an entitlement to control these women and is based in patriarchy and traditional ideologies of female inferiority and servitude.

In addition to in-law-perpetrated emotional abuse of South Asian women, our qualitative research also documents in-laws' direct physical abuse of these women and in-laws' tolerance, support, and even incitement of male-perpetrated IPV among our South Asian sample of IPV victims. These findings are consistent with those reported in previous studies of South Asian victims of IPV (Mehotra, 1999; Supriya, 1996). These findings in conjunction with the quantitative findings discussed above suggest that abusive in-laws and batterers may be reinforcing one another's abuse of women as they support one another's entitlement. Notably, this in-law abuse and reinforcement seems to most often stem from the mother-in-law. These findings are consistent with previous research with South Asian women and with women from Arab, other Asian, and Hispanic immigrant groups (Haj-Yahia, 2000; Huisman, 1996; Leung, Kung, Lam, Leung, & Ho, 2002; Mehotra, 1999; Morash, Bui, & Santiago, 2000; Supriya, 1996). Increased likelihood of abuse by mothers-in-law as compared with other family members may be attributable, in part, to a female hierarchy based on generation and being the mother of a son. Rather than bond formation among women in the family context, a pecking order may become established, providing the mother-in-law with the greatest power and entitlement among women and the daughter-in-law with the least power.

Notably, despite mass media emphasis on the role of joint family systems in abuse of South Asian women by in-laws (Mehra, 2003; "Pregnant Woman," 2004; Singh, 2000), none of the 9.5% of women residing with their in-laws at the time of the survey reported a history of abuse from in-laws. Joint family systems are a South Asian cultural norm and involve the woman leaving her home at marriage and residing in the home of her husband with his parents, younger siblings, and potentially adult brothers and their families. Although several in-depth interviewees (all of whom were victims of IPV) did describe experiences of abuse from in-laws while they were residing with the in-laws, they also described experiences of abuse from in-laws during their visits to the in-laws' homes and their in-laws' visits to their homes and of intentionally inciting the husband's abuse of his wife through telephone conversations. Thus, it appears that although the joint family system, which requires patrilocality (residence closer to the husband's family) and inhibits matrilocality (residence closer to the wife's family), may facilitate opportunities for abuse of women by in-laws, it is not a necessary context for this mistreatment. These findings are consistent with our previous work demonstrating that neither residing in a joint family system nor having an arranged marriage is related to higher rates of IPV among South Asian women (Raj & Silverman, 2002a).

Findings from the current study must be considered in the context of certain limitations; these include the use of English-only instruments (although many

interviewers were bilingual and able to translate) and convenience samples. Community outreach may not have effectively reached the most isolated women, among whom IPV and abuse from in-laws may be more prevalent. Our higher socioeconomic status, predominantly Asian Indian samples may limit generalizability of study findings, although immigration patterns in the United States maintain Asian Indians as the largest South Asian immigrant group, with more highly educated Asian Indians being more likely than less educated Asian Indians to come to the United States (Barnes & Bennett, 2002; IACFPA, 2004). Nonetheless, the English-based assessments likely resulted in overrepresentation of highly educated, English-fluent women in this study.

Additional limitations of the quantitative survey study include the use of cross-sectional data analyses that cannot be used to indicate causality but can only be used to assess the association between IPV and emotional abuse by in-laws. Regression analysis using such a small sample size must also be interpreted conservatively, as the wide CI because of this sample size and low frequency of emotional abuse by in-laws provide a less stable estimate. Finally, both abuse by in-laws and IPV prevalence rates were likely underestimated among this sample because of reliance on self-report, reliance on community outreach for recruitment, and use of more conservative measures (i.e., single items with more limited definitions of IPV and use of a single item that measured only emotional abuse from in-laws). In a previous study with a similar sample (see Raj & Silverman, 2002b), a more extensive measure of IPV was used, contextualizing these experiences as "fighting" rather than "abuse," and a far higher (40% vs. 23%) prevalence of IPV was observed. Additional research, both qualitative and quantitative, with larger and more representative samples from South Asian and other racial/ethnic groups is needed to provide further insight into this issue.

Despite these limitations, findings from this study for the first time demonstrate the interrelationship between abuse by in-laws and IPV among South Asian women and document the varied forms of emotional and physical abuse perpetrated, particularly by mothers-in-law, against South Asian female victims of IPV. These findings have critical implications for current IPV prevention and intervention efforts with South Asian women. Although current intervention practices focus on promoting and restoring a woman's safety from her husband, findings from this and other studies mentioned above indicate the need to assess the role of a woman's in-laws in the abuse, regardless of whether she is currently residing with her in-laws. South Asian community advocates, domestic violence advocates, and legal, health, and social service providers should recognize that abuse by in-laws is not likely a result of South Asian cultural practices, such as the joint family system and arranged marriage, but is rather supported through traditional, patriarchal ideologies that promote female submission and servitude to the husband and his family. Characterizing South Asian cultural practices, such as living with in-laws, as abusive not only is inaccurate but also, without considering a broader context, will likely alienate South Asian women in crisis, resulting in missed opportunities for effective outreach and community dialogue.

Some specific areas in which consideration of in-law abuse can be incorporated into existing prevention and intervention practices are as follows:

- Build a health care response by screening for abuse by in-laws among South Asian women identified as being in an abusive relationship. By expanding screening and safety assessments to include a larger familial context, the health care response to IPV in this community may improve.
- Identify types of legal assistance available for women reporting abuse by in-laws. These can include restraining orders and options for prosecution of in-laws who have been physically abusive.
- Provide social services and counseling services that consider abuse by in-laws and its effects on women who have been victimized by IPV. Such services should assess the role of and relationships with the in-laws throughout the marriage for better insight into how the relationship with the in-laws affected the IPV.
- Provide community education on IPV that includes discussions of abuse by in-laws. These discussions, of course, must maintain the perspective that although certain cultural practices, such as the joint family system, may facilitate abuse by in-laws, abusive behavior toward the daughter-in-law can only occur with in-laws who hold patriarchal ideologies and entitlement to abuse.

Although the current study dealt specifically with South Asian women, the importance of these findings is unlikely limited to this population, as abuse by in-laws has been documented in White, African American, Hispanic, Arab, and other Asian communities as well (Haj-Yahia, 2000; Huisman, 1996; Leung et al., 2002; Mehotra, 1999; Morash et al., 2000; Zink, Elder, Jacobson, & Klostermann, 2004). Further research is needed to better understand the prevalence of abuse by in-laws, its relation to IPV, and the forms of this abuse within diverse racial/ethnic populations.

References

Abraham, M. (1999). Sexual abuse in South Asian immigrant marriages. *Violence Against Women, 5*, 591–618.

Asian American Federation of New York. (2004). *Census information center: Census 2000 detailed Asian groups in the United States*. Retrieved December 28, 2004, from http://www.aafny.org/cic/table/ust.asp

Barnes, J. S., & Bennett, C. E. (2002). *The Asian population: 2000. Census 2000 Brief*. Retrieved July 27, 2004, from http://www.census.gov/prod/2002pubs/c2kbr01-16.pdf

Carey, J. W., Morhan, M., & Oxtoby, M. J. (1996). Intecoder agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *Cultural Anthropology Methods Journal, 8*(3), 1-5.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Haj-Yahia, M. M. (2000). Wife abuse and battering in the sociocultural context of Arab society. *Family Process, 39*, 237-255.

Hathaway, J. E., Mucci, L. A., Silverman, J. G., Brooks, D. R., Mathews, R., & Pavlos, C. A. (2000). Health status and health care use of Massachusetts women reporting partner abuse. *American Journal of Preventive Medicine, 19*, 302-307.

Huisman, K. A. (1996). Wife battering in Asian American communities: Identifying the service needs of an overlooked segment of the U.S. population. *Violence Against Women, 2*, 260-283.

Indian American Center for Political Awareness. (2004). *Inmigration.* Retrieved April 28, 2004, from http://www.iacfpa.org/immigration/immigration.htm

Kapur, S. (Director). (1995). *Bandit queen.* [Motion picture]. United States: Arrow Releasing.

Leung, W. C., Kung, F., Lam, J., Leung, T. W., & Ho, P. C. (2002). Domestic violence and postnatal depression in a Chinese community. *International Journal of Gynecology and Obstetrics, 79*, 159-166.

Martin, S. L., Tsui, A. O., Maitra, K., & Marinshaw, R. (1999). Domestic violence in northern India. *American Journal of Epidemiology, 150*, 417–426.

Massachusetts Department of Public Health. (2000). *Massachusetts behavioral risk factor surveillance 2000.* Boston: Center for Health Information, Statistics, Research and Evaluation.

Mehotra, M. (1999). The social construction of wife abuse: Experiences of Asian Indian women in the United States. *Violence Against Women, 5*, 619-640.

Mehra, B. (2003, November 3). *Dowry and demands: What's and how's.* Retrieved July 30, 2004, from http://www.sulekha.com/expressions/column.asp?cid=305887

Mehta, D. (Director). (1997). *Fire.* [Motion picture]. United States: New Yorker Films.

Migration Information Source. (2005). *Global data center.* Retrieved January 11, 2005, from http://www.migrationinformation.org/GlobalData/countrydata/data.cfm

Morash, M., Bui, M., & Santiago, A. (2000). Gender specific ideology of domestic violence in Mexican origin families. *International Review of Victimology, 1-3*, 67-91.

National Coalition Against Domestic Violence. (2004). *The problem. What is battering?* Retrieved July 18, 2004, from http://www.ncadv.org/problem/what.htm

Pregnant woman stripped, beaten up by husband, in-laws. (2004, January 30). *Newindpress.com.* Retrieved July 29, 2004, from http://www.newindpress.com

Raj, A., & Silverman, J. G. (2002a). Intimate partner violence against immigrant women: The roles of immigrant culture, context, and legal status. *Violence Against Women, 8*, 367-398.

Raj, A., & Silverman, J. G. (2002b). Intimate partner violence against South Asian women in greater Boston. *Journal of the American Medical Women's Association, 57*, 111-116.

Raj, A., Silverman, J. G., McCleary-Sills, J., & Liu, R. (2005). Immigration policies increase South Asian immigrant women's vulnerability to intimate partner violence. *Journal of the American Medical Women's Association, 60*, 26-32.

Ramanathan, S. (1996). Violence against women. *International Medical Journal, 3*, 145-148.

Rianon, N. J., & Shelton, A. J. (2003). Perception of spousal abuse expressed by married Bangladeshi immigrant women in Houston. *Journal of Immigrant Health, 5*, 37-44.

Singh, D. (2000, August 24). They still burn their wives for money! *IndiaNest.com.* Retrieved July 30, 2004, from http://www.boloji.com/women/ds07.htm

Strauss, A. (1987). *Qualitative analysis for social scientists.* New York: Cambridge University Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research.* Newbury Park, CA: Sage.

Supriya, K. E. (1996). Confessionals, testimonials: Women's speech in/and contexts of violence. *Hypatia, 11*(4), 92-106.

Tjaden, P., & Thoenes, N. (2000). *Extent, nature and consequences of intimate partner violence: Findings from the National Violence Against Women Survey.* Washington, DC: U.S. Department of Justice, National Institute of Justice.

U.S. Department of Commerce, Bureau of the Census. (2003, December). *The foreign-born population: 2000.* Washington, DC: Author.

Verma, R., & Collumbien, M. (2003). Wife beating and the link with poor sexual and risk behavior among men in urban slums in India. *Journal of Comparative Family Issues, 34*, 61-74.

Zink, T., Elder, N. A., Jacobson, J., & Klostermann, B. (2004). Medical management of intimate partner violence considering the stages of change: Precontemplation and contemplation. *Annals of Family Medicine, 2*, 231-239.

Anita Raj, PhD, is an associate professor at Boston University School of Public Health. Her research focuses on sexual risk, violence against women, and immigrant and racial/ethnic minority health. She was principal investigator on the current study and has other peer-reviewed published work on U.S. South Asian women's experiences of intimate partner violence, their vulnerabilities to this violence, and their increased health risk as a consequence of victimization. She is also a member of and advisor to Saheli, a community-based organization in greater Boston that serves South Asian women in crisis.

Kaí N. Livramento is a project coordinator on a research study at Boston University School of Public Health. Her interests are in the areas of sexual health and minority health research; she is currently a student in the master's of public health program at Boston University.

M. Christina Santana, MPH, is a project manager on research studies at Boston University School of Public Health. Her work has been focused on women's health and racial/ethnic minority health, including both quantitative and qualitative research.

Jhumka Gupta, MPH, is a doctoral student at Harvard University School of Public Health. Her research interest is primarily in the area of violence against women. Her previous work includes assessment of violence against women and other health issues in the U.S. South Asian community.

Jay G. Silverman, PhD, is an assistant professor at Harvard University School of Public Health. His research focuses on intimate partner violence and other forms of gender-based violence across the life span. He has numerous peer-reviewed publications on these subjects, including work showcasing heightened vulnerability to domestic violence among immigrant women. He is also an advisor to Saheli, a community-based organization in greater Boston that serves South Asian women in crisis.

Perpetration of Intimate Partner Violence Associated With Sexual Risk Behaviors Among Young Adult Men

Anita Raj, PhD, M. Christina Santana, MPH, Ana La Marche, BA, Hortensia Amaro, PhD, Kevin Cranston, MDiv, and Jay G. Silverman, PhD

Objective. We assessed the association between intimate partner violence (IPV) perpetration and sexual risk behaviors and fatherhood (having fathered children) among young men.

Methods. Sexually active men aged 18 to 35 years who visited an urban community health center and who reported having sexual intercourse with a steady female partner during the past 3 months (N = 283) completed a brief self-administered survey about sexual risk behaviors, IPV perpetration, and demographics. We conducted logistic regression analyses adjusted for demographics to assess associations between IPV and sexual risk behaviors and fatherhood.

Results. Participants were predominantly Hispanic (74.9%) and Black (21.9%). Participants who reported IPV perpetration during the past year (41.3%) were significantly more likely to report (1) inconsistent or no condom use during vaginal and anal sexual intercourse, (2) forcing sexual intercourse without a condom, (3) having sexual intercourse with other women, and (4) having fathered 3 or more children.

Conclusion. IPV perpetration was common among our sample and was associated with increased sexual risk behaviors. Urban community health centers may offer an important venue for reaching this at-risk population. (*Am J Public Health*. 2006;96:1873–1878. doi:10.2105/AJPH.2005.081554)

Extensive evidence from diverse populations of women has shown that intimate partner violence (IPV) victimization—a health issue estimated to affect 1 in 4 US women[1–7]—is significantly associated with low contraceptive and condom use and adverse sexual and reproductive health outcomes (e.g., pelvic pain, menstrual abnormalities, sexually transmitted disease (STD)/HIV, unwanted pregnancy, and multiple abortions).[8–10] This research has primarily been limited to studies of female IPV victims and has rarely included reports from male perpetrators. Although research that has included women's and girls' reports of male partner behavior shows a potential association between high rates of IPV perpetration and sexual risk behaviors among young adult men,[11–18] there has been little direct study of whether abusive male partners pose a greater sexual risk to women because of the men's own risky sexual behaviors. Within qualitative studies, battered women have reported that abusive male partners prevented them from using contraception and thus, forced them to have unprotected sex, sometimes for the purposes of conception.[19,20] The sole study of sexual risk behaviors and IPV that was conducted with men showed that IPV perpetration was associated with sexual infidelity, multiple sexual partners, and unprotected anal sexual intercourse.[21] This sample was drawn from a methadone treatment clinic and thus limited generalizability of the study findings. No published study to date has assessed the association between IPV perpetration and fatherhood (i.e., having fathered children) among men. Our goal was to build upon the previous studies by assessing the association of IPV perpetration with sexual risk behaviors and fatherhood among a sample of young adult men who attended a large urban community health center.

METHODS

English- and/or Spanish-speaking men aged 18 to 35 years who reported having sexual intercourse with a female partner during the past 3 months were recruited from a large urban community health center in Boston, Mass, that primarily serves lower-income Hispanic and Black clients. On the basis of these inclusion criteria, men who entered the health center were screened at registration by trained research staff who were fluent in both Spanish and English. Men were screened if they came to the health center for their own care *or* if they were accompanying someone else. Those who agreed to participate in a brief, anonymous men's health survey were then escorted by research staff to a private room, where individuals were screened for a second time to verify eligibility. Upon obtaining oral consent, the self-report paper survey was administered; oral rather than written consent was used to eliminate the need for participants' signatures and to better ensure anonymity. The consent procedures, informed consent information sheet, and survey were offered in either English or Spanish; the Spanish versions were professionally back-translated for use in our study. After survey completion (approximately 20 minutes), participants were given $15 for their time and were informed about health center services, including HIV counseling and testing, STD testing, and social services related to substance abuse and IPV.

Participation

Participants were recruited from April 2004 to February 2005. Of the 432 men who were approached, 354 were eligible; 29 refused to participate, which resulted in a 92% participation rate. Forty-eight percent of the participants were at the health center for their own health care, 46% were accompanying a female partner or child to appointments for their own heath care, and 6% were attending a health fair. After the surveys were reviewed, 18 of the 325 survey participants

were excluded because of their age (n=6) or because they had not had sexual intercourse with a female partner during the past 3 months (n=12). Of the remaining 307 participants, 92.2% (n=283) reported that they had a steady female partner and that they had had vaginal sexual intercourse during the past 3 months with this partner. Our analyses were limited to these individuals.

Survey Measures

Single items assessed participants' age, race/ethnicity, education level, income, employment, relationship status, length of relationship, English language fluency, nativity, and length of residence in the continental United States. Single items also assessed sexual risk behaviors with the reported main female sexual partner during the past 3 months (inconsistent or no condom use during vaginal sexual intercourse; inconsistent or no condom use during anal sexual intercourse; vaginal or anal sexual intercourse, or both, with other female sexual partners; and inconsistent or no condom use with non–main female sexual partners). All these items were created for use in previous research that was conducted with young women of similar demographics who were recruited within the same health center as our study.[22,23]

Forced sexual intercourse without a condom during the past year was assessed with a single item from the Conflict Tactics Scale-2 (CTS-2), a 39-item inventory of abusive behaviors.[24] Fatherhood was assessed with a single question about the number of children respondents had fathered, and 2 variables were created from this question: having fathered any children and having fathered 3 or more children. Having fathered 3 or more children was created as a variable to indicate higher than average fertility in accordance with US Census 2000 data, which showed that the average number of children in both US and Massachusetts households with children was 1.9,[25] and the average number of minor children was close to equivalent across racial/ethnic groups.[26]

Participants' perpetration of physical violence and sexual violence during the past year and ever were assessed with the perpetration items from the CTS-2,[24] which was developed to assess psychological, physical, and sexual aggression by partners who are in dating, cohabitating, or marital relationships. We used the CTS-2 because of its reliability and validity with diverse samples of men and women, including Hispanic and Black men, and with diverse languages, including English and Spanish.[27] The CTS-2 was used in a population-based study of IPV in the United States,[7] and it was used to assess IPV perpetration in a community clinic–based study of US men.[28]

For regression analyses, we summed and dichotomized responses as IPV perpetration or no IPV perpetration during the past year; for descriptive analyses, we summed and dichotomized responses as IPV perpetration ever or never. Consistent with previous research that used this measure across diverse populations,[24,27,28] the CTS-2 showed strong internal reliability with our sample; Cronbach alphas were 0.93 for IPV perpetration during the past year and 0.96 for IPV perpetration ever. The item that assessed forced sexual intercourse without a condom was not included in this scale to allow for assessment of this item as a sexual risk outcome; it is the only item in the scale that assesses a sexual risk behavior.

Data Analyses

Frequencies were generated for IPV perpetration, sexual risk behaviors and fatherhood variables, and demographics. Crude logistic regression analyses assessed the bivariate associations between past-year IPV perpetration and outcome variables, sexual risk behaviors during the past 3 months (unprotected vaginal sexual intercourse with primary partner, unprotected anal sexual intercourse with primary partner, other female sexual partners in addition to primary partner), forced unprotected sexual intercourse during the past year, and fatherhood (having fathered any children, having fathered 3 or more children). We then conducted adjusted logistic regression analyses to assess associations between past-year IPV perpetration and sexual risk behaviors and fatherhood after we adjusted for demographics (age, race/ethnicity, income, continental US nativity, length of residence in the continental United States, and length of relationship). We used adjusted odds ratios (OR) and 95% confidence intervals (CI) to assess significance in final models.

RESULTS

Sample Demographics

The median age of participants was 24 years; 74.9% of participants were Hispanic and 21.9% were Black. The majority of the sample was born in the continental United States (44.5%) or Latin America (53.4%). Almost one third of participants (29.3%) was born in the Dominican Republic; 16.3% were born in Puerto Rico; 7.8% were born in Mexico, South America, Central America, or Cuba. Of those who were not born in the continental United States, 10.2% had lived in the United States for 1 year or less, and 65.0% had lived in the United States for more than 5 years. More than one third (37.5%) were unemployed; 53.4% reported an income of $800 or less per month; and 28.0% did not have a high-school degree or general equivalency diploma (high rates of unemployment and low income and low education level may in part be attributable to the young sample, which likely included high-school students). Approximately 1 in 6 (15.2%) were married; the median length of relationship for the sample was 2 years, and 65% reported having been in their relationship for 1 year or longer.

Sexual Risk Behaviors and Having Fathered Children

Inconsistent or no condom use was reported by the majority who reported vaginal sexual intercourse (80.2%) and anal sexual intercourse (79.2%) with their main female partner. One quarter of participants (24%) reported having forced sexual intercourse without a condom; 16.3% reported engaging in this behavior within the past year. Forty-three percent reported sexual intercourse with a non–main female partner during the past 3 months; 49.2% reported inconsistent or no condom use with these partners. Although sexual intercourse with a male partner was less commonly reported (6.9%) than sexual intercourse with a non–main female partner, 12 of the 19 men who reported sexual intercourse with a male partner also reported sexual intercourse with a non–main female

TABLE 1—Sexual Risk Behaviors and Pregnancy Involvement Among Men in a Steady Relationship With a Female Partner (N = 283)

	%
Condom use during vaginal sexual intercourse with main partner in past 3 months	
None	45.6
Inconsistent	34.6
Consistent	19.8
Condom use during anal sexual intercourse with main partner[a] in past 3 months	
None	61.3
Inconsistent	17.9
Consistent	20.8
Forced sexual intercourse without condom during past year	17.5
Number of non-main female sexual partners during past 3 months	
0	54.1
1	19.1
2 or more	24.0
Number of non-main male sexual partners during past 3 months	
0	85.5
1	3.9
2 or more	3.0
Having fathered children, ever	49.1
Having fathered 3 or more children	16.3

Note. Responses do not add up to 100%, because 2.8% of participants did not respond to the question on number of other female sexual partners in the past 3 months and 7.6% of participants did not respond to the question on male sexual partners in the past 3 months.
[a]Among those who reported anal sexual intercourse (n = 106).

TABLE 2—IPV Perpetration Among Men in a Steady Relationship With a Female Partner (N = 283)

	Past Year, %	Ever, %
Physical IPV perpetration		
Pushed or shoved partner	14.1	23.7
Twisted partner's arm or hair	12.4	18.7
Threw something at partner that could hurt her	11.3	20.8
Grabbed partner	11.0	16.6
Threatened to hit or throw something at partner	8.1	12.7
Slapped partner	7.8	12.7
Slammed partner	6.0	12.4
Punched or hit partner with something that could hurt her	5.7	11.7
Choked partner	4.9	9.9
Used a knife or gun on partner	3.5	7.8
Kicked partner	3.5	7.8
Beat up partner	3.5	7.1
Total physical IPV perpetration	27.6	41.3
Sexual IPV perpetration		
Insisted partner have oral or anal sexual intercourse but did not use physical force	20.5	31.4
Insisted on sexual intercourse when partner did not want to but did not use physical force	14.8	28.6
Used force[a] to make partner have oral or anal sexual intercourse	6.7	9.9
Used threats to make partner have sexual intercourse	5.7	6.7
Used threats to make partner have oral or anal sexual intercourse	4.9	7.1
Used force[a] to make partner have sexual intercourse	4.2	6.7
Total sexual IPV perpetration	28.3	43.8
IPV perpetration resulting in injury/need for medical services		
Partner had small cut, sprain, or bruise due to fight with participant	8.8	16.6
Partner passed out when hit in the head during fight with participant	6.0	6.7
Partner went to doctor due to fight with participant	4.9	8.8
Partner needed to go to doctor due to fight with participant but did not	4.2	7.4
Partner still felt physical pain the next day due to fight with participant	3.9	10.0
Partner had broken bone due to fight with participant	3.5	4.9
Burned or scalded partner on purpose	3.2	6.0
Total IPV perpetration resulting in injury/need for medical services	13.8	22.6

Note. IPV = intimate partner violence.
[a]Force described as "like hitting, holding down, or using a weapon."

partner during the past 3 months. Half of the sample (49.1%) reported having fathered at least 1 child; 16.3% reported having fathered 3 or more children (Table 1).

Partner Violence and Sexual Assault

IPV perpetration of some kind (physical, sexual, injury-related, required medical services) during the past year was reported by 41.3% of the sample; 58.7% reported IPV perpetration ever. Physical abuse of a partner during the past year was reported by 27.6% of the sample; physical abuse of a partner ever was reported by 41.3%. The most common types of reported physical IPV perpetration were pushing or shoving a partner (past year = 14.1%; ever = 23.7%), twisted arm or hair (past year = 12.4%; ever = 18.7%), threw something at partner that could hurt her (past year = 11.3%; ever = 20.8%), and grabbed a partner (past year = 11.0%; ever = 16.6%) (Table 2).

Sexual abuse of a partner during the past year was reported by 28.3% of the sample; sexual abuse of a partner ever was reported by 43.8%. The most common types of reported sexual IPV perpetration were insisting on but not forcing oral or anal sexual intercourse (past year = 20.5%; ever = 31.4%) and insisting on but not forcing sexual intercourse when a partner did not want to have sexual intercourse (past year = 14.8%; ever = 28.6%). One in 10 participants (9.9%) reported a history of having forced a partner to have oral or anal sexual intercourse, and 1 in 16 (6.7%) reporting having forced a partner to have vaginal sexual intercourse.

Partner's injury from, or need for medical services because of, participant's abuse during the past year was reported by 13.8% of the sample; 22.6% reported ever perpetrating

IPV that resulted in their partner's injury or need for medical services. The most common types of reported IPV-related partner injuries or need for medical services during the past year included partner's cut, sprain, or bruise (8.8%) and partner's passing out because of a hit on the head (6.0%); the most common types of partner injuries or need for medical services as a result of participants' IPV ever included partner's cut, sprain, or bruise (16.6%) and partner's pain the day after a fight (10%).

Associations Between IPV and Sexual Risk Behaviors and Having Fathered Children

Crude regression analyses showed that men who reported IPV perpetration during the past year were significantly more likely to report forced sexual intercourse without a condom during the past year (OR=4.6; 95% CI=2.3, 9.3) and sexual intercourse with at least 1 other woman during the past 3 months (OR=2.0; 95% CI=1.2, 9.3). Other assessed outcomes were not significantly associated with IPV perpetration in the crude analyses. Adjusted logistic regression analyses showed that participants who reported IPV perpetration during the past year were significantly more likely to report inconsistent or no condom use during vaginal sexual intercourse (OR_{adj}=2.4; 95% CI=1.1, 4.9) and anal sexual intercourse (OR_{adj}=3.3; 95% CI=1.1, 10.1) during the past 3 months, forcing sexual intercourse without a condom during the past year (OR_{adj}=5.2; 95% CI=2.5, 10.9), sexual intercourse with other women during the past 3 months (OR_{adj}=2.2; 95% CI=1.3, 3.7), and having fathered 3 or more children (OR_{adj}=2.5; 95% CI=1.2, 5.5) (Table 3).

DISCUSSION

Findings from our study show that men who reported IPV perpetration during the past year were more likely than those who did not report such perpetration to engage in risky sexual behaviors with main female partners, including unprotected vaginal and anal sexual intercourse, forced unprotected sexual intercourse, and sexual intercourse with other women. These findings among lower-income urban men support previous work that has documented higher rates of sexual infidelity and unprotected anal sexual intercourse among men who were recruited from a methadone treatment facility and who reported IPV.[21] Overall, these findings show a notable association between IPV perpetration and sexual risk behaviors among young men, and they support previous studies with women that suggested abusive male partners may pose greater STD/HIV risk to women compared with nonabusive men.[12,23]

A novel finding from our study is that male perpetrators of IPV were more likely to report having fathered 3 or more children compared with those who reported no IPV during the past year. Quantitative research with women has documented associations between IPV and unwanted and rapid repeat pregnancies,[29–32] and qualitative research has documented a link between IPV and forced pregnancy.[19,20] Hence, these findings from studies with women suggest that a greater number of offspring by abusive men may be a consequence of these men blocking their female partners' reproductive control. However, our findings did not directly assess forced pregnancy; thus, it remains unclear as to why young men who reported IPV perpetration were more likely to have fathered a greater number of children. This issue warrants further exploration and should include an examination of whether men are more likely to report having a greater number of children within the context of an abusive relationship, particularly because of the evidence that there is an association between women's IPV experiences and poorer maternal and child health outcomes.[33–44]

Although further research with larger and more generalizable samples is needed to confirm our findings, additional study also is needed to clarify why these findings may exist. There is some evidence that young men's traditional masculine gender role ideologies—particularly ideas about male hypersexuality, impregnation as a sign of masculinity, and adversarial heterosexual dyadic norms—are associated with IPV perpetration, unprotected sex, and multiple sex partners.[45–48] Larger-scale research with diverse samples is needed to understand the extent to which and how masculine gender role ideologies may be associated with men's perpetration of IPV and sexual risk behaviors within steady relationships with female partners. Understanding such associations will be critical to developing effective prevention programming in this area.

Although findings from our study show an association between IPV perpetration and sexual risk behaviors among young men, crude analyses did not yield significant findings for either unprotected vaginal and anal sexual intercourse or having fathered 3 or more children. Only adjusted analyses showed significant findings for these variables, which indicates that demographics may obscure the association between IPV and some sexual risk

TABLE 3—Logistic Regression Analyses, Crude and Adjusted for Demographics, to Assess Associations Between IPV Perpetration During the Past Year and Sexual Risk Behaviors and Fatherhood

Sexual Risk Behaviors or Fatherhood	OR (95% CI)	OR_{adj} (95% CI)[a]
Inconsistent or no condom use during vaginal sexual intercourse with main partner in past 3 months	1.6 (0.9, 3.0)	2.4 (1.1, 4.9)
Inconsistent or no condom use during anal sexual intercourse with main partner in past 3 months	2.0 (0.8, 5.3)	3.3 (1.1, 10.1)
Forced sexual intercourse without a condom during past year	4.6 (2.3, 9.3)	5.2 (2.5, 10.9)
Sexual intercourse with other women during past 3 months	2.0 (1.2, 9.3)	2.2 (1.3, 3.7)
Having fathered children, ever	0.8 (0.5, 1.4)	0.9 (0.5, 1.5)
Having fathered 3 or more children	1.6 (0.8, 3.0)	2.5 (1.2, 5.5)

Note. IPV = intimate partner violence; OR = odds ratio; CI = confidence interval.
Fatherhood is defined as having fathered children.
[a]Adjusted for age, Hispanic ethnicity, income, length of continental US residency, and length of relationship.

behaviors. Even for those sexual risk behaviors that were significantly associated with IPV perpetration in the crude analyses (i.e., forced unprotected sexual intercourse and sexual intercourse with other women), the point estimate changed notably between crude and adjusted analyses. Our findings are consistent with findings from previous racially/ethnically diverse population-based research of sexual risk behaviors that showed age, relationship status, and cultural factors are major correlates of men's sexual risk behaviors.[49]

Although the associations between IPV and sexual risk behaviors and fatherhood are notable, the importance of these findings is amplified by the pervasiveness of IPV perpetration that was reported by our sample. More than half of our participants (59%) reported that they had perpetrated IPV against a female partner at some point in their lifetime. A previous study of IPV in a health care setting identified a 14% past-year IPV perpetration prevalence rate[28]; in contrast, 41% of our health center sample reported IPV perpetration during the past year. Higher rates of IPV among our sample compared with the previous study of a health center sample is likely a consequence of our sample being younger and urban, i.e., demographic groups that have an elevated risk for IPV perpetration.[50]

Limitations

A major limitation to our study is generalizability of study findings, which is exemplified by our substantially higher rate of IPV perpetration compared with the previous study.[28] Use of a single community health center that serves predominantly lower-income Hispanic and Black men in an urban area within the Northeast likely limits generalizability of findings to other populations. Furthermore, although our health center is typical of other urban community health centers within the region in terms of its location in a lower-income area and its predominantly racial/ethnic minority and lower-income client population, it reaches a larger segment of Hispanic immigrants than many other health centers. Additionally, our study included men who either sought care at the health center or accompanied others; therefore, our findings cannot be generalized to those who sought care.

In addition to generalizability limitations, there are a number of study design limitations. Our research was cross-sectional; thus, causality cannot be inferred from the findings. Reliance on self-reported data made our data subject to social desirability and recall biases, and lack of data from female partners further inhibited verification of the self-reports. However, these biases would likely result in underreporting rather than overreporting of sensitive issues, such as perpetration of IPV, unprotected sexual intercourse, and sexual infidelity. Because of the nature of the questions, we were unable to assess whether the reported sexual risk behaviors and fatherhood occurred within the context of an abusive relationship. A previous study with an antenatal clinic-based sample of young women in South Africa found that abusive men were more likely than nonabusive men to infect female partners with HIV,[12] which suggests that sexual risk behaviors occur within the context of abusive relationships. Longitudinal study of these issues with men and heterosexual couples is needed; future research also must include relationship-specific assessments about sexual risk behaviors and IPV to more directly assess these associations.

Conclusions

Male perpetrators of recent IPV were more likely than other men to have engaged in risky sexual behaviors and to have fathered 3 or more children, which placed these men and their partners at increased risk for STD/HIV. High rates of having fathered children among abusive men was consistent with reported lack of reproductive control among abused women,[19,20,29–32] and thus must be further explored to both understand and address these associations. Our findings support previous research with women that documented higher rates of sexual risk behaviors among abusive male partners, which showcases the need for interventions that integrate IPV and STD/HIV prevention.

The high rates of IPV and sexual risk behaviors in our sample also show that community health centers may be an important venue for reaching men who are at risk for both IPV perpetration and STD/HIV. Previous studies have recommended screening and referral for IPV perpetration among clinic-based samples of men[28,51–53] and HIV interventions for men in urban health care settings.[54,55] However, clinic-based interventions that integrate IPV and STD/HIV prevention among US men are absent from published literature. These interventions must be developed and evaluated, because IPV and STD/HIV are important public health issues.

About the Authors

Anita Raj and M. Christina Santana are with the Department of Social and Behavioral Sciences, Boston University School of Public Health, Mass. Ana La Marche is with the Martha Eliot Health Center, Boston. Hortensia Amaro is with the Bouve College of Health Sciences, Northeastern University, Boston. Kevin Cranston is with the HIV/AIDS Bureau, Department of Public Health, Boston. Jay G. Silverman is with the Department of Society, Human Development, and Health and the Division of Public Health Practice, Harvard University School of Public Health, Boston.

Requests for reprints should be sent to Anita Raj, PhD, Boston University School of Public Health, 715 Albany St, T2W, Boston, MA 02118 (e-mail: anitaraj@bu.edu).

This article was accepted November 9, 2005.

Contributors

A. Raj originated the study, wrote the paper, and analyzed the data; she also was principal investigator of the study from which these data were obtained. M.C. Santana oversaw all data collection and assisted with writing the paper and interpreting study findings. A. La Marche, H. Amaro, and K. Cranston assisted with interpreting study findings and developing the discussion section of this paper. J. Silverman assisted with originating the study, writing the paper, and analyzing the data; he also provided IPV expertise to the study from which these data were obtained.

Acknowledgments

This project was funded through a grant from the Massachusetts Department of Public Health. We would like to thank all the staff and providers at the Martha Eliot Health Center who helped facilitate study recruitment for this project; we would particularly like to thank David Holder, Catherine MacAuley, and Ana Ortiz for their support and guidance in implementing the study at the Martha Eliot Health Center.

Human Participant Protection

This study was approved by the institutional review boards of the Boston University Medical Center and the Children's Hospital of Boston.

References

1. Centers for Disease Control and Prevention. Lifetime and annual incidence of intimate partner violence and resulting injuries—Georgia, 1995. *MMWR Morb Mortal Wkly Rep.* 1998;47:849–853.

2. Centers for Disease Control and Prevention. Intimate partner violence among men and women—South Carolina, 1998. *MMWR Morb Mortal Wkly Rep.* 2000; 49:691–694.

3. Centers for Disease Control and Prevention. Prevalence of intimate partner violence and injuries—Washington, 1998. *MMWR Morb Mortal Wkly Rep.* 2000;49:589–592.

4. Plichta SB. Violence, health and use of health services. In: Falik MM, Collins KS, eds. *Women's Health and Care Seeking Behavior: The Commonwealth Fund Survey.* Baltimore, Md: Johns Hopkins University Press; 1996:237–270.

5. Schafer J, Caetano R, Cook CL. Rates of intimate partner violence in the United States. *Am J Public Health.* 1998;88:1702–1704.

6. Straus MA, Gelles RJ. Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys. *J Marriage Fam.* 1986;48:465–479.

7. Tjaden P, Thoennes N. Full report of the prevalence, incidence, and consequences of violence against women: findings from the National Violence Against Women Survey. Available at: http://www.ncjrs.org/pdffiles/172837.pdf. Accessed May 26, 2005.

8. Amaro H, Raj A. Theoretical and measurement issues in the study of women's relational power in HIV risk reduction. *Sex Roles.* 2000;42:723–749.

9. Heise L, Ellsberg M. Ending violence against women. Available at: http://www.infoforhealth.org/pr/l11edsum.shtml. Accessed May 26, 2005.

10. Schmuel E, Schenker JG. Violence against women: the physician's role. *Eur J Obstet Gynecol Reprod Biol.* 1998;80:239–245.

11. Bauer HM, Gibson P, Hernandez M, Kent C, Klausner J, Bolan G. Intimate partner violence and high-risk sexual behaviors among female patients with sexually transmitted diseases. *Sex Transm Dis.* 2002; 29:411–416.

12. Dunkle KL, Jewkes RK, Brown HC, Gray GE, McIntryre JA, Harlow SD. Gender-based violence, relationship power, and risk of HIV infection in women attending antenatal clinics in South Africa. *Lancet.* 2004;363:1415–1421.

13. He H, McCoy HV, Stevens SJ, Stark MJ. Violence and HIV sexual risk behaviors among female sex partners of male drug users. *Women Health.* 1998;27:161–75.

14. Rickert VI, Wiemann CM, Harrykissoon SD, Berenson AB, Kolb E. The relationship among demographics, reproductive characteristics, and intimate partner violence. *Am J Obstet Gynecol.* 2002;187: 1002–1007.

15. Roberts TA, Auinger P, Klein JD. Intimate partner abuse and the reproductive health of sexually active female adolescents. *J Adolesc Health.* 2005;36:380–385.

16. Silverman JG, Raj A, Mucci LA, Hathaway JE. Dating violence against adolescent girls and associated substance use, unhealthy weight control, sexual risk behavior, pregnancy, and suicidality. *JAMA.* 2001;286: 572–579.

17. Silverman JG, Raj A, Clements K. Dating violence and sexual risk in a representative sample of high school students. *Pediatrics.* 2004;114:e220–225.

18. Wingood GM, DiClemente RJ. Consequences of having a physically abusive partner on the condom use and sexual negotiation practices of young adult African-American women. *Am J Public Health.* 1997; 87:1016–1018.

19. Hathaway JE, Willis G, Zimmer B, Silverman JG. Impact of partner abuse on women's reproductive lives. *J Am Med Women's Assoc.* 2005;60:42–45.

20. Raj A, Liu R, McCleary-Sills J, Silverman JG. South Asian victims of intimate partner violence more likely than non-victims to report sexual health concerns. *J Immigr Health.* 2005;7:85–91.

21. El-Bassel N, Fontdevila J, Gilbert L, Voisin D, Richman BL, Pitchell P. HIV risks of men in methadone maintenance programs who abuse their intimate partners: a forgotten issue. *J Subst Abuse.* 2001;12:29–43.

22. Raj A, Amaro H, Cranston K, Martin B, Cabral H, Navarro A, Conron K. Is a general health program as effective as an HIV program in reducing HIV risk among Latinas? *Pub Health Rep.* 2002;116:599–607.

23. Raj A, Silverman JG, Amaro H. Abused women report greater male partner risk and gender-based risk for HIV: findings from a community-based study with Hispanic women. *AIDS Care.* 2004;16:519–529.

24. Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The revised Conflict Tactics Scales (CTS2). *J Fam Issues.* 1996;17:283–316.

25. US Census 2000. Table ST-FI-2000. Average number of children per family and per family with children, by state. Available at: http://www.census.gov/population/socdemo/hh-fam/tabst-f1-2000.pdf. Accessed October 24, 2005.

26. US Census 2000. TableAVG2. Average number of people per family household by race and Hispanic origin, marital status, age and education of householder. Available at: http://www.census.gov/population/socdemo/hh-fam/cps2001/tabavg2.pdf. Accessed October 24, 2005.

27. Straus M. Cross-cultural reliability and validity of the Revised Conflict Tactics Scales: a study of university student dating couples in 17 nations. *Cross-Cult Res.* 2005;38:407–432.

28. Oriel KA, Fleming MF. Screening men for partner violence in a primary care setting. A new strategy for detecting domestic violence [see comment]. *J Fam Pract.* 1998;46:493–498.

29. Pallitto CC, O'Campo P. Relationship between intimate partner violence and unintended pregnancy: analysis of a national sample from Colombia. *Intl Fam Plan Persp.* 2004;30:165–173.

30. Hillard PJA. Physical abuse in pregnancy. *Obstet Gynecol.* 1985;66:185–190.

31. Stewart DE, Cecutti A. Physical abuse in pregnancy. *CMAJ.* 1993;149:1257–1263.

32. Jacoby M, Gorenflo D, Black E, Wunderlich C, Eyler AE. Rapid repeat pregnancy and experiences of interpersonal violence among low-income adolescents. *Am J Prev Med.* 1999;16:318–321.

33. Connelly CD, Newton RR, Landsverk J, Aarons GA. Assessment of intimate partner violence among high-risk postpartum mothers: concordance of clinical measures. *Women Health.* 2000;31:21–37.

34. Neggers Y, Goldenberg R, Cliver S, Hauth J. Effects of domestic violence on preterm birth and low birthweight. *Acta Obstet Gynecol Scand.* 2004;83:455–460.

35. Altarac M, Strobino D. Abuse during pregnancy and stress because of abuse during pregnancy and birthweight. *J Am Med Womens Assoc.* 2002;57:208–214.

36. Boy A, Salihu HM. Intimate partner violence and birth outcomes: a systematic review. *Int J Fertil.* 2004; 49:159–163.

37. Campbell J, Jones AS, Dienemann J, et al. Intimate partner violence and physical health consequences. *Arch Intern Med.* 2002;162:1157–1163.

38. Campbell JC, Woods AB, Chouaf KL, Parker B. Reproductive health consequences of intimate partner violence. A nursing research review. *Clin Nurs Res.* 2000;9:217–237.

39. Cokkinides VE, Coker AL, Sanderson M, Addy C, Bethea L. Physical violence during pregnancy: maternal complications and birth outcomes. *Obstet Gynecol.* 1999;93:661–666.

40. Covington DL, Hage M, Hall T, Mathis M. Preterm delivery and the severity of violence during pregnancy. *J Reprod Med.* 2001;46:1031–1039.

41. Curry MA, Perrin N, Wall E. Effects of abuse on maternal complications and birthweight in adult and adolescent women. *Obstet Gynecol.* 1998;92:530–534.

42. Grimstad H, Schei B, Backe B, Jacobsen G. Physical abuse and low birthweight: a case-control study. *Br J Obstet Gynecol.* 1997;104:1281–1287.

43. Jagoe J, Magann EF, Chauhan SP, Morrison JC. The effects of physical abuse on pregnancy outcomes in a low-risk obstetric population. *Am J Obstet Gynecol.* 2000;182:1067–1069.

44. Martin SL, Mackie L, Kupper LL, Buescher PA, Moracco KE. Physical abuse of women before, during and after pregnancy. *JAMA.* 2001;285:1581–1584.

45. Pleck JH, O'Donnell LN. Gender attitudes and health risk behaviors in African-American and Latino early adolescents. *Matern Child Health J.* 2001;5: 265–272.

46. Pleck JH, Sonenstein FL, Ku LC. Masculinity ideology: its impact on adolescent males' heterosexual relationships. *J Soc Issues.* 1993;49:11–29.

47. Shearer CL, Hosterman SJ, Gillen MM, Lefkowitz ES. Are traditional gender role attitudes associated with risky sexual behavior and condom-related beliefs. *Sex Roles.* 2005;52:311–324.

48. Moore TM, Stuart GL. A review of the literature on masculinity and partner violence. *Psychol Men Masculinity.* 2005;6:46–61.

49. Catania J, Coates T, Kegeles S, et al. Condom use in multi-ethnic neighborhoods of San Francisco: the population-based AMEN (AIDS in Multi-Ethnic Neighborhoods) study. *Am J Public Health.* 1992;82: 284–287.

50. Gorman-Smith D, Tolan PH, Sheidow AJ, Henry DB. Partner violence and street violence among urban adolescents: do the same family factors relate? *J Res Adolesc.* 2001;11:273–295.

51. Chelmowski M, Hamberger LK. Screening men for domestic violence in your medical practice. *Wis Med J.* 1994;93:623–626.

52. Conway T. The internist's role in addressing violence. *Arch Intern Med.* 1996;156:931–956.

53. Mintz HA, Cornett FW. When your patient is a batterer: what you need to know before treating perpetrators of domestic violence. *Postgrad Med.* 1997;101: 219–228.

54. Kalichman SC, Rompa D, Coley B. Lack of positive outcomes from a cognitive-behavioral HIV and AIDS prevention intervention for inner-city men: lessons from a controlled pilot study. *AIDS Educ Prev.* 1997;9:299–313.

55. Kalichman SC, Cherry C, Browne-Sperling F. Effectiveness of a video-based motivational skills-building HIV risk-reduction intervention for inner-city African American men. *J Consult Clin Psychol.* 1999;67: 959–966.

Medical Teacher, Vol. 28, No. 1, 2006, pp. 19–23

Taylor & Francis
Taylor & Francis Group

TWELVE TIPS

Twelve tips to promote excellence in medical teaching

SUBHA RAMANI

Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, USA

ABSTRACT *For medical teachers around the world, teaching duties have expanded beyond the classroom and include teaching small groups, assessment, providing instructional materials beyond the syllabus, problem-based learning, learner-centred teaching, clinical teaching on-the-fly—and the list goes on. Faculty development is essential to train medical faculty in essential educational theory and specific teaching skills as well as to encourage a flexible and learner-centred approach to teaching. Finally, self-reflection and critique of teaching techniques are vital to propel medical schools towards promoting and aiming for uncompromising excellence in medical education. The twelve tips described in this article relating to educating teachers, evaluating teaching and eradicating institutional apathy are simple measures that educational leaders can apply to promote excellence in teaching at their parent institutions. The tips introduce a multi-dimensional approach to improving the overall quality of medical education consisting of measures aimed at individual teachers and those aimed at overhauling the teaching climate at medical institutions.*

Introduction

It has been stated that the majority of academic faculty are not formally trained, therefore not qualified in one of their primary responsibilities: teaching (Cannon & Widodo, 1994). Medical school faculty are assuming multiple teaching roles: teacher, administrator, lecturer, small-group facilitator, assessor, role model to name a few (Harden & Crosby, 2000). To help faculty succeed at these teaching tasks, faculty development is key (Wilkerson & Irby, 1998). Reducing faculty burnout in the midst of numerous career responsibilities is another reason to provide faculty with opportunities to improve their teaching skills (Pololi *et al.*, 2001). The British General Medical Council, in a recent edition of its publication Tomorrow's Doctors, includes the following attributes of a medical practitioner (General Medical Council, 2002):

- recognition of the obligation to teach others, particularly doctors in training;
- recognition that teaching skills are not necessarily innate but can be learned;
- recognition that the example of the teacher is the most powerful influence upon the standards of conduct and practice of trainees.

The following twelve tips describe simple steps that can help committed medical educators introduce measures to improve the quality of teaching, and integrate an evidence-based approach and scholarship into their framework of medical education thus promoting overall excellence in teaching at their parent institutions.

Tip 1: Outcome-based education

Establish explicit learning and teaching outcomes to help teachers plan their teaching

Teachers should be informed of and have easy access to written learning outcomes for their courses so that they can plan their teaching strategies and methods. Simultaneously, there should be specific teaching outcomes expected of medical educators so that their teaching can be evaluated using objective criteria. In order that objective criteria can be laid down for promotion of educators, we need to establish the outcomes for clinical teachers that can be monitored and documented. The expected competences of clinical teachers can be modelled on the three-circle Dundee model (Harden *et al.*, 1999). This model serves as an excellent substrate and outcomes can be modified easily for different institutional needs.

Tip 2: Implement best evidence medical education (BEME)

Make the link between evidence-based medical practice and evidence-based medical education

Around the world there are calls for moving away from opinion-based teaching to evidence-based teaching (BEME group, 2000; Harden *et al.*, 2000). Medical education should be subject to the same strict standards of scientific scrutiny as patient care and research (Hart & Harden, 2000). The scientific criteria used to judge medical research should apply to medical education as well (Norman, 2000; van der Vleuten *et al.*, 2000) and if evidence is not available we should be willing to research the area

Correspondence: Subha Ramani, MD MPH MMedEd, Boston University School of Medicine, Section of General Internal Medicine, Department of Medicine, 715 Albany Street, EB33A, Boston, MA 02118, USA. Tel: 617-638-7985; fax: 617-414-4676; email: sramani@bu.edu

ISSN 0142-159X print/ISSN 1466-187X online/06/010019–5 © 2006 Taylor & Francis
DOI: 10.1080/01421590500441786

(Davies, 1999). Evidence-based medicine seminars can educate medical teachers in the fundamental principles of appraisal of quantitative and qualitative research and serve as a good foundation for further review of educational literature. Wolf has summarized the lessons learned from evidence-based medicine and suggested ways in which evidence-based education could benefit from them (Wolf, 2000). The QUESTS dimensions include quality, utility, extent, strength, target or validity and the setting or relevance of the available evidence (Harden *et al.*, 2000). These dimensions can guide educators to critically appraise published educational methods, decide whether the published results are applicable to their teaching situation and if their educational strategies need to be changed in light of the evidence.

Tip 3: Journal clubs and review of existing literature

Journal clubs for faculty to critically appraise articles on medical education

Although journal clubs have been held for decades to teach faculty and trainees critical appraisal of medical literature, similar sessions for critical appraisal of literature on medical education are still rare. Educational journal clubs can expose faculty to educational literature just as they are traditionally used to promote critical appraisal of research literature in medicine. Guidelines have been written for systematic searching and review of educational literature (Haig & Dozier, 2003a, 2003b). During these journal clubs, individual faculty can select an article on a specific area of medical education, review the literature for their colleagues, critically appraise the research methodology and discuss whether the results are applicable to their own setting. Such sessions will also elevate the scientific importance of educational literature within institutions.

Tip 4: Faculty development

Faculty development is an important instrument to create a positive learning environment for teachers

Experts in higher education have made several recommendations to elicit staff needs when designing faculty development programmes (Hitchcock *et al.*, 1993; Steinert, 2000). Examples of such steps include: developing a specific mission for the programme based on the needs of the institution and individual faculty members, identifying institutional or departmental strengths and weaknesses, consulting experts in designing programmes, involving faculty in planning the programme and appointing an effective leader for the project. Carroll (1993) has suggested the following universal considerations in designing programmes for higher education faculty based on Knowles's adult learning principles (Knowles and Associates, 1984):

- Faculty need to know why they should learn something.
- Faculty should demonstrate self-directed learning.
- Faculty possess experience that can be incorporated into the learning resources.
- Faculty development should be task centred.

Institutions need to plan faculty development programmes that will recognize the limited time available to clinical faculty to participate in professional development programmes and maximize the effectiveness given the time constraints (Gelula & Yudkowsky, 2002).

Tip 5: Evaluation of teaching

Tell the teachers how they taught

The evaluation of clinical teaching can be an important source of support and motivation for teachers. In the context of faculty development, teachers can be evaluated before and after participation in such programmes to assess their impact on the teaching skills of faculty. Teaching evaluation can also provide evidence that programme goals are being achieved, teaching standards are being met or exceeded and allow for reflection on, and evolution of, the curriculum. It is important to follow the basic 'rules' of any assessment method: clear goals, high levels of validity, reliability, efficiency and feasibility. Where possible, more than one source of information should be used. The evaluation may involve looking at the perceptions of teaching, the actual teaching process, or the 'product' or a measurable outcome of teaching (Snell *et al.*, 2000). Methods of evaluation could include:

- trainee evaluation of teachers using standardized teacher-rating forms (Snell *et al.*, 2000) or focus-group interviews (Fontana & Frey, 1994). The feedback can be used to make decisions on promotions and allocating teaching responsibilities within departments (Copeland & Hewson, 2000);
- self-assessment by teachers using variables such as teaching effectiveness, professional effectiveness other than teaching and enjoyment of teaching;
- peer reviews of teaching, which can provide informed, valuable and diagnostic evaluation of the clinical teacher (Irby, 1983; Horowitz *et al.*, 1998; Beckman *et al.*, 2004). Videotaping allows for subsequent analysis by an educational consultant, by peer physicians or by groups of faculty, residents and students (Stanley & Wright, 1981);
- simulation of teaching encounters. This has been used to assess teaching skills (Gelula & Yudkowsky, 2002). An OSTE, or objective structured teaching encounter, with a simulated learners or peer observers can be used to explore a variety of one-on-one or group clinical teaching activities;
- teaching portfolios. Faculty members can document their teaching activities, evaluations and curriculum and evaluation development in a teaching dossier or portfolio (Edgerton *et al.*, 1991).

Tip 6: Evaluate impact of teaching

Find out how the teaching affected learners

Determining the impact of effective teaching is a more difficult task as several factors other than the teaching alone may influence trainee performance (Snell *et al.*, 2000, Steinert, 2000, Steinert *et al.*, 2003). This impact may be measured as educational outcomes (e.g. student learning), practice outcomes (e.g. a change in trainee practice) or health outcomes (e.g. an effect on patient or population health).

Institutions can also examine whether current teaching methods have a positive impact on their learners in all three domains of knowledge, skills and attitudes. This can be evaluated using the following methods:

- OSCE (objective structured clinical examination) before and after the faculty development programme. If there is a positive change in the scores, it can be extrapolated that one of the influences of this change is more effective teaching;
- observed assessment of students during real patient encounters using specific defined objectives. The observations would be carried out by faculty or peers in the context of daily clinical care;
- questionnaires given to learners to explore their satisfaction with the educational programme and teaching effectiveness. Such questionnaires can be completed before and after implementation of the faculty development programme to compare changes in learning attitudes and satisfaction.

Tip 7: Mentoring

Appoint a panel of senior educators as mentors for junior faculty

The importance of mentoring throughout one's career has been emphasized, especially during professional transitions (Bligh, 1999; Levy *et al.*, 2004). Studies have shown that faculty members who identified a mentor felt more confident, were more likely to have a productive research career, and reported greater career satisfaction (Palepu *et al.*, 1998; Ramanan *et al.*, 2002). Other reported benefits for mentees include: socialization into the profession; help with choice and fulfilment of career path; meaningful involvement in academic activities; and the development of close collaborative relationships (Pololi *et al.*, 2002). A panel of senior educators can be established whose roles would include but not limited to:

- junior faculty mentoring on professional growth and development;
- advising their mentees regarding other internal and external resources, and helping them network with other educators within or outside the institution;
- reviewing their learner evaluations with them;
- facilitating self-reflection and formulation of future goals;
- observing them in their teaching interactions and providing feedback;
- helping them navigate the confusing structure of a very large department thus ensuring a group of dedicated clinicians who will become the forerunners of scientific educational research in the department.

Tip 8: Institutional funding for educational research and development

Encourage educational research by awarding small departmental or institutional grants

Institutions should be willing to make provision in their budget for start-up funds to encourage faculty to engage in educational projects (Steinert, 2000). Many institutions with established faculty development or teaching scholars' programmes provide small grants to encourage their scholars to conduct educational research. They should also help such educators with their research by encouraging affiliations with external educators and other researchers and assisting them in grant writing.

Tip 9: Promote an institutional culture that values teaching highly

Elevate the importance of teaching equivalent to that of research

The culture in academic medical institutions has to change to one where teachers are shown that education is valued. Currently, the culture is clearly one where basic science research and clinical research are held in higher esteem and the value units are grants and publications. This mentality needs to be altered so that clinical teachers feel they are being valued at the same level. The climate should promote self-reflection and self-development of teachers and provide incentives for educators who wish to participate in faculty development programmes. Ultimately, for patients to be benefited by spectacular scientific advances the frontline physicians—namely the trainees—need to be exposed to high-quality educators.

Tip 10: Rewarding excellence in teaching

Reward teachers who improve their teaching techniques based on best available evidence

An ideal institutional climate should reward positive changes in teaching behaviours (Nieman *et al.*, 1997). Faculty whose teaching is further shown to have a positive impact on learners could be rewarded in any of the following ways:

- monetary rewards;
- certificates of honour;
- teaching awards;
- promotion up the academic ladder.

This would inspire all educators to work hard to improve their own teaching techniques and engage in educational projects using an evidence-based approach.

Tip 11: Recognize scholarship in teaching including best evidence medical education

Introduce all medical educators to the scholarship of teaching

Institutions and departments should apply scholarship criteria (Glassick, 2000) to create a distinct academic track for clinician educators. Academic advancement should be based on their teaching impact, dissemination of new curricula and self-reflective teaching. Teachers should feel that they have an equal chance of being promoted within the institution to that of their investigator colleagues. Requirements of new clinical faculty applying for an educator track position should be clear so that they know what is expected of them and how they will be evaluated (Fincher *et al.*, 2000). This can help them establish individual goals and plan their professional activities. Departments should appoint a few faculty formally trained in medical education

with advanced degrees. These educators can help guide faculty development, faculty evaluation, programme evaluation and educational research.

Tip 12: Participate in the BEME collaboration

Join the BEME collaboration and participate in topic reviews

Institutions should encourage their faculty to join the international BEME collaboration. Educators should be assisted in collaborating with others outside their institution and form topic review groups to study important areas in medical education. Topics that need to be systematically reviewed were ranked and listed at an annual meeting of the Society of Directors of Research in Medical Education (Wolf *et al.*, 2001). The highest ranked topics could be categorized under four major areas: curricular design, learning and instructional methods, testing and assessment, and outcomes. BEME is gaining momentum with growing numbers of people becoming involved as well as an increased number of pertinent workshops, publications and websites.

Conclusions

Although the move towards educator development and best evidence medical education is gathering momentum worldwide, there are several medical schools that are still unaware of this movement. At many institutions teaching remains intuition based and opinion based and teachers feel that teaching duties are simply added on to an already heavy workload. Senior educators in leadership positions need to create a climate emphasizing the importance of excellent teaching, and reward their teachers for high-quality teaching, scholarship in teaching and self-development efforts in education. They also need to establish staff development initiatives in teaching and assessment to help their teachers to be reflective educators. The twelve tips described in this article are some techniques teaching institutions can use to change their value systems to ratchet up the importance of teaching in academe, a movement that will, it is hoped, gather momentum and be self-sustaining in the long run.

Notes on contributor

SUBHA RAMANI, MD MPH MMedEd is a general internist with a major interest in medical education. She is currently an Assistant Professor in the Department of Medicine at the Boston University School of Medicine.

References

BECKMAN, T.J., GHOSH, A.K., COOK, D.A., ERWIN, P.J. & MANDREKAR, J.N. (2004) How reliable are assessments of clinical teaching?: a review of the published instruments, *Journal of General Internal Medicine*, 19, pp. 971–977.

BEME GROUP (2000) Best Evidence Medical Education (BEME): report of meeting—3–5 December 1999, London, UK, *Medical Teacher*, 22, pp. 242–245.

BLIGH, J. (1999) Mentoring: an invisible support network, *Academic Medicine*, 2002, pp. 377–384.

CANNON, R.A. & WIDODO, S. (1994) Improving the quality of teaching and learning in Indonesian universities: issues and challenges, *Higher Education Research and Development*, 13, pp. 99–110.

CARROLL, R.G. (1993) Implication of adult learning theories for medical school faculty development programmes, *Medical Teacher*, 15, pp. 163–170.

COPELAND, H.L. & HEWSON, M.G. (2000) Developing and testing a clinical teaching effectiveness instrument in an academic medical centre, *Academic Medicine*, 75, pp. 161–166.

DAVIES, P. (1999) What is evidence-based education?, *British Journal of Educational Studies*, 47, pp. 108–121.

EDGERTON, R., HUTCHINGS, P. & QUINLAN, K. (1991) *The Teaching Portfolio: Capturing the Scholarship in Teaching* (Washington, DC, American Association for Higher Education).

FINCHER, R.E., SIMPSON, D.E., MENNIN, S.P., ROSENFELD, G.C., ROTHMAN, A., MCGREW, M.C., HANSEN, P.A., MAZMANIAN, P.E. & TURNBULL, J.M. (2000) Scholarship in teaching: an imperative for the 21st century, *Academic Medicine*, 75, pp. 887–894.

FONTANA, A. & FREY, J.H. (1994) Interviewing: the art of science, in: N.K. DENZIN & Y.S. LINCOLN (Eds) *Handbook of Qualitative Research* (Thousand Oaks, CA, Sage Publications).

GELULA, M.H. & YUDKOWSKY, R. (2002) Microteaching and standardized students support faculty development for clinical teaching, *Academic Medicine*, 77, p. 941.

GENERAL MEDICAL COUNCIL (2002) *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education*, Version 2 (London, General Medical Council).

GLASSICK, C.E. (2000) Boyer's Expanded Definitions of Scholarship, the standards for assessing scholarship, and the elusiveness of the scholarship of teaching, *Academic Medicine*, 75, pp. 877–880.

HAIG, A. & DOZIER, M. (2003a) BEME Guide No 3: systematic searching for evidence in medical education—Part 1: sources of information, *Medical Teacher*, 25, pp. 352–363.

HAIG, A. & DOZIER, M. (2003b) BEME Guide No 3: systematic searching for evidence in medical education— Part 2: constructing searches, *Medical Teacher*, 25, pp. 463–484.

HARDEN, R.M. & CROSBY, J.R. (2000) AMEE guide No 20: the good teacher is more than a lecturer—the twelve roles of the teacher, *Medical Teacher*, 22, pp. 334–347.

HARDEN, R.M., CROSBY, J.R., DAVIS, M.H. & FRIEDMAN, M. (1999) AMEE guide No 14: outcome-based education: part 1—An introduction to outcome-based education, *Medical Teacher*, 21, pp. 7–14.

HARDEN, R.M., GRANT, G., BUCKLEY, G. & HART, I.R. (2000) Best evidence medical education, *Advances in Health Sciences Education*, 5, pp. 71–90.

HART, I.R. & HARDEN, R.M. (2000) Best evidence medical education (BEME): a plan for action, *Medical Teacher*, 22, pp. 131–135.

HITCHCOCK, M.A., STRITTER, F.T. & BLAND, C.J. (1993) Faculty development in the health professions: conclusions and recommendations, *Medical Teacher*, 14, pp. 295–309.

HOROWITZ, S., VAN EYCK, S. & ALBANESE, M. (1998) Successful peer review of courses: a case study, *Academic Medicine*, 73, pp. 266–271.

IRBY, D.M. (1983) Peer review of teaching in medicine, *Journal of Medical Education*, 58, pp. 457–461.

KNOWLES, M.S. AND ASSOCIATES (1984) *Andragogy in Action: Applying Modern Principles of Adult Learning* (San Francisco, Jossey-Bass).

LEVY, B.D., KATZ, J.T., WOLF, M.A., SILLMAN, J.S., HANDIN, R.I. & DZAU, V.J. (2004) An initiative in mentoring to promote residents' and faculty members' careers, *Academic Medicine*, 79, pp. 845–850.

NIEMAN, L.Z., DONOGHUE, G.D., ROSS, L.L. & MORAHAN, P.S. (1997) Implementing a comprehensive approach to managing faculty roles, rewards, and development in an era of change, *Academic Medicine*, 72, pp. 496–504.

NORMAN, G.R. (2000) Reflections on BEME, *Medical Teacher*, 22, pp. 141–144.

PALEPU, A., FRIEDMAN, R.H., BARNETT, R.C., CARR, P.L., ASH, A.S., MOSKOWITZ, M.A. & SZALACHA, L. (1996) Medical faculty with mentors are more satisfied, *Journal of General Internal Medicine*, 1, p. 107.

POLOLI, L., CLAY, M.C., LIPKIN, M. JR, HEWSON, M., KAPLAN, C. & FRANKEL, R.M. (2001) Reflections on integrating theories of adult education into a medical school faculty development course, *Medical Teacher*, 23, pp. 276–283.

POLOLI, L.H., KNIGHT, S.M., DENNIS, K. & FRANKEL, R.M. (2002) Helping medical school faculty realize their dreams: an innovative, collaborative mentoring program, *Academic Medicine*, 77, pp. 377–384.

RAMANAN, R.A., PHILLIPS, R.S., DAVIS, R.B., SILEN, W. & REEDE, J.Y. (2002) Mentoring in medicine: keys to satisfaction, *American Journal of Medicine*, 112, pp. 336–341.

SNELL, L., TALLETT, S., HAIST, S. *et al.* (2000) A review of the evaluation of clinical teaching: new perspectives and challenges, *Medical Education*, 34, pp. 862–869.

STANLEY, I.M. & WRIGHT, H.J. (1981) A method of evaluating opportunistic teaching in clinical medicine, *Medical Education*, 15, pp. 389–391.

STEINERT, Y. (2000) Faculty development in the new millennium: key challenges and future directions, *Medical Teacher*, 22, pp. 44–50.

STEINERT, Y., NASMITH, L., MCLEOD, P.J. & CONOCHIE, L. (2003) A teaching scholars program to develop leaders in medical education, *Academic Medicine*, 78, pp. 142–149.

VAN DER VLEUTEN, C.P.M., DOLMANS, D.H.J.M. & SCHERPBIER, A.J.J.A. (2000) The need for evidence in education, *Medical Teacher*, 22, pp. 246–256.

WILKERSON, L. & IRBY, D.M. (1998) Strategies for improving teaching practices: a comprehensive approach to faculty development, *Academic Medicine*, 73, pp. 387–396.

WOLF, F.M. (2000) Lessons to be learned from evidence-based medicine: practice and promise of evidence-based medicine and evidence-based education, *Medical Teacher*, 22, pp. 251–259.

WOLF, F.M., SHEA, J.A. & ALBANESE, M.A. (2001) Toward setting a research agenda for systematic reviews of evidence of the effects of medical education, *Teaching and Learning in Medicine*, 13, pp. 53–60.

Medical Teacher, Vol. 28, No. 5, 2006, pp. 404–408

informa
healthcare

TWELVE TIPS

Twelve tips for developing effective mentors

SUBHA RAMANI[1], LARRY GRUPPEN[2] & ELIZABETH KRAJIC KACHUR[3]

[1]Boston University School of Medicine, USA; [2]University of Michigan Medical School, USA; [3]New York, USA

ABSTRACT *Mentoring is often identified as a crucial step in achieving career success. However, not all medical trainees or educators recognize the value of a mentoring relationship. Since medical educators rarely receive training on the mentoring process, they are often ill equipped to face challenges when taking on major mentoring responsibilities. This article is based on half-day workshops presented at the 11th Ottawa International Conference on Medical Education in Barcelona on 5 July 2004 and the annual meeting of the Association of American Medical Colleges in Boston on 10 November 2004 as well as a review of literature. Thirteen medical faculty participated in the former and 30 in the latter. Most participants held leadership positions at their institutions and mentored trainees as well as supervised mentoring programs. The workshops reviewed skills of mentoring and strategies for designing effective mentoring programs. Participants engaged in brainstorming and interactive discussions to: (a) review different types of mentoring programs; (b) discuss measures of success and failure of mentoring relationships and programs; and (c) examine the influence of gender and cultural differences on mentoring. Participants were also asked to develop an implementation plan for a mentoring program for medical students and faculty. They had to identify student and faculty mentoring needs, and describe methods to recruit mentors as well as institutional reward systems to encourage and support mentoring.*

Introduction

Many professionals identify a mentoring relationship as an essential step for achieving success in politics, business and academia (Roche, 1979). Indeed, most successful people in different areas of human endeavor can point to a mentor who was crucial to their career growth and success. The importance of mentoring throughout one's career has been emphasized, especially during professional transitions (Bligh, 1999; Freeman, 2000; Grainger, 2002; Levy *et al.*, 2004). Studies have shown that faculty members who identified a mentor felt more confident, were more likely to have a productive research career and reported greater career satisfaction (Palepu *et al.*, 1998; Ramanan *et al.*, 2002; Levy *et al.*, 2004). Other reported benefits for mentees include: socialization into the profession; help with choice and fulfillment of career path; meaningful involvement in academic activities; and the development of close collaborative relationships (Morzinski *et al.*, 1996; Pololi *et al.*, 2002). Self-reported benefits for mentors include pride in developing the next generation, building a network of professional collaborators within an institution and being able to disseminate their expertise and skills to a group of mentees. From a mentoring program perspective faculty retention has been reported as a positive outcome (Benson *et al.*, 2002). Despite these benefits, many early career clinicians and investigators have difficulty in finding appropriate mentors. Women and clinician-educator faculty in particular are at risk of inadequate mentoring relationships (Chew *et al.*, 2003).

The mentoring relationship usually develops between an older professional, the 'mentor,' and a younger colleague, the 'mentee' (Grainger, 2002). In the *Odyssey*, Mentor was a trusted friend of Odysseus, who entrusted Mentor with the care of his house and the education of his son, Telemachus, when he set out for the Trojan War. From this epic arose the use of the word mentor as a wise and faithful counselor. Today, a mentor is someone who is a counselor and a teacher and instructs, admonishes and assists a junior trainee or colleague in attaining success.

The 12 tips described below are a summary of participant discussions at the Ottawa conference and AAMC annual meeting workshops from a slightly different angle, namely the needs of mentors themselves (Table 1).

Tip 1: Mentors need clear expectations of their roles and enhanced listening and feedback skills

Mentors are not born but developed

Research reports have listed some valuable characteristics of effective mentors (Bhagia & Tinsley, 2000; Grainger, 2002, Hesketh *et al.*, 2003; Jackson *et al.*, 2003; Levy *et al.*, 2004). These include being knowledgeable and respected in their field, being responsive and available to their mentees, interest in the mentoring relationship, being knowledgeable of the mentee's capabilities and potential, motivating mentees to appropriately challenge themselves and acting as advocates for their mentees. Some key skills required when mentoring others include listening and the ability to give positive as well as negative feedback.

Many educators are not born with these skills and would benefit from institutional staff development programs on mentoring skills. Such programs could highlight the key

Correspondence: Subha Ramani, MD MPH DipMedEd, Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, 715 Albany Street, EB33A, Boston, MA 02118, USA. Tel: 617-638-7985; email: sramani@bu.edu

ISSN 0142–159X print/ISSN 1466–187X online/06/050404–5 © 2006 Informa UK Ltd.
DOI: 10.1080/01421590600825326

Table 1. Tips to promote effective mentors: three domains.

Developing mentors	Rewarding mentors	Supporting mentors
Mentor staff development	Academic recognition	A peer-support group
Heighten awareness of gender and culture issues	Protected time	Mentors for mentors
Education on professional boundaries	Financial and non-financial rewards	Referral panel: study skills counsellors, psychologists etc.

responsibilities of a mentor, skills required for an effective mentoring relationship and strategies to recognize problems in a relationship (Benson *et al.*, 2002). These workshops would be most effective if they used a combination of educational strategies that allowed prospective mentors to engage in practical exercises such as watching videotaped scenarios and role-plays (Connor *et al.*, 2000). One such program at the University of Leeds used simulated GPs (general practitioners) with roles based on real mentoring experiences as a learning tool for improving mentoring skills (Sloan & McMillan, 2003). There were opportunities for the GP mentors to practice their skills on three different simulated mentees followed by an in-depth discussion and feedback. This proved to be an invaluable developmental process for the GP mentors. It is to be emphasized, however, that the actual outcomes of such staff development programs should be measured in real mentoring settings. Examples of outcomes might include trainee satisfaction, observation or videotaping of staff during their mentoring sessions with peer feedback or evaluation of staff in an objective structured teaching evaluation (OSTE) format.

Tip 2: Mentors need awareness of culture and gender issues

Mentor and mentee matching by gender and culture should not be mandatory, but available for those who desire it

Although differences in gender and culture have been considered relative barriers to an effective relationship, literature reports have documented that these have not been viewed by most mentees as real barriers (Jackson *et al.*, 2003). In fact, our workshop participants thought that mentors can support mentees of different cultures and gender by having zero tolerance for discrimination. Gender and cultural differences can foster greater mutual growth of the mentor and mentee as they gain knowledge of each other's cultures. It has been recommended that mentors be aware of their own gender and culture biases as this knowledge could possibly help people overcome innate prejudices. It is also thought that faculty development workshops can help all mentors become comfortable and competent in working with students from different backgrounds (Parker, 2002).

Two issues were raised at the workshops in relation to cross-gender mentoring. The first was that of personal boundaries and the second, lack of understanding of the other gender's domestic responsibilities. Despite these concerns, most mentees did not feel the need for having a same gender mentor. The opinions of our participants reflected those reported in the literature. They felt that relationships across cultures and gender would promote more acceptance of differences and lowering of biases. They stated that institutions should not actively try to pair mentors and mentees based on gender and culture and mentors should be equipped with the skills required to understand issues related to their mentees' gender and ethnicity. However, if individual trainees report discrimination or significant barriers to meaningful mentoring based on these characteristics/variables, the institution should find them mentors who can put them more at ease and better fulfill their mentoring needs.

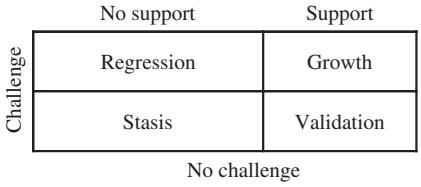


Figure 1. Support vs. challenge.
Source: Figure adapted from Daloz (1986).

Tip 3: Mentors need to support their mentees, but challenge them too

Balance support and challenge

Daloz (1986) states that effective mentor–protégé(e) relationships should balance three elements: support, challenge and a vision of the protégé(e)'s future. If mentors are overly supportive without challenging mentees, the mentees do not grow professionally; on the other hand, challenging without supporting causes mentees to regress in their professional development (Figure 1). Effective mentors balance support with challenge by providing opportunities and setting positive expectations (Bower *et al.*, 1998).

Tip 4: Mentors need a forum to express their uncertainties and problems

Mentors have problems too

It is often assumed that once faculty become mentors, they become all-knowing and do not need any further attention from the program. However, many mentors expressed the need to have a mechanism by which they could discuss problems in their mentoring relationships and get advice. Given that mentors often have more than one mentee and each interpersonal relationship is likely to be different, skills that are effective in one may be ineffective in the other. If they can interact with mentoring colleagues, they might discover solutions to each other's challenges. While discussing

challenges in their mentoring relationships and seeking solutions, it must be remembered that details regarding specific individuals must remain confidential (Freeman, 1997). Institutions can schedule periodic mentor meetings led by senior educators along with external consultants who are knowledgeable about methods for troubleshooting problems in mentoring relationships. Such meetings could provide a forum for mentors to report their successes and failures, and to receive feedback from their peers and the experts. These discussions should include only essential details of the mentoring issues and mentee names and other details must remain confidential.

Tip 5: Mentors need to be aware of professional boundaries

Mentors should stick to mentoring

There are several types of boundaries that need to be considered in a mentoring relationship where the personal contact between mentors and mentees is much closer than in other professional relationships such as a student with a teacher, advisor or role model. Moreover, personal issues and problems may be discussed by a mentee, which could lead to one or both of them harboring inappropriately intimate emotions towards the other (Palepu *et al.*, 1996; Jackson *et al.*, 2003). Mentees could become excessively dependent on their mentors for personal and professional support, which may become a drain on the mentor's energy. As one of our participants stated: ''I had a mentee who expected me to mother him throughout his training period and that was emotionally exhausting.'' Mentor training should include knowledge of professional boundaries, and recognition of psychosocial problems that need referral to professionals such as psychologists or counselors.

Tip 6: Mentors also need mentoring

Mentors for mentors

There were some senior faculty among the workshop participants with vast experience in teaching and mentoring different levels of trainees. All of them felt abandoned by the system once they assumed leadership positions within their institutions. Educational institutions often do not provide mentors for senior teachers. Our participants felt that even the most senior educators need to be mentored as they may wish to change their career focus or professional path while they already hold high positions within their organizations.

Tip 7: Mentors need recognition

Raise the value of mentoring

At most educational institutions around the world, mentors usually perform their mentoring duties not because they are reimbursed for it but because they consider it a rewarding aspect of their profession. However, they usually carry out their mentoring privately and neither their peers nor their superiors are even aware of the mentoring load they carry, let alone laud their efforts. To convince the entire institution that mentoring is one of the most important duties at medical schools, institutional leaders should publicly recognize their group of mentors as an 'elite' group of faculty who are highly valued and appreciated for their work (Palepu *et al.*, 1998). They can be given special honors within the institution and their names announced at major university events and openly appreciated.

Tip 8: Mentors need to be rewarded

Mentors can be rewarded in different ways

Educational institutions can reward their core group of mentors in several innovative ways. Mentor retreats or dinners can be held periodically. At retreats or even just occasional dinners mentors can interact with their colleagues, share their experiences and techniques, both effective and ineffective. Another method to reward special mentors would be to give them extra conference funding. These rewards can easily be given by institutions even in times of economic shortfalls. Additionally mentoring can become a criterion for promotion (Benson *et al.*, 2002).

Tip 9: Mentoring needs protected time

Mentoring cannot be done 'on the fly'

Institutions should recognize that mentoring is one of the key activities of faculty at any educational institution. Faculty who mentor several trainees should be allocated some degree of protected time to perform this important duty effectively. Just adding this important duty to the existing workload is a recipe for poor mentoring relationships.

Tip 10: Mentors need support

Mentors should not be expected to tackle personal or psychological problems

Some mentees' problems may overstep the boundaries of the usual mentor–mentee relationships and discussions. Mentees may be clinically depressed, have personality problems, have substance abuse problems or just academic problems. Mentors should be able to recognize when they feel unable to resolve such problems and should be supported by a network of specialists such as study counselors and psychologists to whom they can refer their mentee. The mentors should not be forced to take on roles in which they do not have expert skills. Once again, the matter of professional boundaries arises.

Tip 11: Encourage peer mentoring

A pyramidal model of mentoring

Medical educators who have studied peer (or near-peer) mentoring suggest that it is a feasible and perhaps more desirable alternative to traditional dyadic mentoring approaches (Woessner *et al.*, 1998; Pololi *et al.*, 2002). Participants identified their peers as 'collaborators' or 'colleagues' (implying a non-hierarchical relationship), while seeking shared insights, experiences, ideas, guidance, problem-solving and support from them. Their reference to peer collaborators reflects a non-hierarchical mentoring process, in contrast to senior–junior mentoring relationships

where characteristics such as power, dominance, dependency and transference have been noted (Pololi *et al.*, 2002).

Pressures on faculty time could be alleviated to a certain extent by creating a pyramidal system of mentoring. Such a model would entail a group of mentees at the bottom of the pyramid who can seek advice from a small group of peers a little higher in the pyramid with the more experienced, senior mentors overseeing and guiding all of them at the top of the pyramid. This pyramidal system would minimize the threat of the power relationship, yet offer the benefit of the valuable experience that senior faculty at the top of the pyramid possess. The advantages of peer mentoring include easier availability, greater understanding of day-to-day problems related to workload stress or conflicts with teachers, and early recognition of serious abuse or emotional problems. Mentees may be more open to sharing their problems with peers than with faculty. The same advantages would apply for faculty mentoring programs as well. It has been shown that faculty may be more willing to share their difficult problems with peer mentors than with senior mentors (Pololi *et al.*, 2002).

Tip 12: Continuously evaluate the effectiveness of the mentoring programs

Mentoring is a work in progress

For mentoring programs to succeed, institutions need to have the mentees and mentors evaluate the program periodically, report the current problems and suggest new approaches to mentoring or changes to the existing program. Evaluation of mentoring should look at process, content and outcomes as noted below (Grainger, 2002):

- Process
  - Clear objectives
  - Regular, purposeful meetings
- Content
  - Feedback
  - Mentee could raise issues and challenge mentor
- Outcome
  - Progress and career development
  - Networking

All mentees and mentors at a given institution should be asked to evaluate their mentoring relationships at least 3–4 times a year. The following items are examples of areas in a mentoring relationship that could be evaluated:

- congruence on professional goals;
- availability of mentor(s);
- mentor giving mentee responsibilities and opportunities;
- mentor involving mentee on committees and other professional activities;
- mentor facilitating networking with internal and external faculty;
- mentor helping mentee integrate work and personal life;
- mentor showing respect for the mentee as a person;
- personal benefits from mentoring.

Institutional leaders could also consult outside experts, particularly at national and international educational meetings, where they could discuss the mentoring challenges at their home sites and take back ideas to overcome those challenges. They could have a committee within their institution that would be responsible for receiving feedback from its mentors and mentees to modify their mentoring system as needed.

Conclusions

Mentoring is a vital cog in the machinery of medical education. Faculty who serve as mentors frequently are not trained in effective mentoring skills or designing mentoring programs. They are most often very busy with their core clinical, research, administrative or educational responsibilities and are expected to squeeze mentoring onto an already full plate. Once they take on mentoring duties, they usually are left to their own devices and have few avenues to discuss problems and challenges in their mentoring programs or relationships. It is evident from the foregoing discussion that faculty need training to be mentors and to benefit from peer mentoring themselves, and must be rewarded for a job well done. Institutions should change their culture to overtly value and reward mentoring so that mentoring does not remain an invisible and only implicitly valuable aspect of their educational programs.

Notes on contributors

Subha Ramani, MD MPH MMedEd, is a general internist and medical educator at the Boston University School of Medicine and its affiliated Boston Medical Center. She is currently an Assistant Professor of Medicine and Director of Faculty Development in Clinical Teaching for the Department of Medicine. She also directs a clinical skills curriculum for Internal Medicine residents. Her chief areas of interest in medical education are staff development in teaching and clinical skills education.

Larry Gruppen, PhD, is Professor and Chair of the Department of Medical Education at the University of Michigan Medical School. He is a cognitive psychologist whose areas of research interest include the fundamental aspects of clinical reasoning, the evolution and assessment of medical expertise, performance assessment, and educational leadership development.

Elizabeth Krajic Kachur, PhD, is a medical education consultant with 22 years of experience throughout the continuum of medical education. She is located in New York City but also works nationally and internationally. Her experience with mentoring schemes ranges from program and mentor development to project evaluation.

References

Benson, C.A., Morahan, P.S., Sachdeva, A.K. & Richman, R.C. (2002) Effective faculty preceptoring and mentoring during reorganization of an academic medical center, *Medical Teacher*, 24, pp. 550–557.

Bhagia, J. & Tinsley, J.A. (2000) The mentoring partnership, *Mayo Clinic Proceedings*, 75(5), pp. 535–537.

Bower, D.J., Diehr, S., Morzinski, J.A. & Simpson, D.E. (1998) Support–challenge–vision: a model for faculty mentoring, *Medical Teacher*, 20, pp. 595–597.

Bligh, J. (1999) Mentoring: an invisible support network, *Academic Medicine*, 77, pp. 377–384.

Chew, L.D., Watanabe, J.M., Buchwald, D. & Lessler, D.S. (2003) Junior faculty's perspectives on mentoring, *Academic Medicine*, 78(6), p. 652.

Connor, M.P., Bynoe, A.G., Redfern, N., Pokora, J. & Clarke, J. (2000) Developing senior doctors as mentors: a form of continuing professional development: report of an initiative to develop a network of senior doctors as mentors: 1994–99, *Medical Education*, 34, pp. 747–753.

Daloz, L.A. (1986) *Effective Teaching and Mentorship: Realizing the Transformational Power of Adult Learning Experiences*, pp. 209–235 (San Francisco, Jossey-Bass).

Freeman, R. (1997) Information shared in mentoring must remain confidential, *British Medical Journal*, 314(7074), p. 149.

Freeman, R. (2000) Faculty mentoring programmes, *Medical Education*, 34, pp. 507–508.

Grainger, C. (2002) Mentoring—supporting doctors at work and play, *BMJ Career Focus*, 324, p. S203.

Hesketh, E.A. & Laidlaw, J.M. (2003) Developing the teaching instinct 5: Mentoring, *Medical Teacher*, 25, pp. 9–12.

Jackson, V.A., Palepu, A., Szalacha, L., Caswell, C., Carr, P.L. & Inui, T. (2003) 'Having the right chemistry': a qualitative study of mentoring in academic medicine, *Academic Medicine*, 78(3), pp. 328–334.

Levy, B.D., Katz, J.T., Wolf, M.A., Sillman, J.S., Handin, R.I. & Dzau, V. J. (2004) An initiative in mentoring to promote residents' and faculty members' careers, *Academic Medicine*, 79(9), pp. 845–850.

Morzinski, J.A, Diehr, S., Bower, D.J. & Simpson, D.E. (1996) A descriptive, cross-sectional study of formal mentoring for faculty, *Family Medicine*, 28, pp. 434–438.

Palepu, A., Friedman, R.H., Barnett, R.C., Carr, P.L., Ash, A.S., Szalacha, L. & Moscowitz, M.A. (1998) Junior faculty members' mentoring relationships and their professional development in US medical schools, *Academic Medicine*, 73, pp. 318–323.

Parker, D.L. (2002) A workshop on mentoring across gender and culture lines, *Academic Medicine*, 77(5), p. 461.

Pololi, L.H., Knight, S.M., Dennis, K. & Frankel, R.M. (2002) Helping medical school faculty realize their dreams: an innovative, collaborative mentoring program, *Academic Medicine*, 77, pp. 377–384.

Ramanan, R.A., Phillips, R.S., Davis, R.B., Silen, W. & Reede, J.Y. (2002) Mentoring in medicine: keys to satisfaction, *American Journal of Medicine*, 112(4), pp. 336–341.

Roche, G.R. (1979) Much ado about mentors, *Harvard Business Review*, 1, pp. 14–31.

Sloan, R.E.G. & McMillan, J. (2003) Developing mentoring skills for general practitioners using a simulated doctor, *Medical Education*, 37, pp. 1044–1045.

Woessner, R., Honold, M., Stehle, I., Stehr, S. & Steudel, W.I. (1998) Faculty mentoring programme – ways of reducing anonymity, *Medical Education*, 32, pp. 441–443.

REGULAR ARTICLES

Teaching Internal Medicine Resident Physicians About Alcoholics Anonymous: A Pilot Study of an Educational Intervention

Adam J. Rose, MD
Melissa R. Stein, MD
Julia H. Arnsten, MD, MPH
Richard Saitz, MD, MPH

ABSTRACT. Greater physician confidence in treating alcoholism is associated with a higher frequency of referring alcoholic patients for treatment, but many physicians have limited experience with Alcoholics Anonymous. We implemented a brief, didactic and experiential educational intervention about AA and evaluated its effect on knowledge and attitudes, using a before-after repeated measures study design. Thirty-six first-year internal medicine resident physicians received an educational intervention, which consisted of a 45-minute lecture about AA, a visit to an AA meeting, and a 30-minute debriefing session the next day. Residents' knowledge and attitudes were assessed by a brief written anonymous survey before and after the educational intervention. Residents reported increases in self-perceived knowledge about AA and had more favorable attitudes towards AA after the intervention. Our pilot study shows that a brief, didactic and experiential course can improve physician knowledge and attitudes about AA, and holds promise for improving physician interface with this commonly used intervention. doi:10.1300/J465v27n03_02

Adam J. Rose is affiliated with the Section of General Internal Medicine & Boston University School of Medicine, and Boston Veterans Administration Medical Center, Boston, MA.

Melissa R. Stein and Julia H. Arnsten are affiliated with the Division of General Internal Medicine, and the Division of Substance Abuse, Albert Einstein College of Medicine, Bronx, NY.

Richard Saitz is affiliated with the Section of General Internal Medicine & Boston University School of Medicine, Clinical Addiction Research and Education Unit, Boston Medical Center, and Youth Alcohol Prevention Center and also the Department of Epidemiology, Boston University School of Public Health, Boston, MA.

Address correspondence to: Adam J. Rose, MD, Section of General Internal Medicine, Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118-2644 (E-mail: adamrose@bu.edu).

The authors thank the subjects for their enthusiastic participation in this study.

This research was supported in part by grants R25 DA13582 (PI: J. Samet) and R25 DA15302 (PI: J. Arnsten) from the National Institute on Drug Abuse (NIDA), and by the Department of Veterans Affairs Office of Academic Affairs. No relevant conflicts of interest exist in relation to this work.

KEYWORDS. Alcoholics Anonymous, medical education, housestaff, alcohol dependence, internal medicine

INTRODUCTION

The widespread nature of alcoholism and its enormous impact on public health are well known (1,2). The challenges inherent in diagnosing and treating alcoholism are made even more daunting by the fact that health insurance coverage and access to high quality care for alcohol dependence are often inadequate (3,4). Although there are no controlled trials comparing Alcoholics Anonymous with no treatment, research to date suggests that AA is efficacious for people with alcoholism either as a sole intervention or a component of a treatment plan (5,6). In one study by Morgenstern et al. (5), affiliation with AA was associated with more abstinence and fewer days of alcohol consumption. AA is free, and local meetings are available worldwide, contributing to its widespread utilization by patients. According to the 2004 National Survey on Drug Use and Health, 3.8 million Americans received treatment for a substance abuse disorder in 2004, and more than half (2.1 million) were treated in the context of a self-help group such as AA, making self-help groups the most commonly used intervention for substance abuse in the United States (3).

The education of future generalists in the diagnosis and management of substance abuse disorders is especially important, because it is uncommon for patients to receive treatment for a substance abuse disorder from a specialist in addiction medicine. In 2004, there were 23.5 million Americans who required treatment for a substance abuse problem, but only 2.3 million of these people received treatment at a facility specializing in the treatment of substance abuse, leaving the great majority to be cared for by generalist physicians (3,7). However, most generalist physicians feel poorly prepared to identify, assess, and address alcoholism in their patients (8-11). Although most physicians ask about alcohol use, they often do not intervene effectively in patients with identified disorders, and they also report dissatisfaction when caring for patients with alcoholism (12). Indeed, a recent study by the RAND Corporation confirmed that the quality of care for identified alcohol dependence is the worst among all chronic conditions. In that study, only 4.6% of patients diagnosed with alcohol dependence were referred for any sort of treatment, including inpatient rehabilitation, outpatient rehabilitation, self-help groups, counseling, or aversion therapy (4).

Clinicians who have higher levels of perceived skill and responsibility for diagnosing and treating alcoholism perform better in these areas. Research by Geller et al. suggests that trainees who feel highly confident in their abilities or highly responsible screen and refer patients for alcohol-related disorders almost twice as often as others (13). One major challenge of medical education, therefore, is to help physicians in training both to develop skills to diagnose and treat alcoholism and to feel comfortable and motivated enough to use these skills effectively. But how is this to be accomplished? Over the past thirty years, the preponderance of evidence has shown that experientially-based educational interventions are the most effective in changing the attitudes of medical trainees towards alcoholics and alcoholism, and in improving their confidence and their practice patterns (14-20). In fact, greater experience with alcoholism is the single greatest predictor of excellent clinical practice in this area, arguing for education through direct experience (21).

Previous studies have evaluated educational interventions to improve physician practice with regard to the diagnosis and treatment of alcoholism. D'Onofrio et al., using a randomized controlled design, showed that a brief curriculum improved self-reported knowledge and practice among emergency medicine residents with regard to screening and intervening for alcohol problems (14). Kahan et al., using a randomized controlled design, showed that a skills-based workshop improved the performance of medical students with a group of standardized patients, although a follow-up survey several months later showed attenuation of the difference (22). Ockene et al., using a pre-post design, showed that attendings, residents, and

nurse practitioners who received a curriculum in patient-centered counseling showed increased counseling skills, preparedness to intervene, perceived importance of intervention, and measured knowledge (23).

However, these prior reports of educational interventions regarding management of alcohol problems have not emphasized knowledge and attitudes regarding AA in particular. This is important because there is a gap between patients and practitioners regarding the emphasis placed on spirituality in the treatment of substance abuse disorders. In particular, Goldfarb et al. have shown that a group of medical students were less likely to emphasize spiritually based approaches to treating substance abuse than recipients of such treatment (24). Indeed, the work of Fazzio et al. suggests that experienced substance abuse faculty and pre-clinical medical students naturally emphasize the importance of spiritual approaches to treating substance abuse, but this emphasis shifts toward the biomedical during third-year clerkships, and may need to be relearned during years of treating substance-abusing patients (25). Chibnall et al. have shown that physicians require specific training to become comfortable with spiritually based treatment modalities (26). There is a need for educational interventions focused on introducing generalist physicians to AA, interventions based upon experiential learning and introducing AA specifically as a spiritually-based treatment program for alcoholism.

To address this need, we designed and piloted a brief, didactic and experiential educational intervention focused on AA for first-year internal medicine resident physicians-in-training (interns). The objective of the educational intervention was to introduce the interns to AA as a resource. The premise was that greater familiarity with and understanding of AA would lead to greater incorporation of AA into management plans for alcoholic patients, by increasing both referrals to AA and support of patients already participating in AA (27). Although we focused our educational intervention on experiential learning through attendance at an AA meeting, the didactic portion of the course was also important for introducing key concepts such as the role of spirituality in AA, which is one feature of AA that is unique among treatment options for alcoholism. We evaluated the effect of the educational intervention using a before-after study design to measure changes in knowledge and attitudes about AA. We hypothesized that a brief, feasible educational intervention would increase knowledge and improve attitudes of resident physicians.

METHODS

Participants

First year internal medicine resident physicians were the participants in the educational intervention. The participants were drawn from the three-year internal medicine residency program at Montefiore Medical Center of Bronx, NY, an affiliate of the Albert Einstein College of Medicine. One hundred percent of the thirty-six interns in the program participated in the study, which was approved by the institutional review board of Montefiore Medical Center.

The Educational Intervention

The course was incorporated into a month-long ambulatory care clinical rotation, as part of a larger substance abuse curriculum. Its contents consisted of 1. a lecture and discussion, 2. attendance at an AA meeting, and 3. a debriefing discussion. One of the authors (AJR) delivered a forty-five minute lecture and discussion to introduce AA. The content of the talk, which was based on previous work about treating alcoholism in primary care, included the history and goals of AA, the structure and function of AA, evidence for the effectiveness of AA and other treatment modalities for alcoholism, information about how to refer a patient to AA, and how to support a patient who is in AA. Possible areas of discomfort with AA, including a perceived emphasis on religion, were included in the discussion (25,26). The twelve steps and the twelve traditions, the ideological bedrock of AA, were circulated and discussed, and the physicians-in-training were oriented to what occurs at an AA meeting.

That evening, the trainees attended a 50-minute AA Beginners Meeting, an open AA

meeting specifically geared toward being accessible to those who have never before attended an AA meeting. This weekly Beginners Meeting was followed by two closed meetings, which the interns did not attend. There were more than 100 people in attendance at these meetings every week, most of whom were members of AA who had come in order to attend the entire evening.

The following morning, a coauthor (AJR) moderated a thirty-minute discussion with the trainees regarding their experiences at the meeting, reactions to what they had seen, and lessons learned. The trainees' experiences and reactions dictated the course of the discussion, but the reactions of the trainees to the perceived spiritual content of AA was explored with each group in addition to any other topics discussed.

Evaluation

The effects of the educational intervention on learner knowledge and attitudes were assessed by written self-report before the lecture and discussion and after the debriefing discussion. The questionnaire assessed demographics, prior exposure to alcoholism and AA, self-assessed knowledge of AA, perceived effectiveness of AA, and comfort with AA. All responses were on a nine-point Likert Scale (1 = strongly disagree, 3 = disagree, 5 = neutral, 7 = agree, 9 = strongly agree). A thorough search of the relevant literature did not reveal any pre-existing instruments that dealt specifically with our study question; therefore, we based our items on previously published substance abuse attitude questions (12,23,28). The questionnaires were anonymous, but were handed out in pairs marked with matching numbers, with the second questionnaire to be retained by the trainee and then filled out at the conclusion of the course.

Statistical Analysis

Paired t-tests were used to compare the mean scores on each survey item before and after the intervention. Alpha was set to the 0.05 level of significance; no adjustments were made for multiple testing. All analyses were performed with SAS version 9.1(29) using PROC FREQ and PROC UNIVARIATE.

RESULTS

Baseline Characteristics of the Sample

Baseline characteristics of the thirty-six participants are summarized in Table 1. A majority of the group consisted of women (22). Caucasians (13) and Asian/Pacific Islanders (11) constituted the bulk of the group. Most of the participants (28) had never previously attended an AA meeting, but the majority of them (20) had some personal contact with a person with a substance abuse problem. The mean age of the participants was 27.7 years old.

Responses to the Questionnaire

The results of the questionnaire are summarized in Table 2. The questions are divided thematically between items relating to knowledge and items relating to attitudes and beliefs. Participants reported increased knowledge of what

TABLE 1. Characteristics of the study subjects (n = 36)

Variable	Number
Gender	
Male	14
Female	22
Race	
White	13
Asian/Pacific Islander	11
South Asian	7
African-American	3
Mixed Race	2
Have you ever been to an AA meeting in the past?	
Yes	8
No	28
Do you have either a family member or a friend with a substance abuse problem?	
Yes	20
No	16

TABLE 2. Results of the questionnaire distributed before and after the educational intervention (n = 36).

Category	Question	Mean Score Before Intervention (SD)[a]	Mean Score After Intervention (SD)[a]	P-Value
Perceived Knowledge	I know what occurs at an AA meeting.	5.2 (2.3)	8.1 (0.9)	< 0.001
Perceived Knowledge	I understand the role of a sponsor in the AA program.	6.3 (2.2)	8.0 (1.0)	< 0.001
Perceived Effectiveness of AA	I believe that AA is an effective treatment option to help alcoholics to remain abstinent.	7.6 (1.2)	8.1 (0.9)	0.008
Comfort with AA	I would be comfortable referring a patient to AA.	7.4 (2.2)	8.4 (0.9)	0.003
Comfort with AA	I would be comfortable asking my alcoholic patient how well AA is working for him.	8.3 (1.2)	8.5 (0.9)	0.4
Comfort with AA	I would be comfortable asking my alcoholic patient about his rapport with his sponsor.	7.6 (1.5)	8.4 (1.0)	0.002
Comfort with AA	I am comfortable with the importance of spirituality in AA.	6.3 (2.4)	7.4 (1.7)	0.008

[a]Nine-point scale where 1 = strongly disagree and 9 = strongly agree

occurs at an AA meeting and increased understanding of the role of an AA sponsor after the intervention. There were also changes in attitudes and beliefs. After the intervention, the participants reported greater belief in the effectiveness of AA, increased comfort in referring patients to AA and in discussing the patients' sponsors, and greater comfort with the role of spirituality in the AA program.

DISCUSSION

Our pilot study demonstrated that a brief educational intervention consisting of a didactic session with discussion, coupled with AA meeting attendance and debriefing, was associated with improvements in knowledge and attitudes about AA. This educational intervention was modest, requiring less than two hours of curricular time, plus an evening spent at the AA meeting. The investigator who implemented the course was a chief resident who had received training in a similar setting, but was not a substance abuse expert. These features, plus the widespread availability of AA meetings, make it feasible to replicate this educational intervention in diverse settings.

Our educational intervention was associated with improvements in both knowledge and attitudes regarding AA. However, it should be noted that only one of the seven questionnaire items failed to show a before-after change: "I would be comfortable asking my alcoholic patient how well AA is working for him." This item had the highest mean score for any item on the pre-test (8.3). As such, substantial improvement was not possible (a ceiling effect). In contrast, respondents had a lower degree of comfort with the task of asking an alcoholic patient about his rapport with his sponsor prior to the intervention (mean score 7.6), allowing room for improvement.

Our study has limitations. We did not assess an actual change in clinical practice on the part of the trainees, but prior studies have linked changes in knowledge and attitudes to changes in practice in this clinical area (13,14). However, as with any study that measures a surrogate endpoint such as a change in physician knowledge and attitudes, our study should not be regarded as definitive with regard to the ultimate goal–practice change–but should point the way for further studies designed to demonstrate changes in physician practice in treating patients with alcoholism or improved clinical outcomes for alcoholic patients. We also did not study actual knowledge through the use of a quiz, but rather studied self-reported knowledge. In addition, despite the anonymous na-

ture of the questionnaire, social desirability bias may have influenced the interns to make unjustified claims about their knowledge and attitudes with regard to AA. However, such a bias would likely have been present both before and after the intervention, thus biasing the results toward the null. Finally, we did not study a comprehensive alcohol curriculum. Rather, we focused on one critical element that holds promise for increasing an intervention that has wide applicability. Other studies are needed to improve education about screening, brief intervention, and additional management of alcoholism in primary care settings.

There are many competing priorities in medical education, and only a limited amount of time to address them. Isaacson et al. showed that in the year 2000, only 51% of residency programs in internal medicine had a formal curriculum in substance abuse at all, and the median amount of time allotted by those programs that did was 5 hours, for all kinds of substance abuse combined (30). However, our pilot study demonstrated that with a minimum of time and effort it is feasible to introduce an experientially based educational intervention regarding the treatment of alcoholism, with an emphasis on AA as a resource. This educational intervention can be led by a generalist or a chief resident, rather than by a substance abuse specialist. In our setting, this intervention increased self-reported knowledge of and comfort with AA, and it has the potential to improve physicians' abilities to interface with AA as a treatment resource.

REFERENCES

1. Moore RD, Bone LR, Geller G, Mamon JA, Stokes EJ, Levine DM: Prevalence, detection, and treatment of alcoholism in hospitalized patients. JAMA 1989; 261(3):403-7.

2. Saitz R: Clinical practice. Unhealthy alcohol use. N Engl J Med 2005;352(6):596-607.

3. Substance Abuse and Mental Health Services Administration: Results from the 2004 National Survey on Drug Use and Health: National Findings. (Office of Applied Studies, NSDUH Series H-28, DHHS Publication No. SMA 05-4062.) Rockville, MD.

4. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA: The quality of health care delivered to adults in the United States. N Engl J Med 2003;348(26):2635-45.

5. Morgenstern J, Labouvie E, McCrady BS, Kahler CW, Frey RM: Affiliation with Alcoholics Anonymous after treatment: a study of its therapeutic effects and mechanisms of action. J Consult Clin Psychol 1997; 65(5):768-77.

6. Walsh DC, Hingson RW, Merrigan DM, Levenson SM, Cupples LA, Heeren T, Coffman GA, Becker CA, Barker TA, Hamilton SK, et al.: A randomized trial of treatment options for alcohol-abusing workers. N Engl J Med 1991;325(11):775-82.

7. Lewis DC: The role of internal medicine in addiction medicine. J Addict Dis 1996;15(1):1-17.

8. Friedmann PD, McCullough D, Chin MH, Saitz R: Screening and intervention for alcohol problems. A national survey of primary care physicians and psychiatrists. J Gen Intern Med 2000;15(2):84-91.

9. Kaner EF, Heather N, McAvoy BR, Lock CA, Gilvarry E: Intervention for excessive alcohol consumption in primary health care: attitudes and practices of English general practitioners. Alcohol Alcohol 1999; 34(4):559-66.

10. Roche AM, Parle MD, Stubbs JM, Hall W, Saunders JB: Management and treatment efficacy of drug and alcohol problems: what do doctors believe? Addiction 1995;90(10):1357-66.

11. Rohman ME, Cleary PD, Warburg M, Delbanco TL, Aronson MD: The response of primary care physicians to problem drinkers. Am J Drug Alcohol Abuse 1987;13(1-2):199-209.

12. Saitz R, Friedmann PD, Sullivan LM, Winter MR, Lloyd-Travaglini C, Moskowitz MA, Samet JH: Professional satisfaction experienced when caring for substance-abusing patients: faculty and resident physician perspectives. J Gen Intern Med 2002;17(5):373-6.

13. Geller G, Levine DM, Mamon JA, Moore RD, Bone LR, Stokes EJ: Knowledge, attitudes, and reported practices of medical students and house staff regarding the diagnosis and treatment of alcoholism. JAMA 1989; 261(21):3115-20.

14. D'Onofrio G, Nadel ES, Degutis LC, Sullivan LM, Casper K, Bernstein E, Samet JH: Improving emergency medicine residents' approach to patients with alcohol problems: a controlled educational trial. Ann Emerg Med 2002;40(1):50-62.

15. Hanlon MJ: A review of the recent literature relating to the training of medical students in alcoholism. J Med Educ 1985;60(8):618-26.

16. Karnitschnig A: Alcoholics anonymous in medical school education. Fam Med 1989;21(2):146-7.

17. Kinney J, Bergen BJ, Price TR: A perspective on medical students' perceptions of alcoholics and alcoholism. J Stud Alcohol 1982;43(5):488-96.

18. Walsh RA: Medical education about alcohol: review of its role and effectiveness. Alcohol Alcohol 1995; 30(6):689-702.

19. Saitz R, Sullivan LM, Samet JH: Training community-based clinicians in screening and brief interven-

tion for substance abuse problems: Translating evidence into practice. Subst Abus 2000;21(1):21-31.

20. Parish SJ, Ramaswamy M, Stein MR, Kachur EK, Arnsten JH: Teaching about substance abuse with objective structured clinical exams. J Gen Intern Med 2006; 21: 453-59.

21. Warburg MM, Cleary PD, Rohman M, Barnes HN, Aronson M, Delbanco TL: Residents' attitudes, knowledge, and behavior regarding diagnosis and treatment of alcoholism. J Med Educ 1987;62(6):497-503.

22. Kahan M, Wilson L, Midmer D, Borsoi D, Martin D: Randomized controlled trial on the effects of a skills-based workshop on medical students' management of problem drinking and alcohol dependence. Subst Abus 2003;24(1):5-16.

23. Ockene JK, Wheeler EV, Adams A, Hurley TG, Hebert J: Provider training for patient-centered alcohol counseling in a primary care setting. Arch Intern Med 1997;157(20):2334-41.

24. Goldfarb LM, Galanter M, McDowell D, Lifshutz H, Dermatis H: Medical student and patient attitudes toward religion and spirituality in the recovery process. Am J Drug Alcohol Abuse 1996;22(4):549-61.

25. Fazzio L, Galanter M, Dermatis H, Levounis P: Evaluation of medical student attitudes toward alcoholics anonymous. Subst Abus 2003;24(3):175-85.

26. Chibnall JT, Duckro PN: Does exposure to issues of spirituality predict medical students' attitudes toward spirituality in medicine? Acad Med 2000;75(6): 661.

27. Friedmann PD, Saitz R, Samet JH: Management of adults recovering from alcohol or other drug problems: relapse prevention in primary care. JAMA 1998; 279(15):1227-31.

28. Chappel JN, Veach TL, Krug RS: The Substance Abuse Attitude Survey: An instrument for measuring attitudes. J Stud Alcohol 1985;46(1):48-52.

29. SAS Version 9.1. SAS Institute, Cary, NC.

30. Isaacson JH, Fleming M, Kraus M, Kahn R, Mundt M: A national survey of training in substance use disorders in residency programs. J Stud Alcohol 2000; 61(6):912-5.

doi:10.1300/J465v27n03_02

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.

Challenges Applying Alcohol Brief Intervention in Diverse Practice Settings: Populations, Outcomes, and Costs

Richard Saitz, Dace Svikis, Gail D'Onofrio, Kevin L. Kraemer, and Harold Perl

This article summarizes the proceedings of a symposium at the 2005 Research Society on Alcoholism, Santa Barbara, California. The purpose of the symposium was to address challenges that arise in translating evidence for efficacy of alcohol brief intervention (BI) into diverse clinical settings and populations by reviewing the literature and describing 4 research studies. Dr. Saitz reviewed the limitations in evidence for efficacy of BIs and then described results of a randomized clinical trial of brief motivational intervention for medical inpatients drinking risky amounts. Dr. Svikis presented alternative methods for identifying pregnant women in prenatal care at risk for alcohol and drug problems (including nicotine and caffeine) and BIs to reduce or eliminate use. Dr. D'Onofrio discussed results of a randomized trial of the brief negotiated interview in emergency department patients. Dr. Kraemer presented results of a decision analytic and computer-simulation model regarding the cost-effectiveness of alcohol screening and intervention in primary care settings. Finally, Dr. Perl discussed the salient issues and suggested future directions for work in the area of alcohol BI.

Key Words: Brief Intervention, Alcohol, Efficacy, Effectiveness, Primary Care, Emergency Department, Prenatal Care, Cost-Effectiveness.

BASED ON EVIDENCE of efficacy, screening, and brief intervention (BI) for unhealthy alcohol use are recommended in national practice guidelines (NIAAA, 2005; Whitlock et al., 2004), but some have begun to question the effectiveness of BI and have noted challenges in disseminating the practice (Beich et al., 2002, 2003). Although clinical trials have found BI to have efficacy, questions remain regarding the magnitude of that efficacy, effects on clinically important outcomes, efficacy in diverse populations and settings, and effectiveness in real-world practice.

This article summarizes the proceedings of a symposium at the 2005 Research Society on Alcoholism, Santa Barbara, California. The purpose of the symposium was to address challenges that arise in translating evidence for efficacy of BI into diverse clinical settings and populations by reviewing the literature and describing 4 research studies. Dr. Saitz reviewed the limitations in evidence for efficacy of BIs and then described results of a randomized clinical trial of brief motivational intervention for medical inpatients drinking risky amounts. Dr. Svikis presented alternative methods for identifying pregnant women in prenatal care at risk for alcohol and drug problems (including nicotine and caffeine) and BIs to reduce or eliminate use. Dr. D'Onofrio discussed results of a randomized trial of the brief negotiated interview in emergency department (ED) patients. Dr. Kraemer presented results of a decision analytic and computer simulation model regarding the cost-effectiveness of alcohol screening and intervention in primary care settings. Finally, Dr. Perl discussed the salient issues and suggested future directions for work in the area of alcohol BI.

LIMITATIONS IN THE EVIDENCE FOR EFFICACY OF BRIEF INTERVENTION

Richard Saitz

There have been at least 11 systematic reviews published addressing the efficacy of BI for unhealthy alcohol use. The US Preventive Services Task Force (USPSTF) reviewed randomized or controlled trials that enrolled

From the Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, and Department of Epidemiology and the Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts (RS); the Department of Psychology, Virginia Commonwealth University, Richmond, Virginia (DS); the Department of Surgery, Yale University Schools of Medicine and Public Health (GD); the Center for Research on Health Care, Division of General, New Haven, Connecticut Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania (KLK); and the Division of Treatment and Recovery Research, National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland (HP).

Received for publication October 3, 2005; accepted November 17, 2005.

RS was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (RO1 AA 12617) and a General Clinical Research Center grant from the National Center for Research Resources (M01 RR00533). DS was supported by NIAAA (R01 AA 11802) and the National Institute of Drug Abuse (R01 DA 11476). GD was supported by NIAAA (RO1 AA 12417).

Reprint requests: Richard Saitz, Section of General Internal Medicine, Boston Medical Center/Boston University Schools of Medicine and Public Health, 91 East Concord Street #200, Boston, MA 02118-2644; E-mail: rsaitz@bu.edu

DOI: 10.1111/j.1530-0277.2006.00038.x

risky or harmful drinkers in primary care settings, not primarily focused on alcohol dependence or patients with psychiatric comorbidity, of at least fair methodological quality, in the English language literature (Whitlock et al., 2004). They identified 12 studies in adults. Results of 7 studies of single brief or very brief (i.e., <5–15 minutes) interventions were mixed. Three studies had positive findings (including one of a very BI). In 2 studies, there were significantly lower proportions of subjects with heavy episodic use and use of risky amounts after BI compared with control subjects, though no differences in drinks per week. In the third study, BI was associated with decreased consumption summarized as grams of alcohol per day. Another large study ($N = 516$) reported a small decrease in risky use (from 27% in the control to 20% in the BI group) of borderline statistical significance between the groups ($p = 0.07$). The remaining 3 studies of single BIs were negative although in 1 case, drinking decreased in both intervention and control groups and twice as much in the intervention group (but differences between groups were not statistically significant).

The USPSTF identified 7 studies of brief multicontact (more than one) interventions. Six of the 7 were of "good" methodological quality and had positive results for the intervention groups on consumption per day or week or exceeding risky amounts. In 6 of the studies that examined alcohol-related problems or consequences, there were no significant differences. Three studies had more dropouts in intervention than in control groups, raising the possibility of subject dissatisfaction or discomfort from the intervention. Of note, in most large, positive studies the BI was done by the patient's health care professional, and outcomes were measured at 1 year.

Other systematic reviews of controlled studies, some detailing and counting each study result and others meta-analyzing the findings yield similar results (Ballesteros et al., 2004; Beich et al., 2003; Bertholet et al., 2005; Bien et al., 1993; Dinh-Zarr et al., 1999; D'Onofrio and Degutis, 2002; Kahan et al., 1995; Moyer et al., 2002; Poikolainen, 1999; Wilk et al., 1997). All reviews that examined consumption outcomes found significant decreases associated with BI in diverse settings. One review found decreases in injury, injury-related death, and motor vehicle crashes (nonsignificant differences) (Dinh-Zarr et al., 1999). Two recent reviews of studies in primary care settings presented the magnitude of the impact of BI in clinically relevant ways. In Beich et al.'s review including 2,784 participants, BI decreased the proportion of drinkers of risky amounts by 12%, from 69% to 57% (Beich et al., 2003). Bertholet et al. (2005) in studies including 5,639 participants found BI decreased consumption by 15% (mean 38 g/wk).

Despite this evidence for efficacy, the literature raises and leaves many questions to be answered by future research on BIs. Although 2 studies have found significant decreases in hospital utilization and nonsignificant decreases in mortality in participants assigned to BI, few studies have reported these outcomes (Fleming et al., 2002; Kristenson et al., 2002). Of note, in 1 of these studies, the serum γ-glutamyltransferase (GGT) level did not differ by randomized group, but alcohol-related deaths (not total mortality) did decrease significantly with BI (Kristenson et al., 2002).

Questions remain regarding efficacy in diverse settings and populations. Efficacy studies with no-treatment control groups have had strict exclusion criteria, including primarily patients not seeking treatment and not including patients with other drug use, psychiatric illness, or alcohol dependence. People with alcohol dependence have been studied, however, in studies in which the control group is more extensive treatment and in which patients are seeking treatment; these studies find BI to have similar efficacy as more extensive treatment (Moyer et al., 2002). In a systematic review of controlled studies of BI in general hospitals, brief intervention did not decrease consumption in any of 6 studies of inpatients, decreased alcohol problems in 4 of 6 studies, and decreased GGT in 2 of 4 studies (Emmen et al., 2004). No randomized study has tested BI delivered by ED clinicians, although studies of brief counseling in the ED have reported decreases in alcohol-related problems but not consumption (Longabaugh et al., 2001; Monti et al., 1999). A randomized trial in hospitalized trauma center patients found a significant decrease in consumption with BI and nonsignificant decreases in repeat emergency care and trauma admissions (Gentilello et al., 1999). One study in older adults found efficacy for BI in decreasing consumption (Fleming et al., 1999). Positive findings in pregnant women and adolescents appear in subgroup analyses of a small number of trials (Chang et al., 1999; Handmaker et al., 1999; Monti et al., 1999). Most adults in the United States have more than 1 behavioral health risk factor. Only 3 studies have addressed alcohol use in this context and in only 1 did alcohol intervention decrease consumption (Coups et al., 2004; Goldstein et al., 2004).

Questions also remain regarding the effectiveness of BI. One meta-analysis asked whether screening and BI was efficient and whether the yield in terms of benefit was worth the effort (Beich et al., 2003). Although cost–benefit analyses have found cost savings, none have used a general metric such as quality-adjusted life-years to allow comparison with other health care interventions (Fleming et al., 2002; Gentilello et al., 2005).

In summary, although the number of subjects studied in total is relatively small compared with efficacy studies for other medical conditions, the evidence for efficacy of BI is strong for modest decreases in consumption in people with nondependent unhealthy alcohol use not seeking treatment. However, the controlled trial evidence in general health care settings is limited with regards to efficacy in adolescents, older adults, people with dependence or comorbidity (e.g., drug dependence, psychiatric illness), and pregnant women and in general hospitals, trauma centers,

and EDs. Questions remain with regard to the impact of BI on outcomes other than consumption (e.g., alcohol-related problems, mortality) and cost-effectiveness. The minimum duration of counseling for effect or gain with increased duration or number of episodes of intervention are not clear. There has been little study of the importance of the type of clinician or relationship of the clinician to the patient in determining efficacy of BI. Finally, how to best achieve both greater efficacy and widespread implementation of BI in clinical practice remains a challenge.

BRIEF INTERVENTION FOR THE SPECTRUM OF ALCOHOL PROBLEMS IN MEDICAL HOSPITAL PATIENTS

Richard Saitz

Brief counseling has some efficacy in outpatient settings among nondependent people with unhealthy alcohol use, but has not been tested in randomized trials for the spectrum of unhealthy use (risky drinking through dependence) in hospitalized patients (Whitlock et al., 2004). Evidence for efficacy is limited and conflicting in general medical hospital settings (Emmen et al., 2004). This section summarizes preliminary results from a clinical trial that addresses these questions.

Medical inpatients in an urban general hospital identified as current risky drinkers by alcohol screening (>14 drinks/wk or >4 drinks/occasion for men; >11 drinks/wk or >3 drinks/occasion for women or the elderly) were randomized to a brief motivational intervention or to usual care. Primary outcomes of interest were: (1) self-reported linkage with alcoholism assistance (residential treatment, treatment including counseling or therapy, Alcoholics Anonymous, employee assistance program, or medication) by 3 months for patients with alcohol dependence and (2) changes in alcohol use at 12 months in all subjects. Regression analyses adjusted for imbalances in randomized groups. Primary analyses of assistance were stratified by gender; exploratory unadjusted analyses were stratified by both age and gender.

Of almost 8,000 hospitalized patients approached, 5,813 were screened for risky drinking, and 986 (17%) were currently drinking at least risky amounts, 524 were eligible, and 341 were enrolled. Most were men (71%), and black (45%), with mean age 44 years; 77% were alcohol dependent. Among dependent subjects, the effect of the intervention on linkage to alcoholism assistance appeared to differ by gender and age. Brief intervention improved linkage with assistance in women and younger men (Saitz et al., 2005). Furthermore, there were few significant effects of the intervention on consumption among all subjects with unhealthy alcohol use.

The prevalence of current alcohol problems in the hospital was high and most patients identified by screening had dependence. Regarding the effectiveness of BI, age and sex appeared to matter. In medical inpatients, BI showed some promise for linking some with alcohol dependence with assistance, but it may not be sufficient for changing consumption. In the hospital setting, effective intervention for alcohol problems may need to be more extensive or tailored to specific patient needs. Additional more complete analyses of these preliminary study results are in preparation.

BRIEF INTERVENTION IN THE PRENATAL SETTING FOR WOMEN USING TOBACCO AND OTHER DRUGS

Dace Svikis

Empirically tested strategies exist to identify pregnant women at risk for alcohol and other drug use. Despite numerous efforts at practitioner education and training, however, incorporation of such strategies into routine clinical practice remains problematic. Barriers cited for adoption of such strategies include lack of time, practitioner discomfort with queries about drinking and drug use, and lack of expertise in handling "yes" responses. This presentation looked at 2 alternative approaches for identifying pregnant women at risk for alcohol and drug problems as well as recent findings from a study of BI for caffeine.

First, "current smoking status" was examined as a screening tool to identify pregnant women at risk of heavy alcohol and/or illicit drug use. Participants were 412 pregnant women seeking care in an urban, hospital-based obstetrics clinic. All women provided informed consent. The sample was predominantly African American (64.7%), single/never married (65.9%), and unemployed (38.7%), with a mean age of 25.7 years (SD 5.7) and 12.5 years (SD 2.3) of education. Over 3/4 (76.8%) reported that their current pregnancy was unplanned. Nearly one-half (41.8%) of the sample reported daily use of tobacco (lifetime) and nearly one-third (32.0%) reported smoking in the 3 months prior to the first prenatal visit. Current smokers were more likely than nonsmokers to report ever having consumed at least 6 drinks (52.3 and 32.9%, $p<0.001$), having had alcohol-related blackouts (36.7 and 24.3%, $p<0.03$), and feeling guilty about their drinking (20.9 and 10.0%, $p<0.017$). Women who reported smoking in the past 3 months were also more likely than nonsmokers to report prenatal use of marijuana (23 and 9%, respectively, $p<0.001$). These findings confirm that current smoking status is in fact a useful screening tool for identifying women at risk for heavy drinking and other drug use.

More recently, our program of research has examined several types of BIs (practitioner advice, behavioral incentives, and motivational interviewing) with alternative target behaviors (e.g., tobacco, caffeine). In all cases, alcohol and other substance use was monitored to see if reductions in smoking were associated with similar reductions in drinking and other drug use. The presentation described a recently completed study of brief practitioner advice targeting caffeine use during pregnancy (Svikis and Jones, 2005).

Caffeine is the most frequently consumed drug in the United States (Senay, 1983) and recent studies have shown

regular, frequent caffeine use to be a form of drug dependence (Hughes et al., 1992). A number of studies have shown adverse consequences of prenatal caffeine use. Despite potential dangers, however, caffeine remains one of the most frequently used nonnutritional substances during pregnancy. The present study examined the effectiveness of a strong physician message to eliminate caffeine use during pregnancy on quantity and frequency of caffeine use. The study also looked at predictors of a woman's ability to quit caffeine use.

Participants were 45 pregnant women who completed a series of questionnaires at the first, second, and sixth prenatal visits as well as a personal interview. The study found that women with a family history of alcoholism and a diagnosis of caffeine dependence (lifetime) were less likely to maintain lower levels of caffeine use during pregnancy. The study also found that this group of women tended to consume more alcohol before pregnancy ($p<0.058$) and were more likely to use other drugs (lifetime) ($p<0.05$). Study findings suggest that less social stigma associated with caffeine use during pregnancy may make it a useful prototype for the study of alcohol and other drug use during the prenatal period.

BRIEF INTERVENTION PERFORMED BY EMERGENCY PRACTITIONERS FOR HAZARDOUS AND HARMFUL DRINKERS IN THE EMERGENCY DEPARTMENT

Gail D'Onofrio

Unhealthy alcohol use is prevalent in ED populations and covers a wide spectrum of misuse, ranging from at-risk or hazardous drinking to dependence (Cherpitel, 1999; Saitz, 2005). Hazardous drinking is defined as exceeding the National Institute on Alcohol Abuse and Alcoholism (NIAAA) guidelines for low-risk drinking (NIAAA, 2005). These drinkers are at risk for future medical, social, or legal consequences. Harmful drinkers are those patients who present with a negative consequence related to alcohol. Previous studies in the ED have demonstrated that BI has decreased negative consequences but that both intervention and control groups similarly decreased their drinking (Longabaugh et al., 2001; Monti et al., 1999). Only one study of hospitalized trauma patients has demonstrated that a 30-minute BI significantly reduced alcohol consumption (Gentilello et al., 1999). All of these studies, however, have used specialized counselors, such as social workers, or masters, or doctoral-level psychologists.

The purpose of this study was to develop and test a BI, namely, the Brief Negotiation Interview (BNI), on harmful and hazardous drinkers performed by existing staff (D'Onofrio et al., 2005b). Emergency practitioners (EPs) were defined as senior emergency medicine residents, faculty, and physician associates. The EPs were trained during a 2-hour skills-based session and at a later date tested for adherence to and competence with the BNI using a standardized patient scenario and a checklist of critical components of the BNI. The intervention was performed in the context of a randomized-controlled trial (RCT) conducted in an urban teaching hospital to test the efficacy of BNI on patient alcohol consumption and negative consequences. Fifty-eight EPs were trained and 53 (91%) passed the proficiency exam; 96% passed after remediation. Two EPs left the institution prior to remediation. During the course of the RCT, 247 BNIs were performed by 47 EPs. The mean number of BNIs performed per EP was 5.28 (SD +4.91, range 0–28) The mean duration was 7.75 minutes (SD ± 3.18, range 4–24). Three BNIs were not performed because of critical illness: a bowel obstruction leading to surgical intervention, an acute myocardial infarction, and an evolving altered mental status because of a subdural hematoma.

The results of the RCT were reported in abstract form (D'Onofrio et al., 2005a). A BI performed by EPs was acceptable and feasible to perform in a real-world setting. However, to date there is a paucity of evidence supporting BI in the ED setting. A recent article by Maio reported that an interactive, computer BI was not effective in decreasing alcohol misuse for adolescents aged 14 to 18 (Maio et al., 2005).

There are many methodological challenges that may influence the results and limit the generalizability of ED studies done to date (D'Onofrio and Degutis, 2004). First, all of the studies have had high refusal rates; as many as 47% of patients refused to participate. Refusal rates are particularly high in the adolescent studies because parents need to give consent. In addition patients in the standard care conditions actually received much more than standard care. This information was imparted both in the lengthy assessments and in the form of advice and handouts that may in fact have functioned as interventions themselves. This contact with counselors and information goes beyond the standard of care commonly seen in ED settings. It is also possible that these studies were not powered sufficiently to find significance in negative consequences and repeat ED and hospital visits, as these events are relatively infrequent.

It is clear that further research is needed to delineate the possible effectiveness of BI in the ED setting. Past studies have included a wide range of ages of participants and inclusion criteria. It is possible for example that the admitted trauma patient, by definition sustaining a severe injury, was more motivated to change than most patients who are treated and released. Perhaps interventions targeted for specific age groups and conditions may be more successful.

COST-EFFECTIVENESS OF BRIEF INTERVENTION FOR UNHEALTHY ALCOHOL USE

Kevin L. Kraemer

Many questions remain about the value and feasibility of extending alcohol screening and BI to diverse practice settings and populations (Beich et al., 2003). Cost-effectiveness modeling can be an important tool for addressing

some of these questions (Fleming et al., 2000, 2002; French, 2001; Gold et al., 1996; Wutzke et al., 2001).

In this study, we estimated the cost-effectiveness of alcohol screening and intervention in primary care using decision analytic and computer simulation techniques. Our analysis closely followed the methodological recommendations of the Panel on Cost-Effectiveness in Health and Medicine (Gold et al., 1996). We designed a Markov decision model to track 6 alcohol-related health states (abstinence, safe drinking, at-risk drinking, alcohol abuse, alcohol dependence, and alcohol dependence in recovery). Model parameters were obtained from published values for alcohol screening sensitivity/specificity, prevalence of alcohol problems in primary care, efficacy of BI, transition between alcohol-related health states, mortality, costs for alcohol screening and intervention, and lifetime health care costs. Where published data were not available, we made simplifying assumptions that reflected actual primary care practices or were biased against alcohol screening and intervention. We obtained standard gamble utility estimates for each alcohol-related health state from a clinic/community sample (Kraemer et al., 2005). We used separate models for men and women because transition probabilities between health states were substantially different by gender. We calculated the incremental cost-effectiveness ratio [cost per quality-adjusted life-year (QALY)] from the societal perspective and discounted costs and benefits at a rate of 3%. Under baseline conditions, the screening and intervention strategy dominated and was cost saving compared with the no screening strategy. Screening and intervention yielded a savings of $300 and a gain of 0.05 QALYs per man or woman screened. Results were robust to a range of alcohol use prevalence, intervention efficacy estimates, costs, utilities, and discount rates.

These preliminary results indicate that screening and intervention for unhealthy alcohol use in primary care can extend quality-adjusted life and save money. Cost-effectiveness analysis can be a useful tool for informing clinical practice, policy, and research related to alcohol prevention and treatment in diverse practice settings and populations.

DISCUSSION

Harold Perl

The participants in this symposium highlighted 3 sets of fundamental and inter-related questions that deserve attention.

1. How can the research community give clinicians the tools that they want, need, and will use? This is not the same as "getting them to do it." Rather, it is a matter of demonstrating effectiveness and benefit at a tolerable and acceptable cost.
2. What are the essential ingredients of BIs? How and why does it work? Would a simpler—or more "stripped-down"—package be more effective or more acceptable?
3. How can BIs for alcohol problems be embedded and incorporated into the broader health care system over time, in a sustainable and viable manner?

The answers will depend on additional research but some observations can be made now.

Alcohol treatment has often been viewed as an episodic intervention for an acute disorder. More recently, we have come to view alcohol problems as more chronic and long term. Consequently, we need to make certain assumptions about the trajectory of this disorder's progression. In other words, an individual may go from a nondrinker to the first drink to a low-risk drinker to a high-risk drinker to harmful drinker to crossing the diagnostic threshold for abuse and/or dependence. An individual may move along this pathway—or not. Another may return to an "earlier" point—or not. One may "respond to treatment"—or not. Whatever the pattern may be, however, we must stop focusing on only the single point in time and consider the interventions we plan—and the outcomes we measure—in terms of longer stretches of these trajectories.

The basic principles of prevention and early intervention suggest that we identify problems and act as early in the pathway as possible. Yet many of the persons who may be most in need of intervention—those with alcohol abuse or dependence—do not even acknowledge themselves to be in need, much less those persons with risky or problem drinking. We cannot expect them to present themselves at the entrance of the alcohol treatment system and request interventions. One alternative strategy would be to cast our lines and nets in different parts of the stream because some fishing holes have more fish than others. Many persons with alcohol problems pass through the medical care system; the mental health care system; prisons, jails, and other correctional settings; college and university settings; the welfare and social service system; and certain occupational and workplace settings. These are settings where screening and intervention holds the promise of bringing new patients in need into care.

The research presented in this symposium demonstrates rather clearly that screening and BI can work, though they do not always work, and challenges abound. For example, one methodological issue needs to be considered in understanding the research on BIs. "Assessment reactivity" is meant to convey the behavior change that is posited to result from the intense assessment protocols that are used to measure drinking behaviors in research studies and is typically put forth as an explanation for observations of positive changes in drinking behaviors across both BI and "comparison" or "control" conditions. However, "assessment reactivity" has begun to take on a nearly pejorative taint, as if it were an adverse event to be avoided at all costs. In fact, such observations may actually be harbingers of a desirable eventuality. That is to say, if the assessment protocol is serving as a form of attenuated treatment, then it may be possible to isolate and identify

what aspects of that "bare-bones" intervention are the most potent and make them the centerpiece of a basic intervention that is truly brief. Presumably, such a potent yet quick intervention would be more likely to be acceptable and sustainable in the medical care and other systems.

Therefore, the next step is to learn how to improve BIs and incorporate them in the current systems of care so they are routine, efficacious, and typical components of the process. They should not be experienced as additional "burdens" to be borne by the most motivated practitioners. Rather, they need to be assimilated by the system so that they are sustainable to the setting (both fiscal and time-management factors) as well as acceptable (even attractive) to policy makers, clinical leaders, and on-the-ground clinical providers.

Current (and future) technologies need to be exploited to facilitate this assimilation, so that the BIs are transparent and seamless. New computer and Web-based technologies may have much to contribute, including electronic reminders, computer-assisted screening tools, and integrating with electronic medical records.

The recently revised NIAAA *Clinicians Guide: Helping Patients Who Drink Too Much* (NIAAA, 2005) is another tool that can be used to facilitate the incorporation of BIs into a variety of settings. Originally designed for use by physicians and other primary medical caregivers, the Guide has been broadened to make it more relevant and useful to mental health caregivers, who often treat persons with unhealthy alcohol use. The Guide recommends a screening protocol that starts with a single question to make the process as easy and quick as possible. Once a patient is identified, then the clinician can more extensively assess the actual level of risk or disorder. The revised Guide also recommends distinct and parallel courses of intervention that can be followed differentially for patients who are at-risk drinkers or have alcohol use disorders. A variety of assessment, intervention, and follow-up tools are also provided.

With further attention to the best balance of cost, efficacy, and acceptability, and with further research on BIs in diverse populations and settings, the promise of routinely and effectively addressing the spectrum of unhealthy alcohol use may be achieved.

REFERENCES

Ballesteros J, Duffy JC, Querejeta I, Arino J, Gonzalez-Pinto A (2004) Efficacy of brief interventions for hazardous drinkers in primary care: systematic review and meta-analyses. Alcohol Clin Exp Res 28:608–618.

Beich A, Gannik D, Malterud K (2002) Screening and brief intervention for excessive alcohol use: qualitative interview study of the experiences of general practitioners. BMJ 325:870–872.

Beich A, Thorsen T, Rollnick S (2003) Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. BMJ 327:536–542.

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

Bien TH, Miller WR, Tonigan JS (1993) Brief interventions for alcohol problems: a review. Addiction 88:315–335.

Chang G, Wilkins-Haug L, Berman S, Goetz MA (1999) Brief intervention for alcohol use in pregnancy: a randomized trial. Addiction 94:1499–1508.

Cherpitel CJ (1999) Drinking patterns and problems: a comparison of primary care with the emergency room. Subst Abus 20:85–95.

Coups EJ, Gaba A, Orleans CT (2004) Physician screening for multiple behavioral health risk factors. Am J Prev Med 27:34–41.

Dinh-Zarr T, Diguiseppi C, Heitman E, Roberts I (1999) Preventing injuries through interventions for problem drinking: a systematic review of randomized controlled trials. Alcohol Alcohol 34:609–621.

D'Onofrio G, Degutis LC (2002) Preventive care in the emergency department: screening and brief intervention for alcohol problems in the emergency department: a systematic review. Acad Emerg Med 9: 627–638.

D'Onofrio G, Degutis LC (2004) Screening and brief intervention in the emergency department. Alcohol Res Health 28:63–71.

D'Onofrio G, Degutis LC, Fiellin DA, Pantalon MV, Busch SH, Chawarski MC, O'Connor PG (2005a) Emergency practitioner performed brief intervention for harmful and hazardous drinkers in the emergency department. Alcohol Clin Exp Res 29(suppl): 179A.

D'Onofrio G, Pantalon MV, Degutis LC, Fiellin DA, O'Connor PG (2005b) Development and implementation of an emergency practitioner-performed brief intervention for hazardous and harmful drinkers in the emergency department. Acad Emerg Med 12:249–256.

Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H (2004) Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systematic review. BMJ 328:318.

Fleming MF, Manwell LB, Barry KL, Adams W, Stauffacher EA (1999) Brief physician advice for alcohol problems in older adults: a randomized community-based trial. J Fam Pract 48:378–384.

Fleming MF, Mundt MP, French MT, Manwell LB, Stauffacher EA, Barry KL (2000) Benefit–cost analysis of brief physician advice with problem drinkers in primary care settings. Med Care 38:7–18.

Fleming MF, Mundt MP, French MT, Manwell LB, Stauffacher EA, Barry KL (2002) Brief physician advice for problem drinkers: long-term efficacy and benefit–cost analysis. Alcohol Clin Exp Res 26: 36–43.

French MT (2001) Economic evaluation of alcohol treatment services. Recent Dev Alcohol 15:209–228.

Gentilello LM, Ebel BE, Wickizer TM, Salkever DS, Rivara FP (2005) Alcohol interventions for trauma patients treated in emergency departments and hospitals: a cost benefit analysis. Ann Surg 241: 541–550.

Gentilello LM, Rivara FP, Donovan DM, Jurkovich GJ, Daranciang E, Dunn CW, Villaveces A, Copass M, Ries RR (1999) Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann Surg 230:473–480.

Gold MR, Russell LB, Seigel JE, Weinstein MC (1996) Cost-Effectiveness in Health and Medicine. Oxford University Press, New York.

Goldstein MG, Whitlock EP, DePue J (2004) Multiple behavioral risk factor interventions in primary care summary of research evidence. Am J Prev Med 27:61–79.

Handmaker NS, Miller WR, Manicke M (1999) Findings of a pilot study of motivational interviewing with pregnant drinkers. J Stud Alcohol 60:285–287.

Hughes JR, Oliveto AH, Helzer JE, Higgins ST, Bickel WK (1992) Should caffeine abuse, dependence, or withdrawal be added to DSM-IV and ICD-10? Am J Psychiatry 149:33–40.

Kahan M, Wilson L, Becker L (1995) Effectiveness of physician-based interventions with problem drinkers: a review. CMAJ 152:851–859.

Kraemer KL, Roberts MS, Horton NJ, Palfai T, Samet JH, Freedner N, Tibbetts N, Saitz R (2005) Health utility ratings for a spectrum of alcohol-related health states. Med Care 43:541–550.

Kristenson H, Osterling A, Nilsson JA, Lindgarde F (2002) Prevention of alcohol-related deaths in middle-aged heavy drinkers. Alcohol Clin Exp Res 26:478–484.

Longabaugh R, Woolard RE, Nirenberg TD, Minugh AP, Becker B, Clifford PR, Carty K, Sparadeo F, Gogineni A (2001) Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. J Stud Alcohol 62:806–816.

Maio RF, Shope JT, Blow FC, Gregor MA, Zakrajsek JS, Weber JE, Nypaver MM (2005) A randomized controlled trial of an emergency department-based interactive computer program to prevent alcohol misuse among injured adolescents. Ann Emerg Med 45:420–429.

Monti PM, Colby SM, Barnett NP, Spirito A, Rohsenow DJ, Myers M, Woolard R, Lewander W (1999) Brief intervention for harm reduction with alcohol-positive older adolescents in a hospital emergency department. J Consult Clin Psychol 67:989–994.

Moyer A, Finney JW, Swearingen CE, Vergun P (2002) Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction 97:279–292.

NIAAA, National Institutes of Health, U.S. Department of Health and Human Services. *Helping Patients Who Drink Too Much: A Clinician's Guide*. NIH Publication No. 05-3769, Revised. Rockville, MD: 2005. Available at: http://www.niaaa.nih.gov/publications/Practitioner/guide.pdf. Accessed September 28, 2005.

Poikolainen K (1999) Effectiveness of brief interventions to reduce alcohol intake in primary health care populations: a meta-analysis. Prev Med 28:503–509.

Saitz R (2005) Unhealthy alcohol use. N Engl J Med 352:596–607.

Saitz R, Palfai TP, Horton NJ, Cheng DM, Freedner N, Dukes K, Kraemer KL, Roberts MS, Samet JH (2005) Brief intervention for the spectrum of alcohol problems on a hospital medical service: age and sex matter. Alcohol Clin Exp Res 29(suppl):178A.

Senay EC (1983) Substance Abuse Disorders in Clinical Practice. 1st ed. Wright, Boston, MA.

Svikis DS, Jones HE (2005) Brief intervention for alcohol in the prenatal setting in women using tobacco and other drugs. Alcohol Clin Exp Res 29(suppl):178A.

Whitlock EP, Polen MR, Green CA, Orleans T, Klein J (2004) Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. preventive services task force. Ann Intern Med 140: 557–568.

Wilk AI, Jensen NM, Havighurst TC (1997) Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. J Gen Intern Med 12:274–283.

Wutzke SE, Shiell A, Gomel MK, Conigrave KM (2001) Cost effectiveness of brief interventions for reducing alcohol consumption. Soc Sci Med 52:863–870.

The Severity of Unhealthy Alcohol Use in Hospitalized Medical Patients

The Spectrum is Narrow

Richard Saitz, MD, MPH,[1,2,3] Naomi Freedner, MPH,[1,2] Tibor P. Palfai, PhD,[2,4] Nicholas J. Horton, ScD,[5] Jeffrey H. Samet, MD, MA, MPH[1,6]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA; [2]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA; [3]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [4]Department of Psychology, College of Arts and Sciences, Boston University, Boston, MA, USA; [5]Department of Mathematics, Smith College, Northampton, MA, USA; [6]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA.

BACKGROUND: Professional organizations recommend screening and brief intervention for unhealthy alcohol use; however, brief intervention has established efficacy only for people without alcohol dependence. Whether many medical inpatients with unhealthy alcohol use have nondependent use, and thus might benefit from brief intervention, is unknown.

OBJECTIVE: To determine the prevalence and spectrum of unhealthy alcohol use in medical inpatients.

DESIGN: Interviews of medical inpatients (March 2001 to June 2003).

SUBJECTS: Adult medical inpatients (5,813) in an urban teaching hospital.

MEASUREMENTS: Proportion drinking risky amounts in the past month (defined by national standards); proportion drinking risky amounts with a current alcohol diagnosis (determined by diagnostic interview).

RESULTS: Seventeen percent (986) were drinking risky amounts; 97% exceeded per occasion limits. Most scored ≥ 8 on the Alcohol Use Disorders Identification Test, strongly correlating with alcohol diagnoses. Most of a subsample of subjects who drank risky amounts and received further evaluation had dependence (77%).

CONCLUSIONS: Drinking risky amounts was common in medical inpatients. Most drinkers of risky amounts had dependence, not the broad spectrum of unhealthy alcohol use anticipated. Screening on a medicine service largely identifies patients with dependence—a group for whom the efficacy of brief intervention (a recommended practice) is not well established.

KEY WORDS: hospital; inpatient; alcohol; screening; brief intervention.
DOI: 10.1111/j.1525-1497.2006.00405.x
J GEN INTERN MED 2006; 21:381–385.

People with unhealthy alcohol use often go unidentified and do not receive timely care despite the existence of brief, valid screening tools.[1] Given the availability of these tools and the magnitude of alcohol-related health problems, professional organizations recommend that clinicians screen for unhealthy alcohol use, and, when indicated, conduct a brief intervention.[2]

In 1990, the Institute of Medicine (IOM) described a spectrum of unhealthy alcohol use in the general population.[3] The spectrum included drinking risky amounts (amounts that increase the risk of health consequences); problem drinking (use associated with consequences); harmful drinking and alcohol abuse (diagnoses characterized by recurrent consequences); and alcohol dependence (alcoholism, the most severe alcohol problem). The IOM emphasized that most people with unhealthy alcohol use were not alcohol dependent and that they would likely benefit from brief intervention.

Few studies have characterized this spectrum in medical inpatients through systematic screening with validated measures. However, determining whether medical inpatients as a whole have a range of unhealthy alcohol use has important treatment implications. It can help establish whether they are likely to benefit from screening and brief intervention—a currently recommended practice that has proven helpful for drinkers with nondependent unhealthy alcohol use but has less established efficacy for those with alcohol dependence.[4]

Therefore, this study aimed to characterize unhealthy alcohol use in medical inpatients. We hypothesized that the prevalence of unhealthy alcohol use would be high and that the spectrum would be broad.

METHODS

Design

We conducted a cross-sectional study to determine the prevalence and spectrum of unhealthy alcohol use in medical inpatients. First, we screened patients for drinking risky amounts. A subgroup of patients who screened positive enrolled in a clinical trial of an alcohol brief intervention (i.e., "enrolled subsample") and was evaluated more extensively. We compared this enrolled subsample to subjects who were drinking risky amounts but did not enroll (i.e., the "nonenrolled subsample") to determine the similarities between these groups and thus make inferences about medical inpatients with unhealthy alcohol use.

The authors have no conflicts of interest to declare.

Preliminary results were presented at the following meetings: Research Society on Alcoholism, June 2004, Vancouver, BC, Canada; Association for Medical Education and Research in Substance Abuse, November 2003, Baltimore, MD; and American Public Health Association, November 2003, San Francisco, CA.

Address correspondence and requests for reprints to Dr. Saitz: Boston Medical Center, 91 East Concord Street, Suite 200, Boston, MA 02118 (e-mail: rsaitz@bu.edu).

Manuscript received September 16, 2005
Initial editorial decision November 3, 2005
Final acceptance December 27, 2005

Subjects

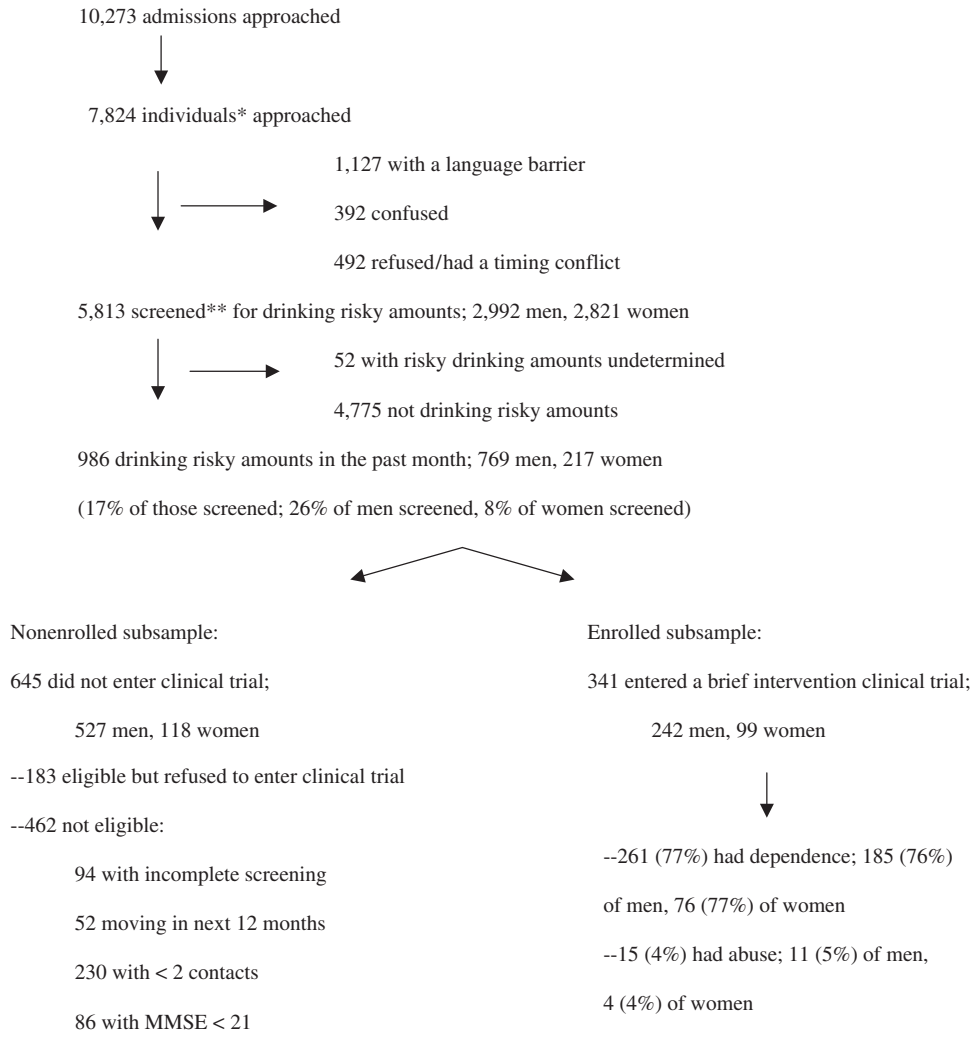
We recruited subjects from the inpatient medicine service of a large, urban teaching hospital between March 2001 and June 2003. On weekdays, trained research associates (RAs) reviewed an admissions database and approached all patients who were ≥ 18 years, hospitalized on the medicine service, and whose physicians permitted contact. Patients who were fluent in English or Spanish and gave oral consent were screened.

The Institutional Review and Privacy Boards of Boston University Medical Center approved this study. We secured additional protection with a Certificate of Confidentiality from the National Institute on Alcohol Abuse and Alcoholism (NIAAA).

Measurements

Research associates collected demographic data and administered the Alcohol Use Disorders Identification Test (AUDIT) by interview[5] (online Appendix A). For the first 7 months, RAs asked subjects with an AUDIT score of ≥ 8 (the validated cut-off) their average number of drinking days per week, average number of drinks consumed on a typical day, and maximum number of drinks consumed on an occasion to better characterize current (past month) alcohol use.[6] For the remaining 22 months of the study, RAs asked the additional questions to all who drank more than "never" in the past 12 months (determined by the AUDIT's first item). We changed our screening approach to maximize identification of drinkers of risky amounts; the newer approach allowed us to identify all of these drinkers, not only those who had reached a specific threshold on the AUDIT. Similar to NIAAA guidelines,[5] we defined risky amounts as >14 standard drinks per week or ≥ 5 drinks per occasion for men (>11 and ≥ 4, respectively, for both women and people ≥ 66 years).

Research associates asked subjects drinking risky amounts about their readiness to change drinking using a visual analogue scale ranging from 0 to 10[7] and then established subjects' eligibility for a brief intervention trial. Eligible subjects (provided 2 contacts, did not anticipate mov-



*For individuals with more than 1 admission, we included only data gathered during our first contact with the individual.
**During the first 7 months of the study, before we changed screening criteria, 22% of the 5,813 subjects were screened.

FIGURE 1. Process of screening medical inpatients and further assessment of subjects enrolled in a study of an alcohol brief intervention.

ing from the area in the next year, had a Mini-Mental State Examination score of ≥21,[8] and provided written informed consent) underwent additional evaluation. Research associates assessed this "enrolled subsample" for current (past year) diagnoses of alcohol abuse and dependence according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition IV* (*DSM IV*) using the Composite International Diagnostic Interview Alcohol Module[9]; alcohol problems with the Short Inventory of Problems[10]; medical comorbidity[11]; and alcohol treatment utilization. Structured medical record review (by author R.S.) determined the principal diagnosis of the current admission and current alcohol-related diagnoses.

Analyses

We conducted analyses with SAS/STAT software.[12] We compared the enrolled and nonenrolled subsamples with the χ^2 test, Fisher's exact test, 2-sample *t*-test, and Wilcoxon rank sum test, as appropriate. Reported *P* values are 2-tailed.

RESULTS

Research associates approached 10,273 admissions (representing 7,824 individuals) and screened 5,813 individuals (2,992 [51%] male; 2,821 [49%] female) (Fig. 1). Of those screened, 986 (17%) reported drinking risky amounts in the past month. A higher percentage of men than women screened positive (26% vs 8%, $P<.0001$).

Almost all drinkers of risky amounts reported exceeding per occasion drinking limits (Table 1). Most had AUDIT scores of ≥8; 38% had scores of ≥20 (indicating need for alcohol treatment). Readiness to change was high.

Among the drinkers of risky amounts, 341 entered a brief intervention trial (i.e., the enrolled subsample). To make inferences about medical inpatients with unhealthy alcohol use, we compared the enrolled subsample with subjects not enrolled in the trial. The enrolled subsample had a lower percentage of men, higher percentage of blacks, and lower mean age. However, both groups were similar on most alcohol use characteristics, including AUDIT scores. When differences occurred (i.e., maximum drinks per occasion and readiness), they were small.

Table 1. Characteristics of Medical Inpatients Drinking Risky Amounts of Alcohol

	Nonenrolled Subsample* (*N*=645)	Enrolled Subsample* (*N*=341)	*P* Value
Male, number (%)	527 (82)	242 (71)	<.001
Race	—	—	<.001
Hispanic, number (%)	89 (14)	30 (9)	—
Black, number (%)	209 (32)	155 (45)	—
White, number (%)	256 (40)	133 (39)	—
Other, number (%)	91 (14)	23 (7)	—
Age, mean y (SD)	48 (12)	44 (11)	<.001
AUDIT	—	—	—
≥8, number (%)	527 (82)	294 (86)	.07
≥20, number (%)	234 (36)	143 (42)	.08
Exceeded maximum per occasion amounts (≥5 drinks for men, ≥4 for women and people 66 y or older), number (%)	624 (97)	337 (99)	.06
Maximum number of drinks per occasion, median (IQR)†	12 (6 to 20)	12 (7 to 24)	.004
Number of drinks per week, median (IQR)	21 (8 to 60)	24 (8 to 72)	.46
Readiness to change, median (IQR)	8 (2 to 10)	8 (5 to 10)	.02
Diagnoses of current alcohol use disorders	—	—	—
Alcohol dependence, number (%)	N/D	261 (77)	—
Alcohol abuse, number (%)	N/D	15 (4)	—
No diagnosis, number (%)	N/D	65 (19)	—
Alcohol-related medical diagnosis, lifetime, self-report, number (%)	N/D	323‡ (95)	—
Alcohol-related medical diagnosis, past 3 mo, self-report, number (%)	N/D	285§ (84)	—
Alcohol problems (≥1), past 3 mo, number (%)	N/D	304 (89)	—
Alcohol-related medical diagnosis, medical record review, number (%)	N/D	156 (46)	—
Principal diagnosis, current admission	—	—	—
Rule out myocardial infarction, number (%)	N/D	61 (18)	—
Alcohol-related, number (%)	N/D	51 (15)	—
Reactive airways diseases, number (%)	N/D	36 (11)	—
Pancreatitis, number (%)	N/D	33 (10)	—
Cellulitis, number (%)	N/D	22 (6)	—
Diabetes, number (%)	N/D	14 (4)	—
Alcohol assistance, past 3 mo	—	86‡ (25)	—
Residential program, number (%)	N/D	11‡ (3)	—
Outpatient treatment (e.g., counseling or therapy), number (%)	N/D	16‡ (5)	—
Mutual-help group (e.g., Alcoholics Anonymous), number (%)	N/D	65 (19)	—
Employee assistance program, number (%)	N/D	1 (0.3)	—
Naltrexone or disulfiram, number (%)	N/D	5 (1)	—
Detoxification program, past 3 mo, number (%)	N/D	56‖ (17)	—

**The nonenrolled subsample underwent screening only; the enrolled subsample entered a clinical trial and underwent both screening and further evaluation.*

†25th and 75th percentiles reported.

‡N=340; §N=338; ‖N=339.

SD, standard deviation; IQR, interquartile range; N≠D, not determined.

We evaluated the enrolled subsample more extensively. Most (77%) had current alcohol dependence. Alcohol-related medical diagnoses were common. In the past 3 months, 25% had received alcohol assistance.

CONCLUSIONS

A substantial proportion of medical inpatients had unhealthy alcohol use. However, contrary to our hypothesis, the spectrum of use was narrow: most patients screening positive had alcohol dependence.

Numerous studies have assessed the prevalence of alcohol problems among hospitalized patients.[13–17] Prevalence figures vary widely—from 2% to 60%[13,14]—and are influenced by many factors (e.g., patient population, definitions, assessments). In the smaller subset of studies that have both examined current alcohol problems among medical inpatients and used validated measures, prevalence ranges from 3% to 47%.[15,16] The prevalence of current drinking of risky amounts in our study falls in the middle range of most commonly reported estimates. Notably, the prevalence of dependence—determined by a diagnostic instrument—among these drinkers of risky amounts is very high.

This study has some limitations and several notable strengths. First, the prevalence of drinking risky amounts was undetermined for some patients (Fig. 1). However, this study—unlike many others—attempted to screen all medical admissions with validated tools that identify the spectrum of unhealthy alcohol use. Second, diagnoses were determined only for the enrolled subsample. However, a similar proportion of the nonenrolled subsample would presumably also have dependence because the groups were quite similar (e.g., the majority of both groups scored ≥ 8 on the AUDIT, strongly correlating with abuse or dependence diagnoses). Third, despite the change in screening criteria, the proportions of subjects identified as drinking risky amounts with the original criteria and the changed criteria were similar (19% and 17%, respectively). Fourth, RAs reviewed admissions databases only on weekdays and may have missed patients admitted on the weekends, although most weekend admissions remain hospitalized on a weekday. Lastly, we defined weekly risky amounts for women as >11 rather than >7 drinks per week, NIAAA's recommended cutoff,[5] to ensure that the trial could detect intervention effects (to avoid floor effects). Still, very few additional women would have been classified as drinkers of risky amounts using the NIAAA cutoff (6 of 2,592 women not drinking >11 drinks per week).

Despite these limitations, our findings have implications for screening and intervention practices. Current practice guidelines recommend screening and brief intervention, when indicated, for unhealthy alcohol use. If widely adopted in general hospitals, screening will reveal that a substantial proportion of patients (17%) has unhealthy alcohol use, and that most of these patients have dependence. However, whether this population may benefit from brief intervention is not well established. Studies on brief intervention's efficacy for inpatients are limited (e.g., variable exclusion criteria, few specific to medical inpatients) and conflicting. A systematic review found that brief intervention did not reduce consumption among inpatients on various units in all 6 trials examined.[18] However, other studies specific to medical inpatients suggest that brief interventions may be useful.[19,20]

A substantial proportion of medical inpatients has unhealthy alcohol use, mainly alcohol dependence. Because the efficacy of known approaches to help these primarily alcohol-dependent medical inpatients is not yet well established—despite the substantial contribution of alcohol to inpatient medical problems—enhanced brief intervention strategies as well as new practical and effective interventions must be developed and disseminated to decrease the direct complications associated with unhealthy alcohol use.

The authors are indebted to Rosanne T. Guerriero, MPH, for efforts in manuscript preparation and editing; the staff and patients of the medicine inpatient service; CARE Unit RAs at Boston Medical Center; the staff and house staff of the Boston University Internal Medicine Residency Training Program; and data management staff at DM-STAT, Malden, MA. This study was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA RO1 12617) and a General Clinical Research Center Grant from the National Center for Research Resources (M01 RR00533).

REFERENCES

1. **Saitz R.** Unhealthy alcohol use. N Engl J Med. 2005;352:596–607.
2. **US Preventive Services Task Force.** Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. Ann Intern Med. 2004;140:554–6.
3. **Institute of Medicine.** Broadening the Base of Treatment for Alcohol Problems: Report of a Study by a Committee of the Institute of Medicine, Division of Mental Health and Behavioral Medicine. Washington, DC: National Academy Press; 1990.
4. **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.
5. **National Institutes of Health.** Helping Patients Who Drink Too Much: A Clinician's Guide. Bethesda, MD: National Institutes of Health; 2005.
6. **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.
7. **Rollnick S.** Behavior change in practice: targeting individuals. Int J Obes. 1996;20(suppl 1):S22–S26.
8. **Folstein MF, Folstein SE, McHugh PR.** "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
9. **American Psychiatric Association.** Diagnostic and Statistical Manual of Mental Disorders. 4th edn. Washington, DC: American Psychiatric Association; 1994.
10. **Miller WR, Tonigan J, Longabaugh R.** The Drinker Inventory of Consequences (DrInC). An Instrument for Assessing Adverse Consequences of Alcohol Abuse. Test Manual. Project MATCH Monograph Series, 4. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1995.
11. **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.
12. **SAS Institute, Inc.** SAS/STAT, Version 8.2 [software]. North Carolina: SAS Institute, Inc.; 1999.
13. **Moen R, Batey R.** Alcohol-related disease in hospital patients. Med J Aust. 1986;144:515–9.
14. **McCusker J, Cherubin CE, Zimberg S.** Prevalence of alcoholism in general municipal hospital population. NY State J Med. 1971;71:751–4.
15. **Corrigan GV, Webb MG, Unwin AR.** Alcohol dependence among general medical inpatients. Br J Addict. 1986;81:237–45.
16. **Dawson NV, Dadheech G, Speroff T, Smith RL, Schubert DS.** The effect of patient gender on the prevalence and recognition of alcoholism on a general medicine inpatient service. J Gen Intern Med. 1992;7:38–45.
17. **Roche AM, Freeman T, Skinner N.** From data to evidence, to action: findings from a systematic review of hospital screening studies for high risk alcohol consumption. Drug Alcohol Depend, In press. Available online November 23, 2005 at: doi:10.1016/j.drugalcdep.2005.10.011. Accessed December 13, 2005.

18. **Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H.** Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systematic review. BMJ. 2004;328:318.
19. **Chick J, Lloyd G, Crombie E.** Counselling problem drinkers in medical wards: a controlled study. BMJ (Clin Res Ed). 1985;290:965–7.
20. **Mcmanus S, Hipkins J, Haddad P, Guthrie E, Creed F.** Implementing an effective intervention for problem drinkers on medical wards. Gen Hosp Psychiatry. 2003;25:332–7.

Supplementary Material

The following supplementary material is available for this article online at www.blackwell-synergy.com

Appendix A.

doi:10.1111/j.1360-0443.2005.01228.x

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]

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.

Correspondence to: Jeffrey H. Samet, Professor of Medicine and Public Health, Boston University Schools of Medicine and Public Health, 91 E. Concord Street, Suite 200, Boston, MA 02118, USA. E-mail: jsamet@bu.edu
Submitted 9 May 2005; initial review completed 24 May 2005; final version accepted 24 May 2005

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 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

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

(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 *TR*aining and *E*ducation for *AM*erica) [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 multi-disciplinary 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 multi-disciplinary 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.

Journal of Urban Health: Bulletin of the New York Academy of Medicine, Vol. 83, No. 4
doi:10.1007/s11524-006-9061-6

Masculine Gender Roles Associated with Increased Sexual Risk and Intimate Partner Violence Perpetration among Young Adult Men

M. Christina Santana, Anita Raj, Michele R. Decker,
Ana La Marche, and Jay G. Silverman

ABSTRACT *This study sought to assess the association between traditional masculine gender role ideologies and sexual risk and intimate partner violence (IPV) perpetration behaviors in young men's heterosexual relationships. Sexually active men age 18–35 years attending an urban community health center in Boston were invited to join a study on men's sexual risk; participants (N = 307) completed a brief self-administered survey on sexual risk (unprotected sex, forced unprotected sex, multiple sex partners) and IPV perpetration (physical, sexual and injury from/need for medical services due to IPV) behaviors, as well as demographics. Current analyses included men reporting sex with a main female partner in the past 3 months (n = 283). Logistic regression analyses adjusted for demographics were used to assess significant associations between male gender role ideologies and the sexual risk and IPV perpetration behaviors. Participants were predominantly Hispanic (74.9%) and Black (21.9%); 55.5% were not born in the continental U.S.; 65% had been in the relationship for more than 1 year. Men reporting more traditional ideologies were significantly more likely to report unprotected vaginal sex in the past 3 months (OR_{adj} = 2.3, 95% CI = 1.2–4.6) and IPV perpetration in the past year (OR_{adj} = 2.1, 95% CI = 1.2–3.6). Findings indicate that masculine gender role ideologies are linked with young men's unprotected vaginal sex and IPV perpetration in relationships, suggesting that such ideologies may be a useful point of sexual risk reduction and IPV prevention intervention with this population.*

KEYWORDS *Sexual risk behaviors, Partner violence, Masculine ideology.*

INTRODUCTION

HIV disease is a leading cause of death for young men in the United States[1] with young men residing in urban centers being at particular risk.[2] Although men predominantly become infected via unprotected sex with a male partner and/or injection drug use,[2] heterosexual risk for HIV/AIDS is a risk for men and is the primary means of HIV infection for women.[2] Epidemiologic study with women suggests that sexual risk for HIV is more likely in the context of an abusive relationship due to both diminished control of sexual protection among abused

Santana and Raj are with the Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA; Decker and Silverman are with the Department of Society, Human Development and Health and Division of Public Health Practice, Harvard University School of Public Health, Boston, MA, USA; La Marche is with the Martha Eliot Health Center, Boston, MA, USA.

Correspondence: M. Christina Santana, MPH, Boston University School of Public Health, 715 Albany St., T2W Boston, MA 02118, USA. (E-mail: csantana@bu.edu)

women and increased likelihood of HIV infection among abusive men.[3] Studies with men also indicate that heterosexual males reporting IPV perpetration are significantly more likely to report higher risk sexual behavior, as compared with men not reporting recent IPV perpetration.[4,5] Researchers have suggested that traditional masculine gender role ideologies may be linking IPV perpetration and sexual risk behaviors among young heterosexual men;[5–7] however, no study to date has directly assessed this. The purpose of the current study is to assess associations between traditional masculine gender role ideologies and both sexual risk behaviors and recent IPV perpetration among young adult heterosexual men recruited from an urban community health center.

IPV Against Women and Sexual Risk for HIV/AIDS in Heterosexual Relationships

Extensive evidence from diverse populations of women demonstrates that IPV victimization, a health issue estimated to affect one in four U.S. women,[8–14] is significantly associated with low contraceptive and condom use as well as adverse sexual and reproductive health outcomes (e.g., pelvic pain, menstrual abnormalities, STD/HIV, unwanted pregnancy and multiple abortions) among women (see [15–17] for reviews). To date this research has primarily been limited to studies of female IPV victims, only very rarely including reports from male perpetrators. While research including women and girls' reports of male partner behavior points to a potential link between high rates of IPV perpetration and sexual risk among young adult men,[3,18–26] there has been little direct study of whether abusive partners pose greater sexual risk to women due to their own risky sexual or controlling behaviors. Within qualitative studies, battered women have reported forced pregnancy and prevention of contraception from their abusive male partners.[27,28] Less research has assessed this issue with men; however, research that has been conducted demonstrates significantly greater risk for sexual infidelity, multiple sex partners, unprotected sex, and forced unprotected sex among those reporting recent IPV perpetration.[4,5]

Men's Traditional Gender Role Ideologies Linked to IPV Perpetration and Sexual Risk Behaviors

Traditional gender role ideologies are the perceptions of how men and women are supposed to think and behave in society and within the context of heterosexual relationships. Extensive research has documented that men with more traditional gender role ideologies are significantly more likely to report sexual coercion and relationship violence.[29–37] Although much of this research has been conducted with middle-class White male samples (e.g., college students), these same identified ideologies (scripts, norms and attitudes) have also been reported in lower income and minority samples, [31,34,37] suggesting that these ideologies are not necessarily unique to specific racial/ethnic or class subgroups. These findings are supported by research indicating similar gender role ideologies among U.S. White, Black and Hispanic young adult males.[38]

Fewer studies have been conducted on associations between male gender roles and sexual risk behaviors. Those which have been conducted reveal similar findings to that seen in the gender-based violence literature; males with more traditional ideologies are significantly more likely to report sexual infidelity, more casual sex partners, unprotected sex, and negative attitudes toward condoms, with these

findings being demonstrated among White, Black and Hispanic young men and adolescent boys.[31,34,39,40–48] Unfortunately comparability of observed gender role associations with IPV perpetration and sexual risk behaviors among men is limited by different measures being used in different fields. Only one male gender roles measure, the Male Role Attitudes Scale (MRAS), [42] has been identified as useful in both the IPV and sexual risk literature.[42,43,45] But there has been no research assessing its association with sexual risk behaviors and IPV perpetration within the same sample.

The current study is designed to assess the associations between male gender role ideologies and sexual risk behaviors as well as IPV perpetration among a young adult sample of men recruited from an urban community health center in Boston. Sexual risk behaviors included in our analyses are unprotected sex and other sex partners in the past 3 months and forced unprotected sex in the past year, based on previous study with this sample of men in heterosexual relationships demonstrating significant associations between past year IPV perpetration and these sexual risk variables.[5]

MATERIALS AND METHODS

English and/or Spanish-speaking men, age 18–35 years, reporting sex with a female partner in the past 3 months were recruited from a large urban community health center in Boston that primarily serves lower income Hispanic and African-American clients. Based on these inclusion criteria, men entering the health center were screened at registration by a trained research staff member bilingual in Spanish–English. Men agreeing to participate in a brief, anonymous men's health survey were escorted by research staff to a private room, where individuals were screened for a second time to verify eligibility. Upon obtaining verbal consent, the self-report paper survey was administered in either English or Spanish, based on the participant's choice. Following survey completion (approximately 20 min), participants were given a $15 cash payment for their time and were informed of health center services related to HIV counseling and testing, STD testing, and social services including substance abuse and IPV. This study was approved by the Institutional Review Boards of Children's Hospital and Boston University Medical Center.

Participation

Participants were recruited from April 2004 to February 2005. Of 432 men approached, 354 were eligible; 29 eligible individuals refused participation, resulting in a participation rate of 92%. Of our eligible and willing participants ($N = 325$), 48% were at the health center for their own health care, 46% were accompanying a female partner or child to appointments for their heath care, and 6% were attending a health fair.

Based on survey reports, 18 of the 325 eligible and willing participants were excluded for not meeting age criteria ($n = 6$) or the criterion regarding sex with a female partner in the past 3 months ($n = 12$); of the remaining 307 participants, 92% ($n = 283$) reported involvement with a main female partner and penile-vaginal sex in the past 3 months with this partner; current analyses are limited to these individuals. Over one-third of these participants (36.7%) took the survey in

Spanish; the remaining 62.3% took the English survey. The Spanish version of the survey was back-translated from the English version by a professional translator unless scales and items were already available and had been tested in Spanish; the Spanish survey was then reviewed for language accuracy and approved by Spanish–English bilingual health center staff.

Survey Measures

Single items assessed demographics, including the participant's age, race/ethnicity, employment, and relationship status and length, as well as their English fluency and nativity to and length of residence in the continental U.S.

Our independent variable was Masculine Gender Role Ideologies as measured by the Male Role Attitudes Scale (MRAS).[42] This eight-item measure uses a four-point response scale ranging from 1 = "agree a lot" to 4 = "disagree a lot" to assess how much participants agree with specified masculine ideologies related to male status in society, male toughness, anti-femininity, and male hypersexuality. This measure has been used with White, Black/African American and Hispanic young men, in Spanish and in English, and does not demonstrate strong differences across racial/ethnic groups.[38,42] Cronbach alpha for our sample was 0.6; this alpha is consistent with previous studies with this measure with racially/ethnically diverse samples of young men.[42,43] Scores on this measure ranged from 1–4 in our sample, with a median score of 3; mean and standard deviation in our sample were 3.0 and 0.5, respectively; these scores are consistent with previous studies with representative samples of adolescent and young adult men.[42,49]

Our sexual risk behavior outcome variables were assessed by single items on unprotected vaginal sex and unprotected anal sex in the past 3 months with a main female partner, as well as an item on sex with other women within the past 3 months (in addition to sex with a main female partner in this timeframe). An item was also taken from the Conflict Tactics Scale-2[50] to assess forced unprotected sex in the past year; this is the only CTS-2 item related to HIV/STD and pregnancy risk.

Our IPV perpetration outcome variable was obtained from the CTS-2[50] as well. This 39-item measure assesses participant's perpetration of physical violence, sexual violence and violence resulting in victim's injury/need for medical services; it uses a seven-point response pattern to assess prevalence rates of violence against their partners ever and in the past year. (For details on items within each subscale, see Raj et al.[5]) For use in analyses, responses were summed and dichotomized as past year IPV perpetration or no past year IPV perpetration; Cronbach alpha for this scale was $\alpha = 0.93$. Note: As the forced unprotected sex item was taken from this scale, it was not included in the creation of the IPV perpetration variable.

Data Analyses

Frequencies were generated for demographics and sexual risk and IPV perpetration variables, as well as items from the MRAS. Chi-square analyses and *t*-tests were used to assess bivariate associations between the MRAS and both demographics (age, education, income, continental U.S. nativity and length of residence, Hispanic ethnicity, marital status, and relationship length) and our outcome variables (unprotected vaginal and anal sex with main female partner, other female sex partners, forced unprotected sex, and IPV perpetration ever). Adjusted logistic regression analyses were then conducted to assess associations between MRAS and each of our outcome variables. Regression analyses were adjusted to control for affects of potential confounders, including age, education (high school graduate),

income, continental U.S. nativity and length of residence, Hispanic ethnicity, marital status, and relationship length; these variables were chosen due to their associations with male sexual risk behaviors and IPV perpetration in this sample (see Raj et al.[5] for details). In an effort to create more parsimonious models for our small sample, we employed methods outlined by Rothman and Greenland[51] to determine which confounders required inclusion in final adjusted models for each outcome. Potential confounders were included based on their altering the point estimate by 10% or greater and being significant predictors of the outcome at $p < 0.20$. Adjusted odds ratios and 95% confidence intervals were used to assess significance in final models.

RESULTS

Sample Demographics

Participants were median age 24 years, 74.9% Hispanic and 21.9% Black. The majority of the sample was born in the continental U.S. (44.5%) or Latin America (53.4%). Almost one-third of participants (29.3%) was born in the Dominican Republic; 16.3% were born in Puerto Rico; 7.8% were born in Mexico, South or Central America, or Cuba. Of those not born in the continental U.S., 10.2% lived in the continental U.S. for 1 year or less, and 65.0% had lived in the U.S. for more than 5 years. More than one-third (37.5%) of participants were unemployed; 53.4% reported an income of $800 or less per month; 28.0% did not hold a high school degree or GED. (Note: high rates of unemployment and low income and education may in part be attributable to the young sample, which likely included high school students.) Approximately one in six men (15.2%) reported being married; 35.7% were living with a partner, and an additional 41.7% were dating someone. Median relationship length for the sample was 2 years; 65.0% had been in their relationship for 1 year or more. Six percent of the sample was not currently involved in a relationship with a woman although they had sex with a main female partner in the past 3 months, suggesting that they had recently ended a relationship with a main female partner.

Sexual Risk Behaviors and IPV Perpetration

Unprotected sex in the past 3 months was reported by the majority of those reporting vaginal sex (80.2%) and anal sex (79.2%) with their main female partner; 16.3% reported having forced unprotected sex in the past year. Forty-three percent of men reported sex with a non-main female partner in the past 3 months; 49.2% reported no or inconsistent condom use with these partners.

Forty percent of the sample (41.3%) reported IPV perpetration (physical, sexual, and/or injury/need for medical services due to IPV) in the past year. Past year physical IPV perpetration was reported by 27.6% of the sample; past year sexual IPV perpetration was reported by 28.3% of the sample; past year perpetration of IPV-related injury or need for medical services was reported by 13.8% of the sample.

Masculine Gender Role Ideologies

In terms of male ideologies, men in this sample most ascribed to roles around men's need for respect; 76.0% of our sample reported that they agreed a lot with the statement "It is essential for a man to get respect from others," and 83.8% agreed a lot with the statement that "A man always deserves the respect of his wife and

TABLE 1. Responses to the male roles attitudes scale (MRAS), (N = 283)

	Disagree a lot 1 (%)	Disagree a little 2 (%)	Agree a little 3 (%)	Agree a lot 4 (%)	Avg. score (Std. dev.)
It is essential for a man to get respect from others	3.2	2.0	18.0	76.0	3.7 (.7)
A man always deserves the respect of his wife and children	1.1	4.6	10.2	83.8	3.8 (.6)
I admire a man who is totally sure of himself	3.8	3.6	19.9	72.5	3.6 (.7)
A man will lose respect if he talks about his problems	46.6	20.5	18.4	13.9	2.0 (1.1)
A young man should be physically tough, even if he is not big	17.7	15.9	29.4	36.4	2.8 (1.1)
It bothers me when a man acts like a woman	19.4	12.7	20.8	46.3	2.9 (1.2)
I do not think a husband should have to do housework	47.7	23.0	13.8	14.8	2.0 (1.1)
Men are always ready for sex	9.9	21.2	25.8	42.4	3.0 (1.0)

children." Many men also expressed expectations that a man be physically tough (36.4%) and not act like a woman (46.3%). Views of men as hypersexual were also endorsed, with 42.4% indicating they strongly agree that men are always ready for sex and an additional 25.8% reporting some agreement with this statement (Table 1).

Bivariate analyses assessing associations between male gender role ideologies and key demographics demonstrated that men who had not graduated from high school or obtained a GED were significantly more likely to endorse more traditional masculine gender role ideologies than men with high school diplomas or GEDs ($p = 0.04$); no other demographics, including race/ethnicity and acculturation variables, were significantly associated with the MRAS. Bivariate analyses assessing associations between male gender role ideologies and sexual risk and IPV perpetration outcome variables demonstrated that the MRAS was significantly associated with past year IPV perpetration ($p = 0.05$), and a trend was seen in the association between the MRAS and unprotected vaginal sex with a main female partner ($p = 0.1$).

Associations between MRAS and Sexual Risk Behaviors and IPV perpetration

Consistent with bivariate associations between male gender role ideologies and sexual and IPV perpetration risk, adjusted regression analyses demonstrated that men reporting more traditional ideologies were significantly more likely to report unprotected vaginal sex with their main partner (OR_{adj} = 2.3, 95% CI = 1.2–4.6) and past year IPV perpetration (OR_{adj} = 2.1, 95% CI = 1.2–3.6). Such ideologies were not associated with unprotected anal sex, forced unprotected sex or multiple sex partners (see Table 2).

TABLE 2. Logistic regression analyses adjusted for demographics to assess associations between male gender role ideologies and sexual risk behaviors and partner violence perpetration, (N = 283)

Sexual risk behaviors or pregnancy involvement	OR_{adj} (95% CI)
Unprotected vaginal sex, past 3 months	2.3 (1.2–4.6)[2]
Unprotected anal sex, past 3 months	1.1 (.4–3.4)[3]
Forced sex without a condom, past year	1.4 (.7–2.8)[4]
Other female sex partners, past 3 months	1.2 (.7–1.9)[1]
IPV perpetration, past year	1.8 (1.1–2.9)[5]

[1]Analysis adjusted for relationship length.
[2]Analysis adjusted for age, relationship length.
[3]Analysis adjusted for age, Hispanic ethnicity.
[4]Analysis adjusted for age, Hispanic ethnicity, non-English speaking.
[5]Analysis adjusted for age, relationship length, non-English speaking.

DISCUSSION

The current study indicates that young men with more traditional masculine gender role ideologies are more likely to report recent unprotected vaginal sex and IPV perpetration within the context of their heterosexual relationships; however, findings demonstrate no significant associations between these ideologies and having other female sex partners, forcing unprotected sex, or engaging in unprotected anal sex. Thus, while these findings lend some support to traditional masculine gender role ideologies being the linchpin explaining previous research findings linking non-condom use and IPV perpetration in men,[4,5] lack of consistent findings in our sample suggest further research exploring this issue is needed.

Lack of consistent findings from our sample may be attributable, in part, to the diversity and simultaneously limited constructs of masculinity used in the MRAS measure. As described previously, the MRAS was designed to measure male status in society (i.e., men's need for respect), male toughness (i.e., expectations men are physically tough and stoic), anti-femininity (i.e., intolerance for men exhibiting traditionally female attributes), and male hypersexuality (i.e., the view that men want sex all the time).[42] The diversity of these attributes may be the cause of the measure's low internal reliability, observed in our own study as well as others.[42,43,45] However, additionally, a number of traditional masculine ideologies seen to be related to sexual assault, multiple partners, and partner violence are not included in the MRAS. Some of these specific ideologies include: 1. *traditional gendered sexual and relationship scripts*, including male control, sexual entitlement, coercive "seduction," and sexual dominance in relationships; 2. *traditional male behavioral norms*, including male aggression, desire for risk/danger, and emotional detachment toward women; 3. *sexually conservative and negative attitudes toward women*, including views that women should be sexually passive and that a woman's victimization from gender-based violence occurs as a consequence of her aggression, sexual teasing of men, or promiscuity; and 4) *adversarial heterosexual relationship norms* (e.g., women lie to get what they want from men).[29–37,48,52–55] More in-depth qualitative and quantitative research is needed to better understand the diversity of masculine ideologies and how these relate to young men's IPV

perpetration and sexual risk behaviors, as well as how these ideologies may link men's IPV perpetration and sexual risk.

While these findings offer important insight into male gender role ideologies and risk behaviors among young urban men, they must be considered in the context of certain study limitations. Generalizability of results are limited due to the use of a single community health center serving predominantly lower income Hispanic and Black men in the urban Northeast. Further, our findings may not even be generalizable to other community health centers in the Northeast. While our health center is typical of other community health centers in Boston in terms of is location within a lower income area and its predominantly racial/ethnic minority and lower income client population, it reaches a larger segment of immigrants and Hispanics than that seen at some of the other health centers in Boston. Our sample was largely Latino and, therefore, may reflect cultural expected sexual gender roles within the Latino community. Additionally, as our study included men seeking care at the health center as well as those accompanying others, findings cannot be generalized to those seeking care even within our collaborating health center. Unfortunately, as reasons for attending the health center were assessed at screening and not linked to survey data, we were unable to compare those attending the health center for their own care as opposed to the care of someone else, to determine whether these groups were comparable.

In addition to limitations related to generalizability, there are a number of other study limitations. This research was cross-sectional, so causality cannot be inferred from findings. Reliance on self-report makes these data subject to social desirability and recall biases; however, these biases would likely result in the under-reporting rather than over-reporting of sensitive issues such as perpetration of partner violence, unprotected sex and sexual infidelity. Due to the nature of the way questions were asked, we were unable to assess whether the reported sexual risk behavior occurred within the context of the abusive relationship. Longitudinal work, as well as studies with relationship-specific questions, will allow for greater elucidation of the nature of the relationships of male gender role ideology with sexual risk and IPV both independently and in conjunction and across greater time periods.

Conclusion and Implications

Young urban men endorsing more traditional masculine gender role ideologies appear to be more likely than those with more egalitarian gender role beliefs to engage in unprotected vaginal sex and IPV within heterosexual relationships. These findings may lend support to the theory that traditional male gender role ideologies constitute a shared risk source for IPV and sexual risk, but the lack of significant findings related to male gender role ideologies and multiple sex partners, as well as forced unprotected sex, highlight the need for further clarification on what different types of ideologies affect diverse sexual risk behaviors in this population. Further research is needed to better clarify the inconsistencies found in this exploratory work. Nonetheless, findings suggest that interventions to promote less traditional masculine gender role ideologies among young men may be helpful in addressing sexual risk and IPV perpetration for this population.

ACKNOWLEDGEMENTS

This project was funded through a grant from the Massachusetts Department of Public Health. We would like to thank all of the staff and providers at Martha Eliot Health Center who helped facilitate study recruitment for this project; we would

particularly like to thank David Holder, Catherine MacAuley, and Ana Ortiz for their support and guidance in implementing the study at Martha Eliot Health Center.

REFERENCES

1. Center for Disease Prevention. Leading causes of death by age group, all races, males—U.S. 2002. *Office of the director. Men's health.* 2005. Available at: http://www.cdc.gov/men/lcod/02all.pdf.
2. Center for Disease Center. Cases of HIV infection and aids in the United States, 2003. *HIV/AIDS surveillance report.* 2005c;15.
3. Dunkle KL, Jewkes RK, Brown HC, et al. Gender-based violence, relationship power, and risk of HIV infection in women attending antenatal clinics in South Africa. *Lancet.* 2004;363(9419):1415–1421.
4. El-Bassel N, Fontdevila J, Gilbert L, et al. HIV risks of men in methadone maintenance programs who abuse their intimate partners: a forgotten issue. *J Subst Abuse Treat.* 2001;12(1–2):29–43.
5. Raj A, Santana MC, La Marche AM, et al. Perpetration of partner violence associated with sexual risk behaviors among young adult men. In Press: *Am J Pub Health.*
6. Sabo D. Men's health studies: origins and trends. *J Am Coll Health Assoc.* 2000; 49(3):133–142.
7. Williams DR. The health of men: structured inequalities and opportunities. *Am J Public Health Nations Health.* 2003;93(5):724–731.
8. Centers for Disease Control and Prevention. Life-time and annual incidence of intimate partner violence and resulting injuries—Georgia, 1995. *MMWR CDC Surveill Summ.* 1998;47(40):849–853.
9. Centers for Disease Control and Prevention. Intimate partner violence among men and women—South Carolina, 1998. *MMWR CDC Surveill Summ.* 2000a;49(30): 691–694.
10. Centers for Disease Control and Prevention. Prevalence of intimate partner violence and injuries—Washington, 1998. *MMWR CDC Surveill Summ.* 2000b;49:849–853.
11. Plichta SB. Violence, health and use of health services. In: Falik MM, Collins KS, eds.*Women's Health and Care Seeking Behavior: The Commonwealth Fund Survey.* Baltimore, Maryland: Johns Hopkins University; 1996.
12. Schafer J, Caetano R, Cook CL. Rates of intimate partner violence in the United States. *Am J Public Health Nations Health.* 1998;88:1702–1704.
13. Straus MA, Gelles RJ. Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys. *J Marriage Fam.* 1986;48:465–479.
14. Tjaden P, Thoennes N. Full report of the prevalence, incidence, and consequences of violence against women: findings from the national violence against women survey. Washington: National Institute of Justice; 2000. Report NCJ 183781.
15. Amaro H, Raj A. Theoretical and measurement issues in the study of women's relational power in HIV risk reduction. *Sex Roles.* 2000;42(7/8):723–749.
16. Heise L, Ellsberg M. Ending violence against women. population reports. Volume Xxvii, Number 4. *Issues in World Health.* December, 1999. Available at: http://www.infoforhealth.org/pr/l11edsum.shtml.
17. Schmuel E, Schenker JG. Violence against women: the physician's role. *Eur J Obstet Gynecol Reprod Biol.* 1998;80(2):239–245.
18. Bauer HM, Gibson P, Hernandez M, et al. Intimate partner violence and high-risk sexual behaviors among female patients with sexually transmitted diseases. *Sex Transm Dis.* 2002;29(7):411–416.
19. Decker MK, Silverman JG, Raj A. Dating violence and STD/HIV testing and diagnosis among adolescent females. *Pediatrics.* 2005;116:e272–e276.
20. He H, McCoy HV, Stevens SJ, et al. Violence and HIV sexual risk behaviors among female sex partners of male drug users. *Women Health.* 1998;27(1–2):161–175.

21. Rickert VI, Wiemann CM, Harrykissoon SD, et al. The relationship among demographics, reproductive characteristics, and intimate partner violence. *Am J Obstet Gynecol.* 2002;187(4):1002–1007.
22. Roberts TA, Auinger P, Klein JD. Intimate partner abuse and the reproductive health of sexually active female adolescents. *J Adolesc Health.* 2005;36(5):380–385.
23. Silverman JG, Raj A, Mucci LA, et al. Dating violence against adolescent girls and associated substance use, unhealthy weight control, sexual risk behavior, pregnancy, and suicidality. *J Am Med Assoc.* 2001;286(5):572–579.
24. Silverman JG, Raj A, Clements K. Dating violence and sexual risk in a representative sample of high school students. *Pediatrics.* 2004:114(2):e220–e225.
25. Wingood GM, DiClemente RJ. Consequences of having a physically abusive partner on the condom use and sexual negotiation practices of young adult African-American women. *Am J Public Health Nations Health.* 1997;87(6):1016–1018.
26. Wingood GM, DiClemente RJ. Partner influences and gender-related factors associated with noncondom use among young adult African–American women. *J Community Psychol.* 1998;26:29–53.
27. Hathaway JE, Willis G, Zimmer B, et al. Impact of partner abuse on women's reproductive lives. *J Am Med Womens Assoc.* 2005;60(1):42–45.
28. Raj A, Liu R, McCleary-Sills J, et al. South Asian victims of intimate partner violence more likely than non-victims to report sexual health concerns. *J Immigr Health.* 2005;7(2):85–91.
29. Andersen B, Cyranowski JM, Espindle, D. Men's sexual self-schema. *J Pers Soc Psychol.* 1999;76:645–661.
30. Anderson VN, Simpson-Taylor D, Hermann DJ. Gender, age and rape-supportive rules. *Sex Roles.* 2004;50(1–2):77–90.
31. Byers ES. How well does the traditional sexual script explain sexual coercion? Review of a program of research. *J Psychol Human Sex.* 1996;8(1):7.
32. Chiroro P, Bohner G, Viki GT, et al. Rape myth acceptance and rape proclivity: expected dominance versus expected arousal as mediators in acquaintance-rape situations. *J Interpers Violence.* 2004;19(4):427–441.
33. Greendlinger V, Byrne D. Coercive sexual fantasies of college men as predictors of self-reported likelihood to rape and overt sexual aggression. *J Sex Res.* 1987;23(1):1–11.
34. Krahe B. Sexual scripts and sexual aggression. In: Eckes T, Trautner HM, eds. *The Developmental Social Psychology of Gender.* Mahwah, New Jersey: Erlbaum; 2000.
35. Malamuth NM, Scokloski RJ, Koss MP, et al. Characteristics of aggressors against women: testing a model using a national sample of college students. *J Consult Clin Psychol.* 1991;59:670–681.
36. Murnen SK, Wright C, Kaluzny G. If "Boys Will Be Boys," then girls will be victims? A meta-analytic review of the research that relates masculine ideology to sexual aggression. *Sex Roles.* 2002;17:359–375.
37. Ryan K. Further evidence for a cognitive component of rape. *Aggress Violent Behav.* 2004;9:579–604.
38. Doss BD, Hopkins JR. The multicultural masculinity ideology scale: validation from three cultural perspectives. *Sex Roles.* 1998;38(9–10):719–741.
39. Bowleg L, Lucas KJ, Tschann JM. "The Ball Was Always in His Court": an exploratory analysis of relationship scripts, sexual scripts and condom use among African American women. *Psychol Women Q.* 2004;28(1):70–82.
40. DeLamater J. Females, males and sexuality. In: Kelley K, ed. *Gender Differences in Sexual Scenarios.* Albany, New York: State University of New York; 1987:127–139.
41. Noar SM, Morokoff PJ. The relationship between masculinity ideology, condom attitudes, and condom use stage of change: a structural equation modeling approach. *Int J Mens Health.* 2002;1(1):43–58.
42. Pleck JH, Sonenstein FL, Ku LC. Masculinity ideology: its impact on adolescent males' heterosexual relationships. *J Soc Issues.* 1993;49(3):11–29.

43. Pleck JH, O'Donnell LN. Gender attitudes and health risk behaviors in African-American and Latino early adolescents. *Matern Child Health J.* 2001;5:265–272.
44. Seal DW, Ehrhardt AA. HIV-prevention-related sexual health promotion for heterosexual men in the United States: pitfalls and recommendations. *Arch Sex Behav.* 2004;33(3): 211–222.
45. Shearer CL, Hosterman SJ, Gillen MM, et al. Are traditional gender role attitudes associated with risky sexual behavior and condom-related beliefs. *Sex Roles.* 2005;52(5/6):311–324.
46. Schoeneberger M, Logan T, Leukefeld, C. Gender roles, HIV risk behaviors, and perceptions of using female condoms among college students. *Popul Res Policy Rev.* 1999;18(1–2):119–136.
47. Spencer M, Fegley S, Harpalani V, et al. Understanding hypermasculinity in context: a theory-driven analysis of urban adolescent males' coping responses. *Res Hum Dev.* 2004;1(4):229–257.
48. Sugarman DB, Frankel SL. Patriarchal ideology and wife assault: a meta-analytic review. *J Fam Violence.* 1996;11:13–40.
49. Forste R, Haas DW. The transition of adolescent males to first intercourse: anticipated or delayed. *Perspect Sex & Repro Health.* 2001;34(4), 2002 http://www.guttmacher.org/pubs/journals/3418402.html.
50. Straus MA, Hamby SL, Boney-McCoy S, et al. The revised conflict tactics scales (Cts2). *J Fam Issues.* 1996:17(3):283–316.
51. Rothman KJ, Greenland S. *Modern Epidemiology.* 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998.
52. Hill MS, Fischer AR. Does entitlement mediate the link between masculinity and rape-related variables? *J Consult Psychol.* 2001;48:39–50.
53. Malamuth NM, Brown LM. Sexually aggressive men's perceptions of women's communications: testing three explanations. *J Pers Soc Psychol.* 1994;67:699–712.
54. Mosher DL, Anderson RD. Macho personality, sexual aggression and reactions to guided imagery of realistic rape. *J Res Pers.* 1986;20:77–94.
55. Mosher DL, Sirkin M. Measuring a macho personality constellation. *J Res Pers.* 1984; 18:153–163.

Strategic Memory in Adults with Anorexia Nervosa: Are There Similarities to Obsessive Compulsive Spectrum Disorders?

Bonnie J. Sherman, PhD[1]
Cary R. Savage, PhD[2]
Kamryn T. Eddy, MA[3]
Mark A. Blais, PsyD[4]
Thilo Deckersbach, PhD[4]
Safia C. Jackson, BS[4]
Debra L. Franko, PhD[4,5]
Scott L. Rauch, MD[4]
David B. Herzog, MD[4]*

ABSTRACT

Objective: There is growing interest in the relationship between anorexia nervosa (AN) and obsessive-compulsive (OC) spectrum disorders (e.g., OCD, body dysmorphic disorder [BDD]). Previous neuropsychological investigations of OC spectrum disorders have identified problems with the efficient use of strategy on complex measures of learning and memory. This study evaluated nonverbal strategic memory in AN outpatients using an approach previously applied to OC spectrum disorders.

Method: Eighteen patients with AN and 19 healthy control participants completed the Rey–Osterrieth Complex Figure Test (RCFT), a widely used measure of nonverbal strategic planning, learning, and memory.

Results: Individuals with AN differed significantly from healthy controls in the organizational strategies used to copy the RCFT figure, and they recalled significantly less information on both immediate and delayed testing. Multiple regression analyses indicated that group differences in learning were mediated by copy organizational strategies.

Conclusion: These results are identical to study findings in OCD and BDD, indicating important shared neuropsychological features among AN and these OC spectrum disorders. As in OCD and BDD, the essential cognitive deficit in AN was impaired use of organizational strategies, which may inform our understanding of the pathophysiology of AN and potentially offer treatment implications.

Keywords: anorexia nervosa, obsessive-compulsive disorder, neuropsycological testing; memory; learning; Rey–Osterrieth Complex Figure Test (RCFT); strategic planning; cognition; organization

(Int J Eat Disord 2006; 39:468–476)

Introduction

Anorexia nervosa (AN) is characterized by food-related and weight-related obsessions and compulsive dietary restriction. These clinical features have prompted links between AN and obsessive-compulsive (OC) spectrum disorders, including obsessive compulsive disorder (OCD), body dysmorphic disorder (BDD), and OC personality disorder (OCPD), for example. These links are supported by three lines of evidence. First, clinical investigations have indicated a high prevalence of OC spectrum disorders in patient groups with AN[1] as well as increased rates of AN in individuals with OC spectrum disorders.[2,3] Similarly, traits associated with OC spectrum disorders, such as perfectionism, are similarly elevated in individuals with AN and OCD, and persist in patients with AN postrecovery.[4–7] These findings lead some investigators to hypothesize that obsessional traits exist premorbidly and contribute to the pathogenesis of eating disorders.[8] Second, family studies, which indicate an increased risk for OC spectrum disorders in the families of patients with AN independent of OC symptoms in the AN probands, provide further support for the links between AN and OC spectrum disorders and suggest a shared diathesis model between the disorders.[9,10]

Imaging and neuropsychological studies of AN and OC spectrum disorders have also indicated simi-

Accepted 20 February 2006
Supported in part by grants from the Rubenstein Foundation, Eli Lilly and Company, Harvard Medical School Eating Disorders Center, MGH David Judah Research Fund.
*Correspondence to: David B. Herzog, MD, Department of Psychiatry, Massachusetts General Hospital, YAW6900, 55 Fruit Street, Boston, MA 02114. E-mail: dherzog@partners.org
[1] Center for Excellence in Women's Health, Boston University Medical Center, Boston, MA
[2] Department of Psychiatry and Behavioral Sciences, University of Kansas Medical Center, Kansas City, KS
[3] Department of Psychology, Boston University, Boston, MA
[4] Department of Psychiatry, Massachusetts General Hospital, Boston, MA
[5] Department of Counseling and Applied Educational Psychology, Northeastern University, Boston, MA
Published online 19 May 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/eat.20300

larities both in neural systems implicated in these disorders and in associated cognitive deficits. For example, functional imaging studies in AN indicate increased activity relative to healthy controls in the caudate nuclei[11–15] and inferior prefrontal territories corresponding to the orbitofrontal cortex.[11,13] Similarly, functional imaging studies in OCD have demonstrated abnormally increased activity in orbitofrontal cortex, anterior cingulate cortex, and caudate nuclei in patients during resting states and during induced states of symptom provocation.[16,17] Further, neuropsychological studies of AN and OC spectrum disorders have indicated similar deficits in executive functioning[18–23] and memory,[18,19,24–28] in particular. Although there is some evidence to suggest that lower weight is associated with poorer cognitive functioning in AN,[20] many studies suggest that these executive functioning and memory deficits in AN persist postrecovery,[18,24,29] which argues that they are not secondary to starvation.

Investigators have explored the relation between executive functioning deficits and memory problems in patients with OC spectrum disorders, including OCD and BDD,[27,30–32] using a measure of nonverbal strategic memory, the Rey–Osterrieth Complex Figure Test (RCFT).[33] Owing to the complexity of the RCFT stimulus, subjects must carefully plan and organize their drawings; these executive organizational strategies invoked when copying the figure have a significant impact on the ability to later recall information about the figure. Using the RCFT, investigators[30–32,34] have found that OCD and BDD patients used less efficient organizational strategies when copying the figure and recalled less on both immediate and delayed testing. Recall was mediated by copy organizational strategy, suggesting that the less efficient planning and organizational strategies employed had negative effects on memory. More recently, these findings have been replicated by other investigators[35,36] and shown to correlate with reduced gray matter volume in a left anterior region of the orbitofrontal cortex of OCD patients.[37]

These findings may help explain the obsessive and compulsive thought/behavior patterns in individuals with OC spectrum disorders. Specifically, individuals with these disorders demonstrate impaired organizational strategies (e.g., they focus on extraneous details without taking in the big picture), which may contribute to the experience of repeated doubt and checking behaviors. For example, an individual with OCD may feel driven to compulsively check whether he has locked the doors in part because he is unable to verify to his satisfaction that he actually did so, which may be related to a deficit in organizing the event (i.e., the initial door locking).[38] It is possible that the OC thought/behavior patterns in individuals with AN (e.g., body image disturbance, body checking, calorie counting) are secondary to similar strategic organization deficits and subsequent memory problems.

To our knowledge, strategic memory has not been directly explored in AN, although previous neuropsychological findings have demonstrated that patients with AN have executive functioning deficits and impaired accuracy on copy and recall of the RCFT.[18,19,25] Thus, the current study was designed to extend previous strategic memory findings in OC spectrum disorders to a group of patients with AN and healthy control subjects, carefully matched for gender, handedness, age, education, and estimated verbal and nonverbal intelligence. Given the identified relationship between AN and OC spectrum disorders, we predicted that (1) patients with AN would show impaired nonverbal memory as measured by the RCFT, and (2) copy organizational strategies would mediate group differences in recall accuracy. Such findings would strengthen the proposed connections between AN and OC spectrum disorders by establishing cognitive similarities among these patient groups using a well-characterized neuropsychological measure of strategic memory.

Method

Participants

Study participants were 18 female outpatients meeting DSM-IV criteria for AN[39] and 19 matched healthy female control participants. Patients were recruited through the Massachusetts General Hospital (MGH) Eating Disorders Unit, the MGH Neuroendocrine Unit, and local advertisement; control participants were recruited through local advertisement. Body Mass Index scores (BMI) ranged from 14.6 to 18.1 in the AN group (*M*, 16.68 ± 1.1). The diagnosis of AN subjects and the nonpsychiatric status of healthy control participants were determined by the Structured Clinical Interview for DSM-IV (SCID),[40] conducted by trained interviewers. Healthy controls had no current or history of any Axis I disorders. Entry criteria for patients with AN included no history of psychotic disorder, neurologic disorder, head injury, current or past substance dependence, or current substance abuse; entry criteria for healthy control subjects included the added requirement of no history of prescribed psychotropic medication or Axis I disorder. Among the AN subjects, 9 had at least one co-morbid Axis I diagnosis, although AN was considered to be the primary diagnosis (reason for seeking treatment) in all cases. Co-morbid conditions

TABLE 1. Demographic, general cognitive, and clinical characteristics of the sample[a]

	AN Patients ($N = 18$)	Healthy Control Subjects ($N = 19$)	t ($n = 35$)	p
Handedness	100% right	100% Right		
Gender	100% female	100% Female		
Age (years)	25.56 (5.8)	25.68 (5.3)	.070	.95
Education (years)	15.72 (2.3)	16.58 (1.3)	1.94	.17
Vocabulary*	12.33 (2.3)	12.84 (1.7)	.59	.45
Matrix Reasoning*	12.89 (1.6)	13.53 (1.4)	1.63	.21
Digit span*	11.00 (2.8)	11.47 (2.0)	.36	.55
Duration of Illness (years)	5.67 (5.5)	N/A		
Binge/purge symptoms	3/18	N/A		
BMI	16.68 (1.1)	22.22 (1.8)	122.41	<.0001
BDI	16.94 (12.1)	2.10 (2.5)	27.35	<.0001
EDI Total	63.50 (38.1)	14.84 (8.6)	29.45	<.0001
EDI Drive for Thinness	11.56 (7.4)	0.47 (0.70)	42.30	<.0001
EDI Bulimia	3.10 (4.4)	0.05 (0.23)	9.32	.004
EDI Body Dissatisfaction	14.22 (8.8)	5.63 (4.8)	13.84	.0007

Note: BMI-body mass index [weight in kg/(height in m^2)]; BDI-Beck Depression Inventory; EDI-Eating Disorder Inventory.
[a] Standard deviation in parentheses.
* WASI scores represent scaled scores.

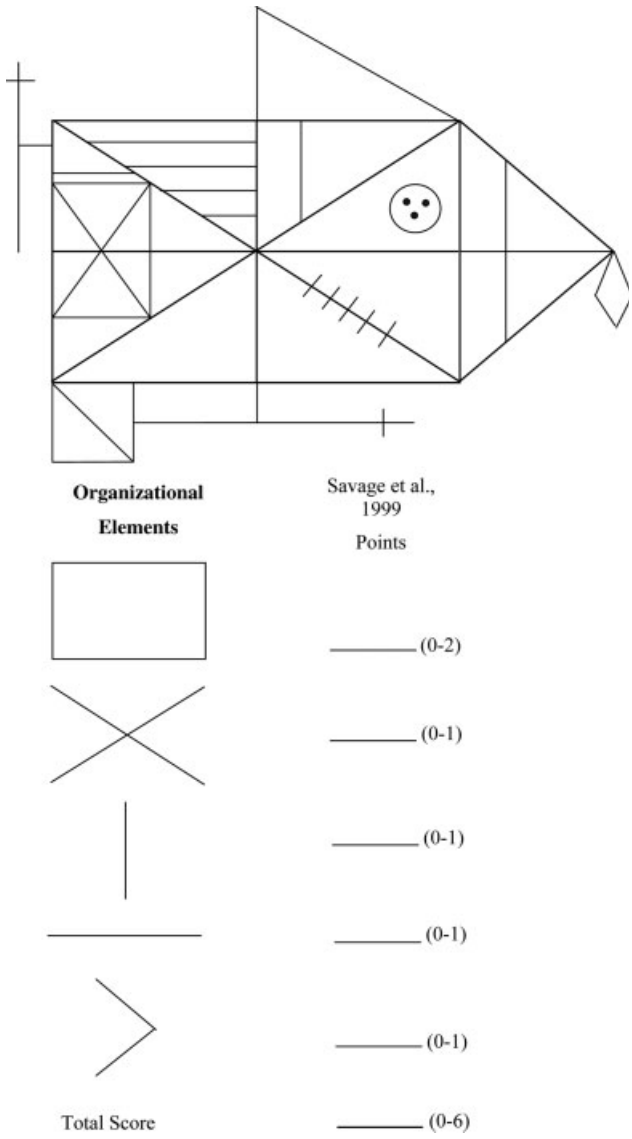
included: major depression ($n = 6$), dysthymia ($n = 3$), and generalized anxiety disorder ($n = 5$). None of the participants in this study had comorbid OCD or BDD. In addition, 11 of the AN subjects were taking psychotropic medication at the time of testing, including antidepressants ($n = 10$), anxiolytics ($n = 4$), and a neuroleptic ($n = 1$). After complete description of the study, written informed consent was obtained from all participants.

Participants completed the Eating Disorder Inventory (EDI)[41] and Beck Depression Inventory (BDI).[42] Verbal and nonverbal intellectual functioning were estimated with the Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence (WASI).[43] Subjects also completed the Digit Span subtest of the Wechsler Memory Scale-III (WMS-III)[44] as a general measure of verbal attention and working memory capacity. Table 1 summarizes the demographic, clinical, and cognitive characteristics of the sample.

Procedure and Materials

Neuropsychological tests were administered and scored by trained examiners. The examiners were trained over several days by a neuropsychologist (TD) with extensive experience in using and interpreting the RCFT. During training, the neuropsychologist observed the examiners as they administered and scored the RCFT on mock subjects. Once competency was achieved, the examiners began administering the RCFT to study participants. Scoring was performed blinded to group membership.

FIGURE 1. Original Savage et al.[30] system for scoring the organizational approach on the Rey–Osterrieth Complex Figure. Five core configurational elements are identified: The large rectangle (2 points), the two diagonals (1 point), the vertical midline (1 point), the horizontal midline (1 point), and the vertex of the triangle to the right (1 point). The scores for each element are summed, resulting in a scale with a range of scores from 0 to 6.



Rey–Osterrieth Complex Figure Test. The RCFT[33] provides measures of construction ability (Copy) and nonverbal immediate and delayed free recall (Immediate Recall, Delayed Recall). In addition, the RCFT can be used to examine organizational strategies used during the copy condition. The RCFT figure was presented to all participants on a sheet of paper presented in horizontal orientation. Participants were instructed to copy the figure (see Figure 1). Once completed, the figure was immediately removed and the participant was asked to reconstruct it from memory. The participant was again asked

TABLE 2. Summary of RCFT means and significance tests in AN patients and healthy control subjects[a]

	AN Patients ($N = 18$) Mean (*SD*)	Healthy Controls ($N = 19$) Mean (*SD*)	t ($n = 35$)	2-tailed p-value	d
Accuracy scores					
Copy	67.72 (4.7)	69.95 (1.8)	−1.93	.06	−.63
Immediate recall	36.94 (14.7)	50.95 (8.6)	−3.56	.001	−1.16
Delayed recall	35.44 (13.5)	50.79 (7.8)	−4.26	.0001	−1.39
Percentage Recall					
Copy to immediate recall	54.8 (22.0)	72.79 (11.8)	−3.12	.003	−1.02
Immediate recall to delayed recall	98.46 (19.1)	100.59 (12.0)	−.41	.69	−.13
Copy organization	3.00 (2.2)	4.63 (1.5)	−2.65	.01	−.87

[a] Standard deviation in parentheses.

to reconstruct the figure after a 30-minute delay. Participants were not given specific time limits in which to complete the figure, rather they were permitted to draw until they indicated that they were finished. For each of the three RCFT trials, participants were provided colored pencils that were changed approximately every 15 seconds to allow subsequent scoring of organization based on the order in which the component parts were constructed. If the subjects were completing a major feature at the end of 15 seconds they were allowed to complete that feature before the pencil was switched.

Two scores were calculated in this study. First, Construction Accuracy was calculated for copy, immediate recall (without intervening distraction), and delayed recall (30-minute) conditions using a scoring system developed by Denman,[45] and described previously in detail.[30,32,34,46] There are three identified criteria for each segment, each assigned one point, resulting in a range of scores from 0 to 72. The focus of scoring is on the participant's ability to draw and recall visual details, rather than on skill in drawing or how the components are organized. Second, Organizational Strategy was evaluated with a quantitative method developed and described in detail by Savage et al.[30] and Deckersbach et al.[47] To score organization, five configural elements of the figure were identified (base rectangle, two diagonals, vertical midline, horizontal midline, vertex of the triangle on right), and the subject received points for constructing each as an unfragmented unit. Previous estimates of reliability for this scoring system have exceeded .90.[30,32,47]

Statistical Analysis

RCFT means were first evaluated by *t*-tests. We predicted that, compared with controls, the AN group would use less systematic organizational strategies and learn significantly less during encoding.

Following these analyses, we tested the hypothesis that differences in free recall would be mediated by strategies used during the RCFT copy condition. The mediation hypotheses were tested via multiple regression, in a three-variable path model,[48] using procedures identical to those used previously.[30,32,34,47] In this analytical approach, the causal ordering is established a priori and a multiple regression equation is computed conditional on the model, with group (AN = 1, Control = 0) as the independent variable, RCFT immediate percentage recall as the dependent variable, and RCFT copy organization as the hypothesized mediator. We predicted that the mediated model would be supported.

Results

RCFT

Mean RCFT accuracy, percentage recall rates, and copy organization scores are presented in **Table 2**.

RCFT accuracy scores were evaluated via two factor (Group × Condition) mixed-model ANOVA, with repeated measures on Condition (Copy, Immediate Recall, Delayed Recall). Analyses indicated a significant main effect for Group, $F(1,35) = 17.19$, $p = .0002$, as well as a significant Group × Condition interaction, $F(2,70) = 9.96$, $p = .0002$. The Group × Condition interaction is illustrated in **Figure 2**.

Analyses indicated that patients with AN and controls differed in copy performance only at a trend level, $t(35) = 1.91$, $p = .06$, but AN subjects recalled significantly less than controls on immediate recall, $t(35) = 3.55$, $p = .001$ and delayed recall, $t(35) = 4.27$, $p = .0001$. In a similar manner, we evaluated mean percentage recall rates (immediate recall × 100/copy; and delayed recall × 100/immediate recall) using two factor (Group × percentage recall between conditions) mixed model ANOVA with repeated measures on percentage recall between conditions (Copy to Immediate Recall, and Immediate to Delayed Recall). These analyses indicated a significant main effect for Group, but not a significant interaction between Group and percentage recall between conditions. We conducted post hoc *t* tests and found that the AN group recalled significantly less from copy than the

FIGURE 2. Group × condition interaction for RCFT accuracy scores. Error bars represent standard error of the mean.

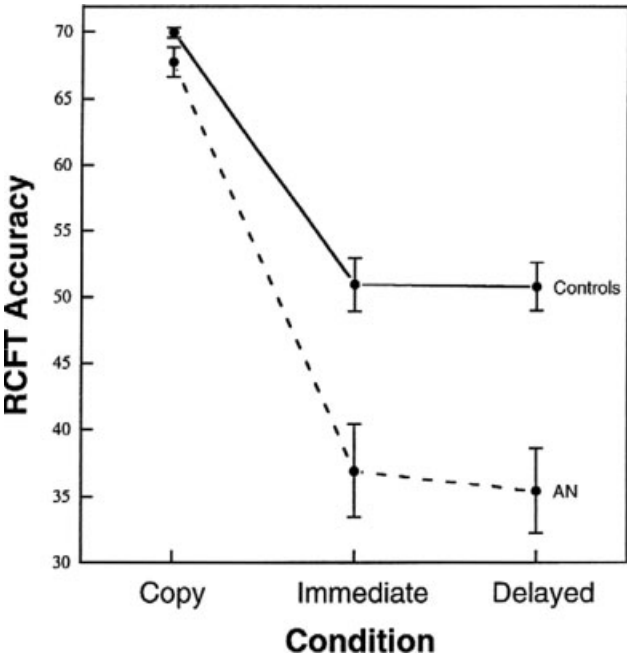
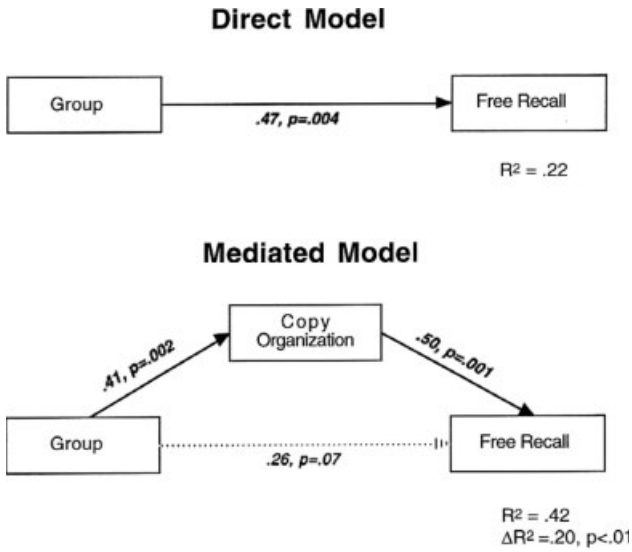


FIGURE 3. Two alternate path models explaining group differences in free recall accuracy (immediate percentage recall) on the RCFT. The top portion represents the direct model, in which the effects of group (AN, Control) are expressed directly in differences in free recall accuracy. The bottom portion illustrates a mediated model, in which the effects of group on free recall accuracy are expressed indirectly, via the influence of copy organizational strategies. Regression analyses supported the mediated model.



control group, $F(1,35) = 9.74$, $p = .004$, but showed normal retention of information between immediate and delayed recall conditions, $F(1,35) = 0.17$, $p = .69$. Finally, patients with AN obtained lower copy organization scores in comparison with controls (**Table 2**), $F(1,35) = 7.04$, $p = .01$, $d = -0.63$.

Path Models

Our a priori hypothesis was that nonverbal memory performance in AN would be mediated by the way organizational strategies were used during learning. We found predicted group differences in learning and strategic processing on the RCFT using ANOVA. Based on these results and a priori hypotheses, we selected Immediate Percentage Recall as the dependent variable in the path analyses testing mediation. This measure was considered the most representative of group differences in free recall on the RCFT, since copy was essentially normal and immediate recall was impaired with no additional loss of information at delayed recall (see **Figure 2**). We also chose this variable because it was the one used in previous OCD studies.[30,34]

Path modeling results for the RCFT are presented following the steps outlined by Baron and Kenny.[48] In the simple regression analyses, Group had a direct effect on the hypothesized mediator, copy organization ($\beta = .41$, $p = .01$), and on the dependent variable percentage recall from copy ($\beta = .47$, $p = .0003$). In the multiple regression equation, copy organization continued to have a strong direct effect on percentage recall ($\beta = .50$, $p < .001$), but the direct effect of Group was reduced in absolute size and dropped to a statistical trend ($\beta = .26$, $p = .07$). When comparing the direct and mediated models, the mediated model provided significantly greater explanatory power ($\Delta R^2 = .20$, $p < .01$). The size and significance of the indirect effect of group through strategic processing can be calculated by multiplying the two standard coefficients; this indirect path is statistically significant ($.41 \times .50 = .21$, $p < .01$).[48] Thus, group differences in RCFT recall are expressed primarily along an indirect path via the effects of group on organizational strategies. **Figure 3** illustrates these results for the RCFT as path diagrams representing "direct" and "mediated" models, with standard coefficients (β) and significance levels provided for each link.

The validity of the mediated model is dependent on the variables selected and their causal ordering. However, it should be noted that there is only one logical causal ordering for this model (Group → Copy Organization → Recall): RCFT performance cannot cause AN (the diagnosis preceded participation in this study) and RCFT immediate recall follows copy organization temporally (poor recall can-

not cause poor copy organization). Nonetheless, it is possible that other unmeasured variables might have an impact on our model. One reasonable variable might be copy accuracy, i.e., information recalled might conceivably be influenced by how accurately the figure was originally constructed. We, therefore, evaluated potential effects for copy accuracy. Copy accuracy was not correlated with immediate percentage recall ($\beta = -.17$, $p = .66$), nor did it add significant explanatory power to the mediated model from **Figure 1** ($\Delta R^2 = .03$, ns). Thus, our results indicate that learning and memory impairment on the RCFT was mediated by impaired organizational strategies in the copy condition, but not by copy accuracy.

Correlations With Clinical Measures

We performed exploratory correlation analyses in the AN group between clinical measures and measures from the RCFT on which Patients with AN were significantly impaired. BMI indices were not significantly related to any of the RCFT scores (all $p > .45$), nor were BDI scores (all $p > .20$). EDI scores were correlated with RCFT copy performance, Total EDI: $r_{16} = -.46$, $p = .05$. Analyses of EDI subscales indicated that only scores on the first three subscales (Drive for Thinness: $r = -0.527$, $p = 0.025$, $n = 18$, Bulimia: $r = -0.766$, $p < 0.001$, $n = 18$, Body Dissatisfaction: $r = -0.548$, $p = 0.018$, $n = 18$) were correlated with RCFT copy performance.

Effects of Comorbidity and Medication

Half the AN sample presented with Axis I comorbidity ($n = 9$) and most were taking psychotropic medication ($n = 11$). To evaluate the potential effects of comorbidity and medication, we performed two additional sets of regression analyses excluding these patients (i.e., those with comorbidity or those taking psychotropic medications) in order to verify the models. These analyses showed no changes in the patterns of results, although the reduced sample sizes limited power and thus rendered some of the findings nonsignificant. Although the effects of comorbidity and medication cannot be ruled out absolutely, results indicate that current findings in the AN group are not likely the result of these factors.

Conclusion

Patients with AN in this study were significantly impaired in comparison with healthy control participants on all RCFT measures of strategic planning and immediate nonverbal memory. We evaluated two different aspects of RCFT performance in this study: organizational strategies used in the initial copy condition, and construction accuracy in the copy and recall conditions. Patients with AN were impaired on both sets of measures. Based on findings in OCD and BDD, we had predicted that organizational strategies used during the initial figure copy would mediate group differences in memory. This hypothesis was supported by a series of regression analyses. In our model (see Figure 3), group differences in immediate memory (immediate percentage recall from copy) arise from the mediating effects of impaired organization in patients with AN during the copy condition. These results indicate that the essential cognitive deficit on the RCFT was impaired organization. This, in turn, affected how much information was encoded and retrieved from memory.

These findings are identical to previous findings in OCD[30,32,34] and BDD[31] and also consistent with earlier investigations in AN using the RCFT.[19,25] Results indicate that nonverbal memory problems, as measured by the RCFT, are strategic in nature. Strategic memory deficits such as these have been identified in neurologic groups with frontal-striatal system dysfunction,[49,50] and the prefrontal cortex has also been implicated in strategic memory processes by functional imaging studies.[51,52] Current findings in AN are, therefore, consistent with previous OC spectrum disorder research and indicate that frontal-striatal dysregulation may similarly underlie nonverbal memory deficits in AN and OC spectrum disorders. Findings from these investigations, taken together, indicate that OCD, BDD, and AN share significant neuropsychological features and at least some overlap in patterns of neural system dysregulation.

We have previously proposed that strategic encoding problems might contribute to clinical features of OC spectrum disorders, such as chronic doubt regarding the adequacy of previous behavior in OCD[30,32] and body image distortions in BDD.[32,46] It is also possible that strategic encoding problems have clinical significance in AN by contributing to body image disturbance, as well as body checking behaviors, compulsive calorie counting, and food rituals, for example. With regard to body image disturbance, patients with AN perceive themselves as "fat" even in the face of extreme wasting. One explanation for this, in light of current results, is that individuals with AN focus their attention primarily on individual body parts, such as the hips.[53] The hips are invariably anatomically

wider than the waist, even in the presence of wasting. Because patients with AN only attend to this one aspect of appearance, they may encode "big hips" and miss the full image of acute starvation and wasting. We have proposed an analogous process in BDD, in which patients focus on one detail of the face and perceive it to be flawed even though the face is normal in overall appearance.[46] Because OC symptoms persist after improvement of eating disorder symptoms, investigators have suggested that OC clinical features predate eating disturbances, and represent a vulnerability factor to these disorders.[8] Our findings raise the possibility that neuropsychological impairment might be one mechanism by which OC features contribute to the development of AN. These phenomenologic connections are, of course, speculative at this point, but suggest future directions for study.

In particular, these findings also offer directions for future intervention research. If executive functioning deficits can be identified as at least partial explanations for some of the OC symptoms in AN, cognitive behavioral interventions may be developed to specifically target these deficits. Preliminary studies have documented that OCD patients improve significantly in strategic verbal[54] and nonverbal[55] memory performance after simple instructional cueing. Further, gaining an understanding of the neural mechanisms underlying AN may also guide pharmacotherapy for these disorders.

Several limitations should be considered when interpreting current findings. First, a significant number of patients with AN had comorbid Axis I diagnoses and/or were taking psychotropic medication. Our analyses of the AN-only and nonmedicated subgroups indicate that these subject characteristics are not likely explanations for our results, and BDI measures of depression did not correlate with any of our cognitive measures. Nonetheless, we cannot completely rule out all confounding effects of medication and comorbidity and some caution is warranted until further replication in more restricted groups. Second, another potential factor in any study of AN is the physiologic and cognitive impact of food restriction and extreme weight loss. Although we cannot directly evaluate the impact of weight loss on our measures, it should be noted that none of the participants in our study were in acute medical crisis at the time of testing and BMI scores did not correlate with cognitive performance in patients with AN. Currently, whether cognitive functioning deficits in AN are secondary to low weight or persist postrecovery is unclear and in need of study (for review, see Tchanturia et al.[29]).

Third, although the evaluators were blind to participant group membership, it is likely that they were aware of an individual's AN status given the associated low weight. Finally, although the study evaluators underwent rigorous training in the administration and scoring of the RCFT, reliability estimates for construction and organizational accuracy were not calculated in this study. However, previous reliability estimates of organization scores have been very high ($>.90$[31,33,48]); to our knowledge reliability estimates for the Denman construction accuracy system have not been published.

In summary, this study found evidence of impaired nonverbal memory performance in a group of patients with AN that was mediated by poor strategies used to copy the figure. These findings were predicted based on potential similarities to OC spectrum disorders and all hypotheses were strongly supported. Current findings in AN are, therefore, consistent with previous OCD and BDD research and indicate that frontal-striatal dysregulation may underlie nonverbal memory deficits in both groups. Several investigators have suggested that AN be conceptualized as an OC spectrum disorder based on symptom similarity and comorbidity, as well as family studies. Our results are consistent with this view of AN and indicate that these patient groups also share important neuropsychological features. Future studies are needed to replicate and extend findings to other domains of neuropsychological functioning and to further evaluate whether these observed deficits persist postrecovery in AN.

References

1. Kaye W, Bulik CM, Thornton L, Barbarich N, Marsters K. Comorbidity of anxiety disorders with anorexia and bulimia nervosa. Am J Psychiatry 2004;161:2215–2221.
2. Kasvikis Y, Sakiris F, Marks IM, Basoglu M, Noshirvani H. Past history of anorexia nervosa in women with obsessive-compulsive disorder. Int J Eat Disord 1986;5:1069–1075.
3. Ruffolo S, Phillips KA, Menard W, Fay C, Weisberg RB. Comorbidity of body dysmorphic disorder and eating disorders: severity of psychopathology and body image disturbance. Int J Eat Disord 2006;39:11–19.
4. Bulik C, Sullivan PF, Fear JL, Pickering A. Outcome of anorexia nervosa: eating attitudes, personality, and parental bonding. Int J Eat Disord 2000;28:139–147.
5. Casper R. Personality features of women with good outcome from restricting anorexia nervosa. Psychosom Med 1990;52: 156–170.
6. Srinivasagam NM, Kaye WH, Plotnicov KH, Greeno C, Weltzin TE, Rao R. Persistent perfectionism, symmetry, and exactness after long-term recovery from anorexia nervosa. Am J Psychiatry 1995; 152:1630–1634.

7. Hsu L, Kaye W, Weltzin T. Are the eating disorders related to obsessive compulsive disorder? Int J Eat Disord 1993;14:305–318.
8. Halmi K, Sunday SR, Strober M, Kaplan A, Woodside DB, Fichter M, et al. Perfectionism in anorexia nervosa: variation by clinical subtype, obsessionality, and pathological eating behavior. Am JPsychiatry 2000;157:1799–1805.
9. Lilenfeld L, Kaye WH, Greeno CG, Merikangas KR, Plotnicov K, Pollice C, et al. A controlled family study of anorexia nervosa and bulimia nervosa: psychiatric disorders in first-degree relatives and effects of proband comorbidity. Arch Gen Psychiatry 1998;55:603–610.
10. Bellodi L, Cavallini MC, Bertelli S, Chiapparino D, Riboldi C, Smeraldi E. Morbidity risk for obsessive-compulsive spectrum disorders in first-degree relatives of patients with eating disorders. Am J Psychiatry 2001;158:563–569.
11. Delvenne V, Goldman S, De Maertelaer V, Simon Y, Luxen A, Lotstra F. Brain hypometabolism of glucose in anorexia nervosa: normalization after weight gain. Biol Psychiatry 1996;40: 761–768.
12. Delvenne V, Goldman S, De Maertelaer V, Wikler D, Damhaut P, Lotstra F. Brain glucose metabolism in anorexia nervosa and affective disorders: influence of weight loss or depressive symptomatology. Psychiatr Res 1997;74:83–92.
13. Delvenne V, Goldman S, De Maertelaer V, Lotstra F. Brain glucose metabolism in eating disorders assessed by positron emission tomography. Int J Eat Disord 1999;25:29–37.
14. Herholz K, Krieg JC, Emrich HM, Pawlik G, Beil C, Pirke KM, et al. Regional cerebral glucose metabolism in anorexia nervosa measured by positron emission tomography. Biol Psychiatry 1987;22:43–51.
15. Krieg J, Holthoff V, Schreiber W, Pirke KM, Herholz K. Glucose metabolism in the caudate nuclei of patients with eating disorders, measured by PET. Eur Arch Psychiatry Clin Neurosci 1991;240:331–333.
16. Rauch SL, Baxter LR. Neuroimaging in obsessive-compulsive disorder and related disorders. In: Jenike MA, Baer L, Minichiello WE, editors. Obsessive-compulsive disorders: practical management (3rd ed). St. Louis, MO: CV Mosby, 1998, p. 289–317.
17. Adler C, McDonough-Ryan P, Sax KW, Holland SK, Arndt S, Strakowski SM. fMRI of neuronal activation with symptom provocation in unmedicated patients with obsessive compulsive disorder. J Psychiatr Res 2000;34:317–324.
18. Kingston K, Szmukler G, Andrewes D, Tress B, Desmond P. Neuropsychological and structural brain changes in anorexia nervosa before and after refeeding. Psychol Med 1996;26:15–28.
19. Jones B, Duncan CC, Brouwers P, Mirsky AF. Cognition in eating disorders. J Clin Exp Neuropsychol 1991;13:711–728.
20. Szmukler G, Andrewes D, Kingston K, Chen L, Stargatt R, Stanley R. Neuropsychological impairment in anorexia nervosa: before and after refeeding. J Clin Exp Neuropsychol 1992;14:347–352.
21. Abbruzzese M, Ferri S, Scarone S. The selective breakdown of frontal functions in patients with obsessive-compulsive disorder and in patients with schizophrenia: a double dissociation experimental finding. Neuropsychologia 1997;35:907–912.
22. Purcell R, Maruff P, Kyrios M, Pantelis C. Neuropsychological deficits in obsessive-compulsive disorder: a comparison with unipolar depression, panic disorder, and normal controls. Arch Gen Psychiatry 1998;55:415–423.
23. Hanes K. Neuropsychological performance in body dysmorphic disorder International J Neuropsychol Soc 1998;4:167–171.
24. Green M, Elliman NA, Wakeling A, Rogers PJ. Cognitive functioning, weight change and therapy in anorexia nervosa. J Psychiatr Res 1996;30:401–410.
25. Mathais J, Kent PS. Neuropsychological consequence of extreme weight loss and dietary restriction in patients with anorexia nervosa. J Clin Exp Neuropsychol 1998;20:548–564.
26. Cohen L, Hollander E, DeCaria CM, Stein DJ, Simeon D, Liebowitz MR, et al. Specificity of neuropsychological impairment in obsessive-compulsive disorder: a comparison with social phobic and normal control subjects. J Neuropsychiatr Clin Neurosci 1996;8:82–85.
27. Deckersbach T, Otto MW, Savage CR, Baer L, Jenike MA. The relationship between semantic organization and memory in obsessive-compulsive disorder. Psychother Psychosom 2000;69: 101–107.
28. Savage C, Keuthen NJ, Jenike MA, Brown HD, Baer L, Kendrick AD, et al. Recall and recognition memory in obsessive-compulsive disorder. J Neuropsychiatr Clin Neurosci 1996;8:99–103.
29. Tchanturia K, Morris RG, Anderluh MB, Collier DA, Nikolaou V, Treasure J. Set shifting in anorexia nervosa: an examination before and after weight gain, in full recovery and relationship to childhood and adult OCPD traits. J Psychiatr Res 2004;38: 545–552.
30. Savage C, Baer L, Keuthen NJ, Brown HD, Rauch SL, Jenike MA. Organizational strategies mediate nonverbal memory impairment in obsessive-compulsive disorder. Biol Psychiatry 1999;45: 905–916.
31. Deckersbach T, Savage CR, Phillips KA, Wilhelm S, Buhlman U, Rauch SL, et al. Characteristics of memory dysfunction in body dysmorphic disorder. J Int Neuropsychol Soc 2000;6:673–681.
32. Savage C, Deckersbach T, Wilhelm S, Rauch SL, Baer L, Reid T, et al. Strategic processing and episodic memory impairment in obsessive compulsive disorder. Neuropsychology 2000;14:141–151.
33. Osterrieth P. Le test de copie d'une figure complex: contribution à l'étude de la perception et de la memoire. Arch Psychol 1944;30:286–350.
34. Mataix-Cols D, Alonso P, Hernandez R, Deckersbach T, Savage CR, Manuel Menchon J, et al. Relation of neurological soft signs to nonverbal memory performance in obsessive-compulsive disorder. J Clin Exp Neuropsychol 2003;25:842–851.
35. Shin M, Park SJ, Kim MS, Lee YH, Ha TH, Kwon JS. Deficits in organizational strategy and visual memory in obsessive-compulsive disorder. Neuropsychology 2004;18:665–672.
36. Penades R, Catalan R, Andres S, Salamero M, Gasto C. Executive function and nonverbal memory in obsessive-compulsive disorder. Psychiatr Res 2005;133:81–91.
37. Choi J, Kang DH, Kim JJ, Ha TH, Lee JM, Youn T, et al. Left anterior subregion of orbitofrontal cortex volume reduction and impaired organizational strategies in obsessive-compulsive disorder. J Psychiatr Res 2004;38:193–199.
38. Savage C. The role of emotion in strategic behavior: insights from psychopathology. In: Salovey FB, editor. The wisdom in feeling. New York: The Guilford Press, 2002, p. 211–236.
39. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV(4th ed). Washington, DC: American Psychiatric Association; 1994.
40. First M, Spitzer RL, Gibbon M, Williams JBW. Structured clinical interview for DSM-IV Axis I Disorders-Patient Edition (SCID-I/D, version 2.0). New York: Biometrics Research Department, New York Psychiatric Institute; 1995.
41. Garner D, Olmstead MP, Polivy J. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. Int J Eat Disord 1983;2:15–34.
42. Beck A, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571.
43. Psychological Corporation. WASI Manual. San Antonio, TX: Harcourt Brace; 1997.
44. Psychological Corporation. WAIS-III, WMS-III Technical Manual. San Antonio, TX: Harcourt Brace; 1997.

45. Denman S. Denman Neuropsychology Memory Scale. Charlestown, SC: Author; 1984.
46. Deckersbach T, Savage CR, Phillips KA, Wilhelm S, Buhlman U, Rauch SL, et al. Characteristics of memory dysfunction in body dysmorphic disorder. J Int Neuropsychol Soc 2000;6:673–681.
47. Deckersbach T, Savage CR, Henin A, Mataix-Cols D, Otto MW, Wilhelm S, et al. Reliability and validity of a scoring system for measuring organizational approach in the Complex Figure Test. J Clin Exp Neuropsychol 2000;22:640–648.
48. Baron RM, Kenny DA. The moderator-mediator variable distinction in social and psychological research: conceptual, strategic, and statistical considerations. J Personality Social Psychol 1986; 51:1173–1182.
49. Savage C. Neuropsychology of subcortical dementias. Psychiatr Clin North Am 1997;20:911–931.
50. Gabrieli J. Cognitive neuroscience of human memory. Annu Rev Psychol 1998;49:87–115.
51. Cabrera A, McNally RJ, Savage CR. Missing the forest for the trees? Deficient memory for linguistic gist in obsessive-compulsive disorder. Psychol Med 2001;31:1089–1094.
52. Savage C, Deckersbach T, Heckers S, Wagner AD, Schacter DL, Alpert NM, et al. Prefrontal regions supporting spontaneous and directed application of verbal learning strategies: evidence from PET. Brain 2001;124(Pt 1):219–231.
53. Cash TF, Deagle EA III. The nature and extent of body-image disturbances in anorexia nervosa and bulimia nervosa: a meta-analysis. Int J Eat Disord 1997;22:107–125.
54. Deckersbach T, Savage CR, Dougherty DD, Bohne A, Loh R, Nierenberg A, et al. Spontaneous and directed application of verbal learning strategies in bipolar disorder and obsessive-compulsive disorder. Bipolar Disord 2005;7:166–175.
55. Buhlman U, Deckersbach T, Engelhard I, Cook LM, Rauch SL, Kathmann N, et al. Cognitive retraining for organizational impairment in obsessive-compulsive disorder. Psychiatr Res (in press).

Risk adjustment and risk-adjusted provider profiles

Michael Shwartz*

Department of Operations and Technology Management,
School of Management, Boston University,
595 Commonwealth Avenue, Boston, Massachusetts 02215, USA
E-mail: mshwartz@bu.edu
*Corresponding author

Arlene Ash

Section of General Internal Medicine, Boston Medical Center,
720 Harrison Avenue, # 1108, Boston, Massachusetts 02118, USA
E-mail: aash@bu.edu

Erol Peköz

Department of Operations and Technology Management,
School of Management, Boston University,
595 Commonwealth Avenue, Boston, Massachusetts 02215, USA
E-mail: pekoz@bu.edu

Abstract: Provider profiles are an important component of efforts to improve both healthcare productivity and quality. Risk adjustment is the attempt to account for differences in the risk for specific outcomes of cases in different groups in order to facilitate more meaningful comparisons. The value of provider profiles, whether used internally or released to the public, depends both on adequate risk adjustment and on distinguishing systematic differences in either the quality or productivity of care from random fluctuations. In this paper, we discuss how risk adjustment systems can be used to predict outcomes for individual cases; methods for measuring and comparing the performance of risk adjustment systems; and how random variation can affect provider performance data and several approaches for addressing random variation in risk-adjusted provider profiles. In the last section, we illustrate the methods discussed by developing and analysing risk-adjusted provider profiles for 69 family practitioners responsible for over 68,000 cases.

Keywords: risk adjustment; provider profiling.

Reference to this paper should be made as follows: Shwartz, M., Ash, A. and Peköz, E. (2006) 'Risk adjustment and risk-adjusted provider profiles', *Int. J. Healthcare Technology and Management*, Vol. 7, Nos. 1/2, pp.15–42.

Biographical notes: Michael Shwartz is Professor of Healthcare and Operations Management at Boston University's School of Management. His research interests include the use of mathematical models of disease processes to analyse screening programmes, studies of the relationship between severity of illness and hospital costs and outcomes, healthcare payment, analyses related to quality of care, small area variation analyses, evaluation of substance abuse programmes, and analyses of large databases. He was Founding Director of the Institute for Healthcare Policy and Research within the Boston Department of Health and Hospitals.

Arlene Ash is a research professor at Boston University's Schools of Medicine (Division of General Internal Medicine) and Public Health (Department of Biostatistics), and a Co-founder and senior scientist at DxCG, Inc., a company that licenses risk-adjustment software. She is an internationally recognised expert in the development and use of risk-adjustment methodologies and an experienced user of national research databases, especially for Medicare, where her diagnostic cost group models have been adapted for use in HMO payment.

Erol Peköz is an associate professor of Operations and Technology Management at Boston University's School of Management. Working in the area of quantitative methods, much of his research has focussed on statistical models for understanding congestion and delay in service operations. Current research focuses on statistical models for geographic variations in healthcare usage and supply-induced demand for hospital beds. Other areas of interest include risk management, the theory of rare events, and Monte Carlo simulation.

1 Introduction

Improving productivity and quality requires changing provider behaviour. Although healthcare providers clearly want to 'do the best' for their patients, real improvement requires major cultural, structural and process changes in the healthcare delivery system. Documenting significant variations in either processes of care or outcomes is an important component of change programmes. As providers usually respond to credible data, many healthcare organisations routinely examine and internally share provider performance data as part of their improvement efforts. As market-based solutions to spiralling healthcare costs also require performance data on providers, 'Intense enthusiasm for public reporting of healthcare performance continues unabated' (Clancey, 2003). The value of provider profiles, whether used internally or released to the public, depends upon their credibility, in particular, how well they account for differences in the inherent risk of their patient panels and how well they avoid mistaking random fluctuations for systematic differences in either the quality or productivity of care.

Creating provider profiles consists of the following. First, one must decide how cases are defined, what outcomes are of interest and how cases are assigned to providers. For example, a case could be a hospital admission; outcomes could be cost or whether a complication or a death occurs; and, cases might be assigned to the initial hospital where care begins or to the hospital to which a patient is transferred. Or, a case could be a 'person year' of healthcare experience; the outcome that person's healthcare costs the following year; and, each case might be assigned to their primary care provider. Second, a risk adjustment model is used to determine a predicted value, PRED, of the outcome, Y, for each case. For example, if Y is cost, then PRED is expected cost; if Y is dichotomous (that is, it equals 1 when a particular outcome, such as death, occurs, and 0 otherwise) then PRED is the estimated probability that the outcome occurs. Third, the individual Ys and PREDs are averaged for each provider, leading to an observed (O = average Y) and expected (E = average PRED) among that provider's cases. Finally, discrepancies between O and E are put in context, that is, they are judged for their size

and statistical significance. Large discrepancies might lead to certain managerial actions, such as publishing 'physician scorecards' and steering patients to better-performing providers.

This paper has four main Sections. Section 2 discusses how risk adjustment systems are used to predict outcomes for individual cases, that is, to determine the PREDs. Section 3 discusses methods for measuring and comparing the performance of risk adjustment systems, specifically examining how close the PREDs are to the Ys. Section 4 discusses how random variation can affect provider performance data and several approaches for addressing random variation in provider profiling. Finally, we apply the methods discussed in the earlier sections to real data and draw conclusions.

Productivity and quality have typically been conceptualised as separate dimensions. Managers have attempted to achieve gains in productivity without sacrificing quality. However, this formulation creates unnecessary tensions between managers and caregivers. Alternatively, the Institute of Medicine report, *Crossing the Quality Chasm,* views healthcare quality itself as consisting of six dimensions: safety, effectiveness, patient-centredness, timeliness, efficiency and equity. Three of these dimensions address productivity: effectiveness (matching care to science, avoiding both overuse of ineffective care and underuse of effective care), efficiency (the reduction of waste, including waste of supplies, equipment, space, capital) and timeliness (reducing waiting times and delays for both patients and those who give care). Aligning productivity with quality shifts managerial and policy focus: high quality care is by definition productive, and efforts to improve productivity raise quality. In this spirit, we consider methods that apply to a range of outcomes, including pure productivity measures, such as cost, as well as traditional quality outcomes, such as mortality.

2 Calculating expected outcomes

Iezzoni (2003) urges prospective users of a risk adjustment system to consider various conceptual and practical issues, including: the clinical dimensions measured, data demands, content and face validity, reliability, outcomes of interest (example, death, functional impairment, resource consumption), the population to be studied (example, older versus younger patients, those with particular conditions, those seen in particular settings), the time period of interest (example, the first 30 days following hospital admission, a calendar year), and the factors whose potential influence on the outcome is of primary interest (example, the type of therapeutic approach used, the provider or the type of institution at which the patient was treated).

After resolving these qualitative issues, a key question remains: how well does the system account for differences in patient risk? Comparing methods is easiest when the performance measure is a single number, where higher (or lower) is better, especially when previous experience suggests that values that exceed some threshold are 'good enough'.

Some risk adjustment systems assign each individual a score that directly reflects the outcome of interest. For example, a version of MedisGroups assigns each individual a predicted probability of death (for those interested in the details of the particular risk adjustment systems mentioned in this article, see Iezzoni (2003)). These predicted probabilities were derived from the database MedisGroups used to develop its models.

When using this system for risk adjustment in another population, the average predicted probability from MedisGroups may not correspond to the actual probability of death in that population. To the extent the MedisGroups database can be considered a standard, a difference between the actual and predicted death rate may indicate higher or lower quality of care. However, new populations typically differ from the particular population used to develop the risk adjustment model in ways both subtle and large that cast doubt on such a conclusion. Thus, even when using a risk score that directly estimates the outcome of interest, some form of 'recalibration,' that is, forcing the average predicted outcome to equal the average actual value in the new population, is usually a good idea. Recalibration, of course, eliminates the ability to compare the new population's outcomes to the external standard. However, it enhances the ability to compare subgroups within the population of interest to the population's own norm.

Some systems provide a dimensionless score that must be calibrated to create a prediction. For example, the DCG (Diagnostic Cost Group) prospective relative risk score (RRS) represents next year's expected total healthcare costs as a multiple of average cost. A RRS of 1.2 means expected resource consumption is 20% greater than average. The easiest way to convert such scores into predictions is to multiply each score (RRS) by a proportionality constant, k, where:

$$k = \text{average value of the outcome} / \text{average value of RRS},$$

with both averages being taken over the entire population of interest. By definition, the prediction PRED = k* RRS is calibrated to the new data and outcome, that is, both PRED and the outcome that it predicts have the same average in that population.

More flexible calibration methods may be especially useful when estimating new outcomes (such as, the probability of death as a function of RRS) or the same outcome in a markedly different setting (such as, the cost of care in a fundamentally different delivery system), when the best predictions may not be a simple multiple of the original score. One useful approach in this case is 'risk score bucketing,' where cases are first ranked in order of their risk scores and then put into 'buckets' or 'bins' with other cases with similar risk scores. For example, buckets could be chosen to contain equal numbers of cases (example, deciles of increasing risk), to focus on high or low-risk subpopulations, or to match categories that have been used before (such as scores that fall within pre-specified ranges). When predicting healthcare costs, buckets containing uneven percentiles of cases are likely to be particularly informative, example, buckets defined by risk scores at the 20th, 50th, 80th, 90th, 95th, 99th and 99.5th percentile. Since the prediction for each case is the average value of the outcome for all cases in the same bucket, buckets should contain enough cases to produce a stable average outcome in each bucket. For a highly skewed outcome like costs, this may require 500 or more cases.

The predictive models that underlie risk adjustment systems generally have one of two basic structures: categorical grouping or regression modelling. Categorical models ('groupers') use the information about each case to place it into exactly one of several categories (groups or buckets). To the extent that cases in the same group have similar morbidity profiles, they may be at similar levels of risk for various health outcomes. The Adjusted Clinical Groups (ACG) system, for example, assigns each person to one of 93 ACG categories, based on a morbidity profile, age and sex (Version 5.0,

http://www.acg.jhsph.edu/). A prediction for each person is then calculated as the average value of the outcome of interest for all the cases in the same group.

In contrast, regression models move directly from patient profiles (containing demographic and clinical descriptors) to predictions. For example, the Diagnostic Cost Group, Hierarchical Condition Category (DCG/HCC) model assigns each person a score by adding values (called coefficients) for whichever condition categories are present for that person to an age/sex coefficient (Version 6.2 models code for 184 condition categories and a limited number of interactions among categories, http://dxcg.com/method/index.html). The Chronic Illness and Disability Payment System (CDPS), widely used by Medicaid programmes, also relies on regression (http://medicine.ucsd.edu/fpm/cdps/doc.html).

The distinctions between groupers and regression models may not matter that much to users, since most vendors of risk adjustment methods enable users to move smoothly from each person's data to a medical profile to a prediction. However, even when presented with a direct prediction of the outcome of interest, the user may still face a calibration problem, which can be solved, as above, by replacing each PRED with NEWPRED = k*PRED, where k is chosen to make the average of NEWPRED equal to the average outcome in the user's population.

It is also relatively easy to re-estimate the models on a new population. With a categorical model, a natural estimate for each case is the average value of all cases in the new population in the same category. The only limitation on this method arises when categories have too few cases to create a trustworthy average, in which case the solution is to combine (or 'pool') small categories with somewhat similar larger ones before calculating averages. A regression model can also be refitted to the new population. However, without 'tweaking', it may yield negative predictions for some cases, or inappropriate coefficients for rare conditions. The easiest way to recalibrate a regression model is via risk score bucketing, as described above. This simple technique enables the user to convert any predictive model into a customised 'grouper' with however many groups, of whatever size, seem most useful.

In general, it is important to recognise that whenever models are developed or coefficients estimated using a particular data set (the development data), its 'predictions' are precisely tailored to the particular outcomes that have already been observed. When these same models are used to actually predict what will happen in new (validation) data, the fit of these predictions to the new outcomes is usually not as good. Relatively simple recalibration of models, based on risk bucketing (or categorical groupings) with many cases per group, are the least subject to this problem.

3 Comparing risk adjustment systems

In this section, we discuss how to calculate and interpret the summary numbers commonly used to characterise the performance of risk adjustment systems.

3.1 Continuous outcomes

The standard measure of how well a risk adjustment system predicts a continuous outcome (like cost or length of hospital stay) is R^2. This number is driven by the ratio of

two sums, each taken over all cases. The first, the 'sum of squares total' or SST, is a measure of total variability in the outcome. SST is a property of the data and not of any model. It is computed as follows. Let Y-bar be the average value of outcome Y. For each case, compute the difference between its outcome and this average (Y – Y-bar) and square it. SST is the sum of these squared differences. The second, the 'sum of squares error' or SSE, is a model-specific measure of the variability of actual values from model predictions. For each case, compute the model's 'error' in predicting the outcome, (Y – PRED), and square it. SSE is the sum of these squared differences. Notice that the closer the model predictions are to the actual values of the outcome, the closer SSE is to zero. SSE/SST measures the proportion of total variability in the data that remains after applying the model. R^2, calculated as 1 – (SSE/SST), is said to measure the proportion of total variability in the data 'explained by the model'. In a calibrated model (where the average PRED equals Y-bar), R^2 can also be calculated as the square of the correlation between the actual outcome for each person and the predicted outcome.

What kind of R^2 can one expect to find? The answer to this question depends to some extent on the particular data set used and to a large extent on what is being predicted. For example, various risk adjusters can use year-1 patient descriptors to predict either year-2 total healthcare costs (prospective modeling) or year-1 costs (concurrent modelling). The Society of Actuaries (Cumming *et al.*, 2002) recently compared the predictive power of 7 major risk adjustment systems that rely on diagnoses found in administrative data. Their data set included almost 750,000 members of commercially insured plans. When the models were calibrated (through regression) to the population, R^2 values for concurrent models ranged from 0.24 (for Medicaid Rx, a system originally developed for a Medicaid population) to 0.47 (for DCGs, version 5.1). When prospective predictions were made, R^2 values ranged from 0.10 (for ACGs, version 4.5) to 0.15 (for 3 of the 7 systems). R^2 values are much higher for concurrent models, which estimate expected costs for treating known problems, than for prospective models, which use the problems being treated this year to predict costs for the (as yet unknown) problems that will arise and require treatment next year.

As extreme values can have a large effect on model estimates, risk adjustment systems are often fitted to modified data, often by dropping 'outlier' cases. For example, Medicare uses Diagnosis Related Group (DRG)-based prospective payment to reimburse hospitals for admissions. The DRG grouper places each admission in exactly one of approximately 500 categories. Within each category, outliers are defined as cases whose costs are more than 3 standard deviations from the mean after transforming data to logs. DRG payment weights are estimated after dropping these outliers. Since dropping outlier cases reduces both SSE and SST, R^2 could either increase or decrease, depending on which term decreases more.

A second approach that we prefer is to retain all useable cases but to reset extreme values to something less extreme. 'Topcoding' and 'winsorising' are technically distinct methods for achieving this, but the terms are often used interchangeably. To topcode a variable at a fixed threshold, T, all values larger than T are given the value T. 'Truncation' is also used in the same context, although in more standard usage, truncated data have cases with extreme values removed. We prefer topcoding values to dropping cases because the most expensive cases are important and should not be completely ignored. Topcoding is particularly well suited to modelling in a system that pays for cases above some threshold out of a separate (reinsurance) pool. The analysis by the

Society of Actuaries examined models to predict actual costs when costs were topcoded at $50,000 and when they were top-coded at $100,000. Since it is easier to predict less-skewed outcomes, R^2 values were higher with topcoding, e.g., up to 0.10 higher for some concurrent predictions when costs were topcoded at $100,000.

An alternative to the above approaches is to transform the dependent variable in order to 'pull in' the outliers, most often done by replacing Y with its natural logarithm. However, no real-world administrator cares about how well we can predict log (dollars); the predictions must be transformed back to dollars (via exponentiation) and, perhaps, calibrated through multiplication by an appropriate constant before asking how well the predictions 'perform'. Duan's smearing estimator is a theoretically attractive way to find the calibrating constant for log-transformed data (Duan, 1983). However, the resulting predictions are often too small. Simply multiplying by the number needed to make the average prediction equal to the actual average value of the outcome will produce better-fitting predictions. However, retransformed predictions, even when perfectly calibrated, often do not perform as well as predictions from ordinary least squares models that simply treat the original data as if it were normal.

Generalised Linear Models (GLM) provide an alternative, comprehensive framework for modelling non-normally distributed data. Buntin and Zaslavsky (2004) evaluate several such models.

Ultimately, R^2 values depend heavily on features of the data set used, in particular, the amount of variation in both the dependent and independent variables. Hence, rather large differences in reported R^2s for different risk adjustment methods may simply reflect the relative difficulty of predicting outcomes in a particular database rather than any inherent difference in the systems. This makes studies (like the Society of Actuaries report) that examine the performance of several different models on the same data particularly valuable.

In calculating R^2, each difference between a person's predicted value and actual value contributes to the model's error and reduces R^2. In many settings, however, the main purpose of risk adjustment is not to predict correctly for each person, but to produce correct average predictions for groups of patients. For example, in provider profiling, we want to know how closely average predictions for providers match with their average actual outcomes. Grouped R^2 statistics, an analog of the traditional (individual) R^2, are single summary measures that answer such questions. A Grouped R^2 can be calculated for any way of partitioning the population which places each person into one and only one group; different partitions lead to different Grouped R^2s. The Grouped R^2 analog of SSE, call it GSSE, is computed as follows: for each group, square the difference between its average actual outcome and its average predicted outcome, multiply this squared difference by the number of people in the group, and sum over all groups. The analog of SST, GSST, is computed as follows: for each group, square the difference between its average actual outcome and the average actual outcome in the population, multiply this squared difference by the number of people in the group, and sum over all groups. The Grouped R^2 is then calculated as 1 – GSSE/GSST. Note that its value depends not only on the data set being used, but also on the way the population is grouped. In calculating the Grouped R^2, multiplying by the size of each group ensures that errors in predicting the average outcome counts less for smaller groups; for groups of equal size, say 10 deciles based on prior cost, the step of 'multiplying by the size of the group' can be ignored.

Although R^2 is widely used as a summary measure of risk adjustment system performance, neither it nor the Grouped R^2 provides much intuitive feel for the ability of a system to discriminate among cases with high and low values of the outcome variable. To provide such insight, we recommend examining actual outcomes within deciles of predicted outcome. That is, array the data from lowest predicted value of the outcome to highest, divide the data into deciles, and calculate the average value of the actual outcome in each decile. We used this approach to compare different ways of predicting hospital Length of Stay (LOS) for pneumonia cases. To illustrate, a model with age, sex and DRG had an R^2 of 0.10. Mean actual LOS in the lowest and highest deciles were 5.2 days and 11.9 days, respectively. The highest R^2 of the systems examined was 0.17. For this system, mean actual LOS in the lowest and highest deciles were 4.1 days and 13.2 days respectively. Thus, the higher R^2 translated into being able to find a low risk 10% of cases with lengths of stay that averaged 1 day less and a high risk 10% that averaged over 1 day more than the DRG-based model.

Finally, suppose that the goal is to identify a small number of high-cost cases (or 'top groups') for disease management programmes or other interventions. Ash *et al.* (2001) and Zhao *et al.* (2003) illustrate some approaches for comparing the ability of different risk adjustment models to identify useful top groups, for example, those that contain

- few 'bad picks', that is, people whose costs will actually be quite low
- many 'good' or 'great' picks (those with the highest costs)
- many people with potentially manageable diseases (example, diabetes or asthma).

3.2 Dichotomous outcomes

Logistic regression models are usually used to predict dichotomous outcomes, such as whether or not someone dies within a specified time frame. The most widely used summary measure of the performance of a logistic regression model is the c statistic. There are several equivalent definitions of the c statistic, one of which is the following: among all possible pairs of cases such that one dies and the other lives, the c statistic is the proportion of pairs in which the predicted probability of death is higher for the person who died. Note that c does not depend on the model's actual predictions, but only on their ranks (for example, if all the predictions were divided by 2, c would not change). Thus, c measures the model's ability to discriminate between those with the event and those without it. c achieves its maximum value of 1.0 when all predicted values for cases with the event are larger than any predicted values for cases without the outcome. When the model has no ability to discriminate (example, probabilities are randomly assigned to cases with and without the event), the expected value of the c statistic is 0.5.

Some of the best risk adjustment models achieve c statistics around 0.9. For example, Knaus *et al.* (1991) reported a c statistic of 0.9 for predicting death using APACHE III, based on the same roughly 17,000 ICU patients on which the model was developed. To examine the quality of surgical care in VA hospitals, Khuri *et al.* (1997) developed logistic regression models to predict 30-day mortality using data from over 87,000 non-cardiac operations. Different versions of the model had c statistics ranging form 0.87 to 0.89. The New York State risk adjustment model for predicting death following CABG surgery has a c-statistic of 0.79 (Hannan *et al.*, 1997).

Models can discriminate well (between those with the event and those without) without calibrating well. For example, consider a sample of cases with a death rate of 10%. A model that predicts a 0.2 probability of death for everyone who lives and 0.3 for all who die discriminates perfectly. However, it is poorly calibrated, since the average predicted death rate in the sample is over 20% while the actual death rate is 10%. Alternatively, a model that predicts a death rate of 0.1 for all cases would be perfectly calibrated, but have no ability to discriminate. Researchers differ over the relative importance of calibration versus discrimination. However, as we saw above, it is easy to recalibrate a model; it is harder to improve a model that doesn't discriminate well.

The most widely used method to check for calibration is the Hosmer-Lemeshow chi square test, which, however, does not check directly for overall calibration, that is, whether the average of the predicted outcomes approximately equals the average of the actual outcomes. Rather, it assesses whether average and predicted rates are similar within subgroups of cases, most commonly, deciles of predicted risk. Specifically, within each subgroup, the following quantity is calculated:

$$(\#\text{ observed alive} - \#\text{ predicted alive})^2 / \#\text{ predicted alive} +$$

$$(\#\text{ observed dead} - \#\text{ predicted dead})^2 / \#\text{ predicted dead.}$$

These quantities are summed over the subgroups and the result, call it X, is compared to a chi square distribution; with ten groups, this reference distribution has eight degrees of freedom. Smaller values of X indicate better concordance between observed and expected values. Formally, the model is accepted if the p-value for this test is reasonably *large* (for example, when the subgroups are deciles, X needs to be less than 13.3 for the p-value to be 0.10 or larger). Unfortunately, this test is very sensitive to sample size, since with many cases, even small deviations of observed from expected yield small p-values (and with few cases, the reverse is true: that is, even large deviations will not be significant). For example, Khuri *et al.* (1997), in connection with the model developed on VA cases to predict 30-day mortality for non-cardiac surgery patients reported: 'The only goodness-of-fit statistics that were statistically significant at the 0.05 level, were for all operations combined and for general surgery, primarily because there were a large number of cases in these categories' (over 87,000). It is, however, possible for large samples to be so well calibrated that they 'pass' this test; for example, the New York State CABG model, fitted to over 57,000 cases, yielded a Hosmer-Lemeshow test p-value of 0.16.

4 Random variation and provider profiling

4.1 The effect of randomness

First, we briefly review the effect of randomness on the nature of conclusions that can be drawn from provider profiles. Initially, we assume that all patients have the same risk for the outcome of interest; that is, we assume that all providers have panels with the same risk and, thus, that risk adjustment is not an issue.

The most widely used summary measure of the amount of variability in a data set is the Standard Deviation (SD), which measures how far away data points are from their average. For variables with bell-shaped distributions, most of the data (often

approximately two-thirds) will be within one SD from the average and few observations will be more than 2 or 3 SDs from the average.

Assume the outcome of interest is total healthcare cost during a year. The usual model underlying provider profiling views the costs of the particular patients treated by a provider as a sample of size n from some 'parent' population of cases that might have been treated by this provider. Let M (for mean) equal the unknown average cost of all patients in this parent population. Envision drawing many samples of size n from the parent population and calculating the average cost of patients in each sample. Let A_i = the average cost calculated from sample i. The A_is calculated from these samples will vary around the true but unknown mean M because costs in the parent population vary. For example, some samples will happen to contain a few more cases with large costs than others. In the same sense that the SD measures how far away data points are from their average, a Standard Error (SE) measures how much a statistic calculated from a sample (example, A_i, the mean of the i^{th} sample) is likely to vary from the number in the parent population that it estimates (in this case, M). The SE of the mean depends both on the inherent variability of the outcome in the parent population (SD) and on the size of the sample (n).

Although we do not know the SD of the parent population, we can estimate it from the variability of costs among the patients of the providers in our sample. A reasonable estimate of the SD of the population, called the 'pooled' estimate of the SD and indicated by SD_p, can be calculated by taking a 'weighted average' of the variance (the SD squared) of each provider's patients' costs and then taking the square root of this quantity. Once SD_p is known, the SE_i for provider i is $SD_p / \sqrt{n_i}$, where n_i is the number of patients seen by that provider. In using the pooled estimate of the SD from a number of providers, we allow for the possibility that 'true' mean costs may differ by provider (i.e., there is a different M for each provider), but assume the inherent variability of costs for each provider is the same.

We observe A_i, the average cost of provider i's patients and want to know how trustworthy it is as an estimate of provider i's 'true' average cost M_i. Statistical theory tells us that if n is large enough, there is approximately a 95% chance that the interval $A_i \pm 2 * SE_i$ will include the true mean M_i; thus, this interval is called a 95% confidence interval. 'Large enough' is often thought of as 30 or more, but when the outcome distribution is very skewed, as healthcare cost data are, sample sizes of several hundred may be required before the interval $A_i \pm 2 * SE_i$ really has a 95% chance of containing M_i. Suppose a provider's panel contains n = 100 patients. If the SD_p is approximately as large as A_i (it is often larger) and if average costs were $1,000, the interval that goes from $A_i - 2 * SE_i$ to $A_i + 2 * SE_i$ would approximately range from $800 to $1,200. This $400 width broadly indicates the range of uncertainty associated with using A_i to estimate M_i.

To highlight the effect of randomness in the context of profiling, let A = observed average cost of all patients treated by all of the providers. We assume that this observed average over the patients of all providers estimates the 'true' but unknown average, M, with essentially no error (since the pooled sample is very large). If all providers really have the same average cost M (estimated by A), there is a 95% chance that provider i's observed average cost A_i will be in the interval $A \pm 2 * SE_i$. If provider i's observed average cost does fall in this interval, provider i is viewed as performing 'as expected.' If provider i's cost falls outside the interval, he or she is assumed to be performing better or worse than expected, depending on whether cost is below or above the bounds of the

interval. To illustrate, assume again that overall average costs are $1,000, SD is also about $1000, and that provider i has 100 patients. If provider i's practice really is 'average,' there is a 95% chance that his or her observed average (based on these 100 patients) will be between $800 and $1,200. As long as provider i's costs fall between these bounds, we usually view his or her practice as 'normal'.

Several issues should be kept in mind when using the above approach. Imagine we profile 100 providers, all of whom treat 100 patients each and all of whom have the same 'true' average cost of $1,000. Any provider whose observed average cost falls outside the interval $800 to $1,200 will be flagged, either as a low cost provider whose practice might serve as a benchmark for others, or a high cost provider who might be the focus of interventions to lower costs. However, because of the way in which the above interval is constructed, 5% of the providers whose true average is $1,000 will fall outside the interval just due to random chance. Among providers who fall outside the interval, it is not clear which really have outlier practices and which have normal practices but samples that looked abnormal due to random chance. In traditional hypothesis testing, mistakenly concluding that a normal provider has an outlier practice is called a type I error. With 95% confidence intervals, 5% of normal practices will receive type I error 'flags'.

The other important type of error is not identifying a provider whose practice really is aberrant. Imagine that one of our provider's costs actually average $1,200, 20% above the rest of the providers. There is about a 50% chance that this provider's observed average will fall below $1,200 (and above $800), and thus an approximately 50% chance that the provider will not be flagged. Failing to identify an outlier provider is called a type II error. The more aberrant the data for the outlier provider, the smaller the chance of a type II error. For example, if the outlier provider had a true average of $1,400, the chance that the observed average would fall below $1,200 is only about 2.5%.

These same considerations apply when examining a dichotomous outcome. Using data on cardiac catheterisation, Luft and Hunt (1986, p.2780) showed that small numbers of patients and relatively low rates of poor outcomes make it difficult to 'be confident in the identification of individual performers'. For example, suppose the death rate is 1%, but a hospital treating 200 patients experiences no deaths. Even using a lenient 10% chance of a type I error (which narrows the interval for concluding the provider is performing as expected), determining whether the hospital had statistically significant better outcomes is impossible. For another example: with an expected death rate of 15%, 5 deaths in 20 patients (25% mortality) would not provide convincing evidence of a problem.

Thus, particularly when sample sizes are small (as they often are in the types of condition-specific provider profiles most useful for improvement), provider comparisons need to address the fact that random chance strongly affects 'raw' rates.

4.2 Adding risk adjustment

It is easy to incorporate into the above framework the fact that providers often treat patients with different risks. As discussed in Section 1, risk adjustment models provide a predicted outcome for each patient. The average of the predicted outcomes for patients seen by a provider is their expected outcome (E). This is compared to the average observed outcome of the patients seen by the provider (O). If a provider's observed (O) is much different than the expected (E), the provider is flagged as an outlier.

Deviations between O and E can be assessed by comparing either (O–E) to zero or the ratio O/E to 1. Neither is inherently superior. Which is worse: a 2% complication rate when only 1% was expected (a 100% higher complication rate but only 1 excess problem per 100 patients), or a 50% complication rate when only 40% was expected (only a 25% higher complication rate, but 10 excess problems per 100 people). In practice, expected differences across providers will be far less than 1% versus 40%. If expected outcomes differ dramatically, the underlying patient populations or other characteristics are likely to be too different for meaningful comparisons; when providers' expected outcomes are roughly similar, difference and ratio measures of performance produce similar judgments about the relative performance of various providers.

Common practice is to focus on O/E (referred to as the 'O to E ratio'). This ratio is 'centered' at 1 (normative values are approximately equal to 1), but could range from 0 to infinity. Sometimes one examines the log (O/E), which stretches the scale, so that, for example, the distance between points with O/E ratios of 0.25, 0.50, 1.00, 2.00 and 4.00 are equally spaced, since each value doubles the preceding one. The O/E ratio is unstable when E is close to 0. When E is an expected number of events equal to 5 or more, O/E is reasonably stable.

Interpreting O/E requires a Standard Error (SE). Assume a regression model has been used to determine the PREDs, from which the Es are calculated. The SE associated with predicting a continuous variable is part of the standard output of regression packages. Let s_j = standard error for the jth observation (note: this is the standard error for the individual observation not the standard error for the expected value of the observed outcome – regression packages typically provide both). Then SE_i for the average of the n_i cases treated by provider i is $\sqrt{\Sigma_j s_j^2} / \sqrt{n_i}$ (where j is summed from 1 to n_i). For a dichotomous outcome, SE_i is $\sqrt{\Sigma_j p_j * (1 - p_j)} / \sqrt{n_i}$, where p_j is the predicted probability for the jth case.

The above approach for estimating the SE has a good theoretical justification when the size of random variation around a predicted value is pretty much the same for all observations. However, for healthcare costs, variation is usually substantially higher among cases with higher predicted costs. If random variation is a constant percentage of cost, then log (cost) satisfies the 'constant variance' (homoscedastic) requirement. However, problems often arise when transforming a predicted log (cost) back to a predicted cost in the dollar scale. Also, just because variation increases with the prediction does not mean that it increases as a constant percentage of cost. In our experience, random variation is a higher percentage of predicted cost when predictions are low. We have found it useful to estimate random variation within bins of predicted values, for example: bin 1 = the lowest 20% of predicted, bin 2 = those with predicteds between the 20th percentile and the 30th percentile, bin 8 = those with predicteds between the 80th percentile and the 90th, bin 9 = those between the 90th and 95th percentile, bin 10 equals those between the 95th and 99th percentile, and bin 11 = those above the 99th percentile. Within each bin, we calculate the standard deviation of actual costs and then assign that SD to each person in the bin. When calculating the SE for a confidence interval, the assigned standard deviation is used rather than the value determined from the regression model.

To portray the results for provider i, one could show the interval $E_i \pm 2*SE_i$ and a point for O_i. If O_i falls in the interval, the provider is assumed to be practising as expected. Alternatively, one could show the point E_i and the 'acceptance interval' that

goes from $O_i - 2*SE_i$ to $O_i + 2*SE_i$. If E_i falls within these bounds, the provider is viewed as practising as expected. Another common practice is to divide the end points of the interval $O_i \pm 2*SE_i$ by E_i, resulting in an approximate 95% confidence interval for O/E. If this interval includes 1, the provider's practice is viewed as unexceptional.

Often profiles are presented by showing the intervals and points for a number of providers. The width of each interval is primarily a function of the number of cases treated. Providers with wider intervals treat fewer cases. In an attempt to simplify, sometimes profiles just show O/E ratios or (worse) just O and use a star to indicate if the O/E ratio is significantly different than 1 or if the O value is significantly different than expected. Unfortunately, this approach creates artificially large distinctions between providers whose observed experiences fall just inside versus just outside their intervals.

An alternative to comparing either O or O/E to some interval which reflects random fluctuations is to measure the difference between O and E in units of standard errors, that is, $z = (O-E) / SE$. For sufficiently large n, this quantity follows a standard normal probability distribution (which is why we call it 'z', the common designation for such a variable). If z has a standard normal distribution, it is easy to calculate the probability, called a p-value, that deviations from expected at least as large as what was observed are due to random chance. If this probability is small, the assumption that the provider practised 'as expected' is rejected. For example, a z-score of 2 corresponds to a p-value of about .05, a common cutpoint for identifying statistically significant findings. Flagging a provider as an outlier when the z-score is greater than 2 has the same theoretical justification as flagging them if their observed or expected falls outside the types of 95% intervals discussed above.

In the above, if a provider's actual outcomes are statistically different from expected (outside the confidence bounds), the provider is flagged as an outlier. Two main reasons (other than 'chance') can cause a provider to be flagged: one, the provider is particularly effective (or ineffective); or two, the provider is the victim (or beneficiary) of case mix differences not accounted for by the risk adjusters in the model. Flagging the provider implies that patient management is the cause. Below, we describe a modification of the above approach that typically widens the confidence bounds, reducing the chance that a provider is flagged as an outlier. In essence, it gives more weight to the presumption that observed differences in provider outcomes are due to unmeasured case mix differences and randomness rather than the effectiveness of patient management.

R^2 indicates the extent to which the independent variables in the model explain variations in the dependent variable. One can also examine the increase in R^2 associated with a particular independent variable or set of independent variables. For example, after including all risk adjusters in a model, how much higher is R^2 when dummy or indicator variables for providers are then added? Closely associated with this idea of 'incremental contribution' to R^2 is the partial F statistic. The F statistic for the entire regression model is used to test the null hypothesis that there is no relationship between the independent variables in the model and the dependent variable. The partial F statistic associated with the provider indicator variables can be used to test the null hypothesis that individual providers do not affect the outcome, after the model has accounted for other independent variables (such as differences in inherent patient risk).

The partial F statistic has a very close relationship to the intraclass correlation coefficient (ICC). Imagine randomly selecting a provider and then randomly selecting two of that provider's patients; the ICC is the correlation between their outcomes. A high

ICC means that knowing the outcome for one of the provider's patients provides information about likely outcomes for other patients of the same provider; that is, that 'provider matters'. If A (n) is the average number of patients treated by providers (calculated by averaging the number of patients treated by each provider), V(n) = variance of the number of patients treated by providers, and N = number of providers, then $n` = A(n) - V(n)/(N*A(n))$. If F_p is the partial F statistic associated with providers, then the ICC can be estimated as

$$(F_p - 1) / (F_p + n` - 1)$$

If outcomes of patients treated by the same provider are dependent, estimates based on a provider panel of n patients are less reliable than if outcomes were independent. The decreased reliability can be measured by the design effect (de), which for each provider with a panel size of n is calculated as 1 + (n–1)*ICC. Assuming that the dependency between outcomes is due to unmeasured case mix differences, the effective sample size for a provider is calculated as n/de. If the ICC is large, the effective sample size for a provider will be much smaller than n, confidence intervals will be wider, and hence fewer providers will be flagged as outliers.

4.3 Hierarchical models

There are several potential problems with the standard approaches to profiling discussed above. One relates to how the 'true' mean value of the outcome is estimated for each provider (i.e., M_i for provider i). Traditionally, each M_i is estimated by A_i, the average outcome of provider i's patients. However, especially for providers with small panels, A_i may not be the best predictor of what will happen to provider i's patients in the future. In a population with a 'historical' 2% problem rate, do we really think that a provider who had 1 problem among 10 patients has a 10% problem rate; or, if there were no problems among 10 patients, the provider has a true problem rate of 0%. Typically, the set of averages is too spread out, with the highest ones being higher than their 'true' values and the lowest ones being lower. In addition, the traditional approach to estimating SEs described above may underestimate the amount of variability that is present, leading to confidence intervals that are too narrow. One reason for this is that traditional methods recognise only one source of variation in the data – random variation of patients within providers. However, it seems reasonable to assume not only that patients vary randomly, but also that providers vary randomly. Provider variability will increase SEs. SEs also may be underestimated because when considering providers, such as hospitals or health plans, patients may cluster within provider, example, by physician within hospital or health plans. Clustered data may result in dependencies between outcomes within the same cluster. As noted above, when analysing units within which there are dependencies in outcomes, effective sample sizes (in terms of the amount of information provided) are less than actual sample sizes. Approaches that do not adjust for clustering may underestimate SEs (Greenfield *et al.*, 2002). Hierarchical models (also called multilevel or random effects models) provide a comprehensive approach for dealing with such problems. Greenland (2000), McNeil, Pedersen and Gatsonis (1992) and Shahian *et al.* (2001) provide non-technical descriptions of multilevel models. Normand, Glickman and Gatsonis (1997) is a good technical discussion in the context of provider profiling).

Imagine a provider whose rate of problems we seek to understand and predict. Under the traditional approach described earlier, we use data from that provider to calculate the percentage of problems. Then, the 95% confidence interval is used to help us understand the accuracy of that % as an estimate of the true provider problem rate. For example, having observed 1 problem out of ten, the estimated problem rate would be 10%, but we would be restrained from viewing this as evidence of a deviation from a population-wide average of 2% by the fact that the lower bound of the 95% confidence interval is just barely above 0%. Absent provider-specific information, the hierarchical modelling framework leads one to believe that each provider is like other providers (technically, this is called the assumption of exchangeability). As more and more provider-specific data become available, we allow these data to (gradually) modify our expectation. For example, having viewed 1 problem in 10, our 'best guess' for M_i would now be a little larger than 2%. The assumption that the data of other providers is an important source of information for predicting the performance of each individual provider is the core of the hierarchical modelling approach. If, for some reason, the assumption does not seem reasonable (at least for some subset of providers), then a hierarchical model may be inappropriate.

Using a hierarchical model, we essentially begin by estimating provider i's problem rate by the observed problem rate in the population of patients (what we earlier called A). Once some data have been collected on provider i, we can calculate A_i. Under common formulations of hierarchical models, provider i's 'true' problem rate is then estimated as a weighted average of the two estimates A_i and A. The weight assigned to the observed rate for provider i (i.e., A_i) becomes larger as the amount of data available on provider i increases. A traditional approach estimates the provider's true average as A_i, regardless of how much or little data are available for provider i; in contrast, the common hierarchical model estimate is a weighted average of the observed rate, which might be very extreme, and the rate for all patients, which is necessarily 'in the middle'. Hence, the hierarchical model estimate is less extreme. If an A_i pertained to a provider with very few cases, the estimate would be far less extreme. In the aggregate, the 'ensemble' set of provider-specific estimates from hierarchical models is less spread out than the A_is. It is in this sense that estimates from hierarchical models are said to 'shrink' traditional estimates. By shrinkage we do not mean that all estimates will be smaller, but that they will be pulled in from the extremes towards a central value.

Hierarchical models provide a comprehensive framework for incorporating variation at different levels of analysis. The 'hierarchy' derives from nesting, which occurs when data are not generated independently but in groups. For example, in some settings patients can be viewed as nested within primary care physicians; primary care physicians may be nested within practice groups (example, physicians who work out of the same clinic); and practice groups may be nested within region. At each level of the hierarchy, there may be different independent variables that one wants to take into account when profiling. Explicit modelling of the hierarchical structure recognises that nested observations may be correlated and that each level of the hierarchy can introduce a source of variation.

Hierarchical models can easily incorporate risk adjustment. Rather than shrinking the estimates of each provider to an overall average, they can be shrunk to the expected average for that provider based on the risk characteristics of patients in their panel. Thus, the estimate is a weighted average of the provider's observed average (O_i) and their

expected average (E_i). The weight associated with O_i depends on the size of the provider's panel.

In reanalysing CABG mortality data from the Pennsylvania Healthcare Cost Containment Council, Localio *et al.* (1997) used simulations to demonstrate 'the dramatic reduction in the number of false outliers with the use of hierarchical statistical models. The hierarchical models maintained adequate statistical power for detecting true departures from expected rates of mortality'.

Hierarchical models allow consideration of outcomes with more policy relevance than just mean outcomes. Normand, Glickman and Ryan (1996) illustrated this in a study profiling hospitals for the HCFA Cooperative Cardiovascular Project in the early 1990s. Outcome measures included: the probability that hospital-specific mortality for average patients was at least 50% greater than median mortality; and the probability that the difference between risk-adjusted mortality (calculated for each hospital using a logistic regression model fitted to the hospital's patients) and standardised mortality (predicted mortality based on a model developed from all patients) was large.

Hierarchical models provide an attractive framework for estimation when profiling providers. Shrunken estimates appropriately adjust for the influence of outliers and for the increased unreliability associated with estimates from smaller samples. Furthermore, the probability intervals from hierarchical models reflect the uncertainty associated with estimates better than traditional confidence intervals. Nevertheless, hierarchical models have generally not been used to profile provider performance outside of research settings. One reason is that they 'require substantial statistical sophistication to implement properly' (Shahian *et al.* (2001), p. 2162). Hierarchical models nonetheless offer substantial advantages, and over the next several years, easier methods for implementing these models will likely appear.

4.4 Comparing outcomes over time

The discussion so far has focussed on cross-sectional analyses, that is, information relating to a single time period. However, a very useful way in which to profile involves examining changes over time, that is, undertaking longitudinal analyses. As Berwick (1996, p.4), a leading healthcare quality improvement expert, observed, 'Pick a measure you are about, and begin to plot it regularly over time. Much good will follow'.

Plotting over time highlights changes. With only a few time periods, it is difficult to determine if changes reflect random variations versus real changes in behaviour. In particular, not much confidence can be placed in any changes occurring over just two time periods. However, as the number of time periods increase, persistent trends increasingly suggest that observations reflect an underlying reality as opposed to random variation. Further, when provider outcomes that had been stable over time change after a managerial or policy intervention, this lends credibility to the effectiveness of the intervention. Unfortunately, since time periods for profiling have to be large enough to have a reasonable sample (example, often a year), longitudinal data for many periods may not be available. And, even when they are, the earlier data may be too old to be relevant for understanding the current situation or predicting the future.

5 An application

5.1 Traditional analysis

To illustrate the techniques discussed above and the challenges of provider profiling, we consider a data set of 68,066 commercially insured individuals treated by 69 family practitioners. For each individual, we know their health plan expenditures (that is, what the insurance plan paid for their care) and their Relative Risk Score (based on the DxCG system). For ease of presentation, we will refer to health plan expenditures as costs.

As discussed, in healthcare cost data there are often a few very large outliers. In this data set, 113 individuals (0.2%) had costs over $50,000. The most expensive case was over $80,000. We began by topcoding costs at $50,000, that is, setting any cost over $50,000 to $50,000. In what follows, all 'cost' references are to topcoded costs.

There were large variations in the size of the physician panels, ranging from 44 individuals to over 1,800 individuals. The vast majority of the panels (over 90%) had over 100 individuals; 75% had over 500 individuals. Estimates from the larger panels are more reliable than those from the smaller panels.

Table 1 shows, for a subgroup of providers, panel size, observed average cost for the panel, the ratio of average panel cost to the average cost over all patients ($1846), and the coefficient of variation (the SD divided by the average). All providers whose average costs were more than 15% above or below the overall average are included in this table (in addition to some other providers, as described below). There were three providers (52, 69 and 54) whose average costs were more than 30% below the overall average and three providers with average costs more than 30% above the overall average. Note that the patient costs varied widely within provider panels; coefficients of variation often exceeded 2.

Table 1 Cost and severity information by provider

Provider Number	*Panel Size*	*Observed Cost*	*Cost/Aver Cost*	*Coefficient Variation*	*RRS/ AverRRS*	*Predicted Cost*	*Observed/ Expected*
Low cost providers							
52	311	1190	0.64	2.41	0.62	1286	0.93
69	82	1221	0.66	3.20	1.04	1909	0.64
54	644	1280	0.69	2.16	0.80	1550	0.83
9	1782	1367	0.74	2.43	0.84	1618	0.84
56	715	1401	0.76	2.85	0.77	1516	0.92
33	607	1406	0.76	2.25	0.84	1613	0.87
48	1750	1455	0.79	2.02	0.94	1758	0.83
32	996	1499	0.81	2.33	0.96	1794	0.84
66	44	1516	0.82	1.88	0.82	1580	0.96
61	92	1524	0.83	1.90	0.92	1724	0.88
7	1779	1563	0.85	2.57	0.98	1823	0.86
6	1831	1598	0.87	2.43	1.07	1950	0.82
29	1223	1607	0.87	2.87	0.98	1820	0.88

Table 1 Cost and severity information by provider (Continued)

Provider Number	*Panel Size*	*Observed Cost*	*Cost/Aver Cost*	*Coefficient Variation*	*RRS/ AverRRS*	*Predicted Cost*	*Observed/ Expected*
55	403	1667	0.90	2.12	1.20	2144	0.78
1	1709	1718	0.93	2.68	1.02	1877	0.92
22	1517	1784	0.97	2.23	1.07	1954	0.91
4	1719	1804	0.98	2.47	1.08	1964	0.92
12	1782	1835	0.99	2.14	1.20	2136	0.86
High cost providers							
19	1625	1805	0.98	2.24	0.86	1646	1.10
38	1454	1867	1.01	2.36	0.89	1682	1.11
16	1351	1891	1.02	2.34	0.91	1709	1.11
17	1131	2058	1.11	2.18	1.01	1863	1.10
24	1515	2061	1.12	2.24	1.03	1893	1.09
51	1108	2101	1.14	1.98	1.03	1895	1.11
45	456	2126	1.15	2.51	0.91	1714	1.24
2	1568	2146	1.16	2.26	1.17	2088	1.03
23	1092	2160	1.17	2.13	1.14	2052	1.05
28	622	2181	1.18	2.40	1.10	1995	1.09
37	1256	2185	1.18	1.97	1.11	2012	1.09
44	1706	2186	1.18	2.13	1.01	1865	1.17
42	916	2209	1.20	2.58	1.00	1847	1.20
15	692	2224	1.20	2.30	1.26	2225	1.00
68	38	2302	1.25	1.53	1.84	3073	0.75
26	1231	2351	1.27	2.12	0.97	1796	1.31
39	802	2380	1.29	1.92	0.99	1832	1.30
13	162	2441	1.32	2.42	1.08	1966	1.24
64	1290	2620	1.42	2.24	1.38	2397	1.09
31	720	2728	1.48	1.96	1.19	2128	1.28

The sixth column shows the ratio of the mean RRS of the panel to the mean RRS over all patients (which was 1.5). Though many of the providers had a case mix near average (that is, ratios near 1), some providers' panels differed greatly from average. For example, provider 52 had a low-risk group of patients, 62% of the average RRS. No doubt, this explains some of the provider's low cost. Provider 68 had a particularly high-risk group of patients, 84% above average. This may explain some of provider 68's higher cost.

To develop a model to adjust for differences in the risk profile of patients in different panels, we ran a linear regression model with cost as the dependent variable and RRS as the independent variable. R^2 was 0.51; that is, over 50% of the variation in cost was accounted for by variation in RRS. We also ran a model that included RRS squared, age and age squared, and gender. Since R^2 increased only slightly, to 0.53, in what follows we use the model with just RRS as the independent variable.

Examining average cost by decile of predicted cost provides insight into the ability of RRS to differentiate groups with very different costs. In the 3 lowest deciles, average costs were \$215, \$358 and \$573 respectively. In the 3 highest deciles, average costs were \$2421, \$3542 and \$9609. Also, the 50% of people with the lowest predicted costs incurred only 9% of the expenses; while the 10% with the highest predicted cost incurred 48% of all expenses; and the top 5% spent 33%. Recall that we are 'predicting' costs from data that 'know' which illnesses were being treated to generate these costs. Prospective models – models that predict next year's costs (before knowing what illnesses will arise) – cannot isolate high cost groups as well.

Columns 7 and 8 of Table 1 show the average of the predicted costs (E) and the little observed cost versus expected cost (O/E) ratio for the patients in each panel. Sometimes, adjusting for patient risk dramatically affected whether a provider looked extreme. For example, provider 52 had actual costs that were 64% of the overall average. However, this provider saw a relatively healthy set of patients. When this risk was taken into account, provider 52's actual costs were only 7% below expected (O/E = 0.93). Provider 69 also had low actual costs compared to the average, 66% of the average. After accounting for the fact that this provider's patients were somewhat less healthy than average, these actual costs were still viewed as low, only 64% of expected. Risk adjustment mattered little for this provider. Comparing columns 4 and 8 shows that risk adjustment affected the perceived efficiency of some providers only a little, while for others it mattered a lot.

In our risk adjustment model (which includes only one independent variable), the SD associated with the predicted cost for an individual varied only slightly across individuals. For almost all people it was 2,980. Thus, the interval within which observed mean costs were expected to fall was 'predicted costs $\pm$ 1.96*2980/$\sqrt{n}$, where n is the provider's panel size. (In this calculation, we used 1.96 rather than 2, the approximate number of SEs above and below the mean used to calculate a 95% confidence interval). Table 2 shows average actual costs, average predicted costs and the cutpoints around predicted costs used to identify providers as outliers. In addition to including providers whose actual average costs were more than 15% from the overall average, both Tables 1 and 2 include all providers flagged because their actual cost was outside the cutpoints (indicated by * in the Table). Of the 10 providers whose costs were more than 15% below average, only 5 would have been flagged because their observed cost was outside the range expected after risk adjustment. Among the 13 providers whose costs were more than 15% above average, only 8 have actual costs higher than expected.

Table 2 Observed costs, predicted costs and cutpoints used to flag providers

Provider Number	*Panel Size*	*Observed Cost*	*Predicted Cost*	*Observed/ Expected*	*Lower Cutpoint*	*Upper Cutpoint*
69	82	1221	1909	0.64	1264*	2554
55	403	1667	2144	0.78	1853*	2435
6	1831	1598	1950	0.82	1814*	2087
54	644	1280	1550	0.83	1320*	1781
48	1750	1455	1758	0.83	1618	1898

Table 2 Observed costs, predicted costs and cutpoints used to flag providers (Continued)

Provider Number	*Panel Size*	*Observed Cost*	*Predicted Cost*	*Observed/ Expected*	*Lower Cutpoint*	*Upper Cutpoint*
32	996	1499	1794	0.84	1609*	1979
9	1782	1367	1618	0.84	1479*	1756
7	1779	1563	1823	0.86	1684*	1961
12	1782	1835	2136	0.86	1998*	2275
33	607	1406	1613	0.87	1376*	1850
29	1223	1607	1820	0.88	1653	1987
61	92	1524	1724	0.88	1115	2333
22	1517	1784	1954	0.91	1804*	2104
1	1709	1718	1877	0.92	1736*	2018
4	1719	1804	1964	0.92	1823*	2105
56	715	1401	1516	0.92	1298	1735
52	311	1190	1286	0.93	955	1617
66	44	1516	1580	0.96	699	2460
68	38	2302	3073	0.75	2125	4020
15	692	2224	2225	1.00	2003	2447
2	1568	2146	2088	1.03	1940	2235
23	1092	2160	2052	1.05	1875	2228
37	1256	2185	2012	1.09	1848	2177*
24	1515	2061	1893	1.09	1742	2043*
64	1290	2620	2397	1.09	2234	2559*
28	622	2181	1995	1.09	1760	2229
19	1625	1805	1646	1.10	1501	1791*
17	1131	2058	1863	1.10	1689	2036*
16	1351	1891	1709	1.11	1550	1868*
51	1108	2101	1895	1.11	1719	2070*
38	1454	1867	1682	1.11	1529	1835*
44	1706	2186	1865	1.17	1724	2007*
42	916	2209	1847	1.20	1654	2040*
45	456	2126	1714	1.24	1441	1988*
13	162	2441	1966	1.24	1507	2425*
31	720	2728	2128	1.28	1910	2346*
39	802	2380	1832	1.30	1625	2038*
26	1231	2351	1796	1.31	1629	1962*

Note: * indicates provider has been flagged as an outlier

Overall, 13 providers (19%) were flagged as low cost providers and 15 (22%) flagged as high cost providers. It is not reasonable or useful to flag over 40% of providers. Closer examination of some of the flags suggests a possible problem. Providers 22, 1 and 4 were

flagged as low cost providers despite the fact that their actual costs were less than 10% below expected; providers 37, 24, 64, 19, and 17 were flagged as high cost providers despite the fact that their costs were within 10% of average. This illustrates a very important point about provider profiling: if sample sizes are large, relatively small differences between observed and expected can be statistically significant. To compensate, an oversight group may wish to adopt a 'practical significance' in addition to a statistical significance standard – only initiating action when the difference between observed and expected is larger than a managerial or policy-relevant cutpoint. For example, if we were to specify that to be flagged, a provider's average costs had to be both statistically significant and more than 15% from average, then 7 providers would be flagged as low (provider 69, 55, 6, 54, 48, 32 and 9) and 7 as high (provider 44, 42, 45, 13, 31, 39 and 26). The downside is that this lessens the incentive for managers of large practices to push for continuous, incremental improvement.

Requiring that findings be statistically significant reduces the risk of falsely flagging small providers, where large deviations may just reflect random chance. However, it may also cause us to miss truly aberrant practices. For example, though the performance of provider 68 differs widely from expected, this provider is not flagged as an outlier.

As an alternative approach to showing confidence boundaries, one that perhaps more easily facilitates consideration of both practical and statistical significance without implication of guilt by flagging, we suggest examining the z-score, that is, (O – E) / SE. Table 3 is Table 2 with z-score added and sorted by z-score. z-scores can be easily converted into a p-value and then used in a traditional hypothesis testing framework. Thus, a z-score of approximately 2 is equivalent to flagging a provider if actual costs are outside the cutpoints defined above. However, ranking on a z-score scale provides some sense of those providers for whom the evidence is strongest that their actual costs do not equal their expected costs. Thus, rather than just a yes/no designation that a provider is outside the cutpoints, z-scores continuously measure the deviation of observed from expected and are easily combined with a managerial cutpoint. For example, one might say we will flag any provider with a z-score greater than 3 and an O/E ratio of less than 0.85 or more than 1.15. In this case, 5 low cost providers would be flagged and 7 high cost providers. This seems like a much more reasonable number of providers to flag.

Table 3 Providers sorted by z score

Provider Number	*Panel Size*	*Observed Cost*	*Predicted Cost*	*Observed/ Expected*	*Lower Cutpoint*	*Upper Cutpoint*	*z-score*
			Low cost providers				
6	1831	1598	1950	0.82	1814*	2087	–5.06
12	1782	1835	2136	0.86	1998*	2275	–4.27
48	1750	1455	1758	0.83	1618*	1898	–4.25
7	1779	1563	1823	0.86	1684*	1961	–3.68
9	1782	1367	1618	0.84	1479*	1756	–3.56
55	403	1667	2144	0.78	1853*	2435	–3.21
32	996	1499	1794	0.84	1609*	1979	–3.12
29	1223	1607	1820	0.88	1653*	1987	–2.49

Table 3 Providers sorted by z score (Continued)

Provider Number	*Panel Size*	*Observed Cost*	*Predicted Cost*	*Observed/ Expected*	*Lower Cutpoint*	*Upper Cutpoint*	*z-score*
54	644	1280	1550	0.83	1320*	1781	−2.31
4	1719	1804	1964	0.92	1823*	2105	−2.24
22	1517	1784	1954	0.91	1804*	2104	−2.23
1	1709	1718	1877	0.92	1736*	2018	−2.20
69	82	1221	1909	0.64	1264*	2554	−2.09
33	607	1406	1613	0.87	1376	1850	−1.71
56	715	1401	1516	0.92	1298	1735	−1.04
61	92	1524	1724	0.88	1115	2333	−0.64
52	311	1190	1286	0.93	955	1617	−0.57
66	44	1516	1580	0.96	699	2460	−0.14
			High cost providers				
68	38	2302	3073	0.75	2125	4020	−1.59
15	692	2224	2225	1.00	2003	2447	−0.01
2	1568	2146	2088	1.03	1940	2235	0.77
23	1092	2160	2052	1.05	1875	2228	1.20
28	622	2181	1995	1.09	1760	2229	1.56
13	162	2441	1966	1.24	1507	2425*	2.03
37	1256	2185	2012	1.09	1848	2177*	2.05
19	1625	1805	1646	1.10	1501	1791*	2.15
17	1131	2058	1863	1.10	1689	2036*	2.20
24	1515	2061	1893	1.09	1742	2043*	2.20
16	1351	1891	1709	1.11	1550	1868*	2.24
51	1108	2101	1895	1.11	1719	2070*	2.31
38	1454	1867	1682	1.11	1529	1835*	2.36
64	1290	2620	2397	1.09	2234	2559*	2.69
45	456	2126	1714	1.24	1441	1988*	2.95
42	916	2209	1847	1.20	1654	2040*	3.68
44	1706	2186	1865	1.17	1724	2007*	4.44
39	802	2380	1832	1.30	1625	2038*	5.22
31	720	2728	2128	1.28	1910	2346*	5.41
26	1231	2351	1796	1.31	1629	1962	6.54

Note: * indicates provider has been flagged as an outlier

Managers need to decide how many providers will be singled out for special treatment. Quantitative measures help rank the providers in terms of which one's panels provide the strongest evidence of substantial deviance. However, strong evidence and large deviance are different dimensions and only one can be used in a single ranking. It is silly to assume that just because an algorithm flags over 20% of providers as high, that it will be

worthwhile to intensively follow-up on all of them. External 'sanctions' should probably emphasize practical importance (large deviance), while internal managers with large panels should continuously push for small (statistically meaningful) improvements.

As discussed earlier, we usually determine the variation in a predicted panel average by assuming that the same variance applies for every case, even though random variations are usually larger for cases that are expected to be more expensive. In our data, for the 20% of cases with the lowest predicted costs, mean costs were $88 with a standard deviation of $298; in the upper 1% of cases in terms of predicted, mean costs were $23,422 with a standard deviation of $15,616. Thus, we recalculated confidence bounds using the binning approach described above. In most cases, the confidence intervals were wider, but usually by under $100 (for 65 of the 69 providers); as a result, two fewer providers (one high, one low) were flagged as having statistically significant costs.

It may well be that providers have little control over the proportion of their panel that actually visits them over some time period. This suggests doing the analysis including only patients with positive costs. When we did this (using the SE from the regression model rather than the binning approach just described), 12 providers were still flagged as statistically significant on the low side and 12 (3 fewer than before) on the high side. Most of the flagged providers were the same as before.

Given the many providers identified as having average costs statistically significantly different than expected, one might think that provider is an important variable explaining differences in costs. However, once RRS is in the model, adding 68 dummy variables for provider only leads to an increase in R^2 of less than .01. The partial F statistic for provider is 5.3 and the intraclass correlation coefficient is .004. Adjustment for the design effect would have little impact on the width of the confidence intervals in this situation.

5.2 Hierarchical modelling (HM)

In the above analysis, the only source of randomness is due to variation in individual patient costs around each provider's mean. As noted, hierarchical modelling allows one to formally incorporate the possibility that providers' mean costs may also vary from their expected due to random factors. In our analysis, we used a relatively simple hierarchical model:

- providers' true mean costs vary normally around their expected costs with some unknown standard deviation (which is the same for all providers)
- the SE associated with the mean costs for a provider is the SD of that provider's patients' costs divided by the square root of the panel size.

The output of the HM analysis provides two quantities of direct interest:

- an estimate of each provider's true mean cost, which is a weighted average of their expected costs and their actual costs
- bounds within which one is 95% sure true mean costs will fall.

Table 4 compares the HM estimates to the traditional approach estimates. The first thing to notice is the extent to which shrinkage toward expected depends on the size of a

provider's panel. Provider 6 had a large panel, over 1800 members. The provider's observed average cost was $1598, the expected $1950, and the HM-estimated cost $1685. The HM estimate was shrunk 25% of the way from the actual to the expected estimate. Provider 66 had a particularly small panel, only 44 members. For that provider, the HM- estimated cost was shrunk over 87% of the way toward from the actual to the expected.

Table 4 Comparison of traditional cutpoints and cutpoints from a hierarchical model (HM)

Provider Number	*Panel Size*	*Observed Cost*	*Predicted Cost*	*HM Cost*	*Traditional Lower Cutpoint*	*Traditional Upper Cutpoint*	*HM Lower Cutpoint*	*HM Upper Cutpoint*
				Low cost provider				
6	1831	1598	1950	1685	1814*	2087	1523*	1841
12	1782	1835	2136	1912	1998*	2275	1748*	2069
48	1750	1455	1758	1505	1618*	1898	1377*	1632
7	1779	1563	1823	1631	1684*	1961	1464*	1793
9	1782	1367	1618	1418	1479*	1756	1277*	1559
55	403	1667	2144	1927	1853*	2435	1679	2160
32	996	1499	1794	1595	1609*	1979	1413*	1775
29	1223	1607	1820	1694	1653*	1987	1488	1896
54	644	1280	1550	1366	1320*	1781	1183*	1546
4	1719	1804	1964	1853	1823*	2105	1676	2028
22	1517	1784	1954	1833	1804*	2104	1663	2000
1	1709	1718	1877	1771	1736*	2018	1594	1950
69	82	1221	1909	1824	1264*	2554	1510	2121
33	607	1406	1613	1486	1376	1850	1287	1681
56	715	1401	1516	1456	1298	1735	1240	1668
61	92	1524	1724	1678	1115	2333	1391	1957
52	311	1190	1286	1239	955	1617	1016	1461
66	44	1516	1580	1572	699	2460	1275	1873
68	38	2302	3073	3015	2125	4020	2697	3323
15	692	2224	2225	2223	2003	2447	1984	2470
2	1568	2146	2088	2123	1940	2235	1932	2315
23	1092	2160	2052	2114	1875	2228	1908	2323
28	622	2181	1995	2063	1760	2229	1812	2316
13	162	2441	1966	2017	1507	2425*	1716	2327
37	1256	2185	2012	2121	1848	2177*	1930	2315
19	1625	1805	1646	1760	1501	1791*	1592	1929
17	1131	2058	1863	1978	1689	2036*	1777	2185
24	1515	2061	1893	2001	1742	2043*	1813	2189
16	1351	1891	1709	1825	1550	1868*	1634	2017

Table 4 Comparison of traditional cutpoints and cutpoints from a hierarchical model (HM) (Continued)

Provider Number	*Panel Size*	*Observed Cost*	*Predicted Cost*	*HM Cost*	*Traditional Lower Cutpoint*	*Traditional Upper Cutpoint*	*HM Lower Cutpoint*	*HM Upper Cutpoint*
51	1108	2101	1895	2023	1719	2070*	1827	2224
38	1454	1867	1682	1803	1529	1835*	1617	1989
64	1290	2620	2397	2506	2234	2559*	2282	2737
45	456	2126	1714	1836	1441	1988*	1572	2115
42	916	2209	1847	1999	1654	2040*	1758	2249
44	1706	2186	1865	2078	1724	2007*	1896	2267*
39	802	2380	1832	2104	1625	2038*	1875	2347*
31	720	2728	2128	2365	1910	2346*	2119	2633
26	1231	2351	1796	2104	1629	1962*	1890	2331*

Note: * Indicates provider has been flagged as an outlier

The HM intervals show the range within which one is 95% sure the provider's true mean cost lies. The interval is approximately centred at the HM-estimated true cost. Consider provider 6. We are pretty sure the provider's true costs were somewhere between $1,523 and $1,841. Based on the profile of the provider's patients, we would have predicted costs of $1,950. Since 1,950 is outside the range 1,523 to 1,841, we are pretty sure this provider's true cost were below predicted. Hence, the provider is flagged as a low cost provider. On the high side, consider provider 26. We are pretty sure this provider's true costs were between $1,890 and $2,331. This is higher than the provider's predicted cost, $1,796. Hence, this provider is flagged as a high cost provider.

As is apparent in Table 4, 7 providers were flagged as low cost providers and 3 as high cost providers. Overall, 14% of the providers were flagged, compared with over 40% using the traditional approach. Fourteen percent seems like a more reasonable number of outlier providers. Formally incorporating random variation among providers in the analysis protects against the tendency of the traditional approach to flag 'too many' providers.

6 Past, present and future

Throughout the 1980s, the primary task in risk adjustment was to develop and test models. Today's researchers, policy analysts and managers can choose from several well-vetted risk adjustment models. In this chapter, we have discussed how to use existing models to create credible predicted values for each observation in a population. In healthcare, of course, individual outcomes are highly variable, and individual predictions are, at best, only 'correct on average'. We have discussed how to examine the accuracy of predictions, both to help in choosing among models and in understanding the strengths, limitations and performance characteristics of a model that is being used.

Risk adjustment becomes policy-relevant when we compare average outcomes (Os) with predicted or expected average outcomes (Es) within important subgroups and seek to interpret the meaning of discrepancies. We have discussed several plausible ways for

doing this. Although we have focussed on provider profiling, our discussion pertains to any attempt to understand the effect of a patient, provider or system characteristic on a health outcome. To learn where the system is, and how to improve it, we must compare Os to Es and act on what we see. The challenge is to neither over-react to the pseudo-precision that makes random noise look like an important 'finding,' nor to be paralysed into inaction by the certainty of uncertainty.

In 2004, the newspapers are filled with stories that rely on credible risk adjustment. The Centers for Medicare and Medicaid Services (CMS) makes risk adjusted payments to HMOs for the Medicare beneficiaries they enroll; the 2003 Medicare drug bill mandates risk adjustment to deal with predictable differences in the expected cost of drugs for individuals; consortia of purchasers rank hospitals, imposing surcharges on consumers for using facilities in 'inefficient' tiers; medical managers of provider groups produce, examine and supply profiling data to individual doctors and medical teams, seeking to identify and promulgate better modes of practice.

Though state-of-the-art risk adjustment models predicting continuous outcomes, such as costs, have R^2 over 0.50 (when used concurrently) and models predicting a dichotomous outcomes, such as death, have c statistics approaching 0.90, much variation in outcomes remains unexplained. The most important concern is the potential for systematic bias. When a provider's actual outcomes differ from expected, we can quantify the likely contributions of randomness, but it can never be entirely clear how much is due to unmeasured risk and how much to differences in provider efficiency or quality of care. Even after expensive and time-consuming medical record reviews, this key question may remain unanswered. Significantly worse outcomes than expected need to be understood; however, they are not synonymous with poor quality or low productivity.

Traditional approaches for identifying outlier providers that measure random variation only at the patient level, especially those that use standard cutpoints (such as $z = 2$) to identify outliers, may flag too many outliers. In our example, despite our attempts to create stable findings, about 40% of the providers were flagged as outliers. Managers should consider both the practical and statistical significance of deviant outcomes; one way is to focus on providers with the most extreme z-scores whose O/E ratio exceeds a managerially relevant cutpoint. This guards against flagging either providers with extreme O/E ratios and small panel sizes (their z-scores will be small) or providers with large z-scores but O/E ratios close to 1 (they won't exceed the cutpoints).

Hierarchical models that formally incorporate variation among providers will be increasingly used in provider profiling. As we have seen, profiling based on such models flags fewer providers as outliers. Given the current status of risk adjustment models and the possibility that deviant O/E ratios are due to unmeasured case mix differences, cautious 'flagging' seems appropriate. Though hierarchical models are harder to implement, improvements in software and greater understanding of their value will likely lead to increased use. Finally, as longitudinal data become more available, longitudinal profiles should enable more reliable distinctions in the quality and efficiency of the care received from various providers.

Productivity and quality improvements in healthcare delivery require changing provider behaviour and credible provider profiles are needed for improvement programmes. However, as we have shown, analytical and conceptual challenges to developing useful profiles remain.

References

Ash, A.S., Zhao, Y., Ellis, R.P. and Kramer, M.S. (2001) 'Finding future high-cost cases: comparing prior cost versus diagnosis-based methods', *Health Services Research*, Vol. 36, pp.194–206.

Berwick, D.M. (1996) 'This year of 'how': New systems for delivering health care', *Quality Connections*, Vol. 5, pp.1–4.

Buntin, M.B. and Zaslavsky, A.M. (2004) 'Too much ado about two-part models and transformations? Comparing methods of modeling health care costs', Forthcoming in *Journal of Health Economics*.

Clancey, C. (2003) 'Quality improvement: Getting to how', *Health Services Research*, Vol. 38, pp.509.

Cumming, R.B., Knutson, D., Cameron, B.A. and Derrick, B. (2002) *A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations*, Society of Actuaries (soa.org/sections/riskadjfinalreport1.pdf)

Duan, N. (1983) 'Smearing estimate: A nonparametric retransformation method', *Journal of the American Statistical Association*, Vol. 78, pp.605–610.

Greenfield, S., Kaplan, S.H., Kahn, R., Ninomiya, J. and Griffith, J.L. (2002) 'Profiling care provided by different groups of physicians: Effects of patient case-mix (bias) and physician-level clustering on quality assessment results', *Annals of Internal Medicine*, Vol. 136, pp.111–121.

Greenland, S. (2000) 'Principles of multilevel modeling', *International Journal of Epidemiology*, Vol. 29, pp.158–167.

Hannan, E.l., Kilburn, H., Racz, M., Shields, E. and Chassin, M.R. (1997) 'Improving the outcomes of coronary artery bypass surgery in New York state', *Journal of the American Medical Association*, Vol. 271, pp.761–766.

Iezzoni, L.I. (Ed.) (2003) *Risk Adjustment for Measuring Health Care Outcomes, Third Edition*, Health Administration Press, Chicago.

Khuri, S.F., Daley, J., Henderson, W., Hur, K., Gibbs, J.O., Barbour, G., *et al.* (1997) 'Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: Results of the national veterans affairs surgical risk study', *Journal of the American College of Surgeons*, Vol. 185, pp.315–327.

Knaus, W.A., Wagner, D.P., Draper, E.A., Zimmerman, J.E., Bergner, M., Bastos, P.G., *et al.* (1991) 'The APACHE III Prognostic System: risk prediction of hospital mortality for critically ill hospitalized adults', *Chest*, Vol. 100, pp.1619–1636.

Localio, A.R., Hamory, B.H., Fisher, A.C. and TenHave, T.R. (1997) 'The public release of hospital and physician mortality data in Pennsylvania: a case study', *Medical Care*, Vol. 35, pp.272–286.

Luft, H.S. and Hunt, S.S. (1986) 'Evaluating individual hospital quality through outcome statistics', *Journal of the American Medical Association*, Vol. 255, pp.2780–2784.

McNeil, B.J., Pedersen, S.H. and Gatsonis, C. (1992) 'Current issues in profiling quality of care', *Inquiry*, Vol. 29, pp.298–307.

Normand, S.L.T., Glickman, M.E. and Ryan, T.J. (1996) 'Modeling mortality rates for elderly heart attack patients: Profiling hospitals in the cooperative cardiovascular project', *Case Studies in Bayesian Statistics*, Vol. 3, Springer, New York.

Normand, S.N.L.T., Glickman, M.E. and Gatsonis, C. (1997), 'Statistical methods for profiling providers of medical care: issues and applications', *Journal of the American Statistical Association*, Vol. 92, pp.803–814.

Available online at www.sciencedirect.com
SCIENCE DIRECT
Drug and Alcohol Dependence 82 (2006) 231–237
DRUG and ALCOHOL DEPENDENCE
www.elsevier.com/locate/drugalcdep

The use of the mini-mental state examination in recruitment for substance abuse research studies

Kristofer L. Smith [a,b], Nicholas J. Horton [c], Richard Saitz [a,b,d], Jeffrey H. Samet [a,b,e,*]

[a] *Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02118-2393, USA*
[b] *Department of Medicine, Boston Medical Center, Boston, MA 02118, USA*
[c] *Department of Mathematics, Smith College, Northampton, MA 01063, USA*
[d] *Department of Epidemiology and the Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA 02118, USA*
[e] *Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA 02118, USA*

Received 10 June 2005; received in revised form 29 September 2005; accepted 29 September 2005

Abstract

Background: Substance abuse is associated with cognitive impairment. Participation in clinical addiction research can be cognitively demanding. Screening tools can identify cognitively impaired subjects. We examined the use of the mini-mental state examination (MMSE) as an entry criterion in three randomized controlled substance abuse clinical trials.

Methods: In each of the three studies, we calculated the proportion of subjects excluded due to MMSE scores (<21) suggestive of cognitive impairment. We estimated the potential impact on enrollment based on the number of excluded subjects. Separately, for two of the studies, we assessed the impact of cognitive function on participation in follow-up using multivariable logistic regression.

Results: Of all persons screened for enrollment, 1.6% (171/10,791) were ineligible based solely on a MMSE score of <21. We estimate that 119 of these 171 ineligible persons would have consented and enrolled. These 119 persons would have represented 9.3% of all enrolled subjects across these studies. For subjects in a study in an inpatient detoxification unit, a higher MMSE score was associated with higher odds (adjusted odds ratio 1.15, 95% CI 1.03–1.30) of completing at least one follow-up assessment. A similar impact on subject follow-up was not observed in a study of medical inpatients with unhealthy alcohol use (adjusted odds ratio 1.01, 95% CI 0.86–1.20).

Conclusion: Screening for cognitive impairment using the MMSE excludes a small, but substantial, number of persons from addiction research studies. Cognitive ability, as captured by the MMSE may impact follow-up. These data support cognitive screening of substance abuse research subjects.

Keywords: Alcoholism; Research design; Clinical trials/mt (methods); Neuropsychological tests; Cognitive impairment

1. Introduction

Substance abuse can significantly impair cognitive ability (Bolla et al., 2000; Duka et al., 2003; Grant, 1987; Loberg and Miller, 1986). These impairments include the areas of visuospatial abstracting, problem solving, planning and organization, new learning, cognitive flexibility and memory skills (Ardila et al., 1991; Glen et al., 1988; Hambridge, 1990; Waugh et al., 1989). Impaired cognitive function is associated with shorter participation in treatment, increased dropout rates, less use of aftercare services, and greater post-treatment unemployment (Donovan et al., 1984; Gregson and Taylor, 1977; Teichner et al., 2002; Walker et al., 1983). Patients with impaired cognitive function, it has been argued, are not as likely to benefit from treatment as their unimpaired peers as they cannot sustain attention necessary to integrate and incorporate new information (Aharonovich et al., 2003; Becker and Jaffe, 1984; Kupke and O'Brien, 1985; Leber et al., 1985; Teichner et al., 2001). The associations between treatment outcomes and cognitive impairment has led to calls for universal cognitive dysfunction screening of all substance abuse patients (Miller and Saucedo, 1983; O'Farrell and Langenbucher, 1985).

In addition to the impact of cognitive function on treatment outcomes, researchers have been increasingly concerned with the ability of impaired subjects to provide informed consent (Grisso et al., 1997; Oldham et al., 1999). In certain popu-

* Corresponding author. Tel.: +1 617 414 7288; 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.2005.09.012

lations cognitive ability is strongly associated with decisional capacity (Moser et al., 2002). Patients with sufficiently substantial deficits do not provide meaningful consent to participate in research (Carpenter et al., 2000). There is also the theoretical concern about the accuracy of information obtained from subjects with cognitive impairment. As a result researchers have been encouraged to include cognitive assessment tools in clinical investigations involving subjects with high probability of having cognitive dysfunction (National Bioethics Advisory Commission, 2001).

The recognition of the need for cognitive screening, however, has not led to agreement as to which tool best combines brevity with high sensitivity and specificity. While several highly specific and sensitive neuropsychological batteries for the detection and delineation of cognitive impairment exist, these approaches are costly and lengthy. Given the sizable number of potential subjects that require screening for many substance abuse clinical research studies, particularly studies involving screening of non-treatment-seeking samples, these batteries are not practical. In the substance abuse treatment setting, a brief cognitive screening device capable of identifying those in need of further evaluation has been the alternative strategy to lengthy and expensive evaluations (Gillen et al., 1991).

The mini-mental state examination (MMSE) is a brief cognitive test widely used in both the inpatient and outpatient medical setting to assess cognitive function (Folstein et al., 1975). Performance on the MMSE agrees closely with results of expert assessments of cognitive ability (Etchells et al., 1999). The actual impact of using the MMSE as a brief cognitive assessment tool in substance abuse research studies, however, has received limited attention. This paper details the impact of using the MMSE to exclude cognitively impaired subjects during screening for eligibility in three alcohol and drug abuse research studies. Separately, we examine the association between cognitive function as measured by the MMSE and follow-up.

2. Methods

We examined data from three randomized controlled alcohol and drug abuse clinical trials, the health evaluation and linkage to primary care (HELP) study; addressing the spectrum of alcohol problems (ASAP); and adherence to drugs for HIV, an experimental randomized enhancement (ADHERE). Common to all three studies, patient were excluded if they planned to leave the local area within 12 months, did not have contact persons to facilitate follow-up, were not fluent in English or Spanish or if they had inadequate mental status to provide a history or informed consent, as indicated by a MMSE score of less than 21. To enable a better assessment of the generalizability of these findings to other substance abuse research study populations, we provide below a brief description of the enrolled subjects from each of the studies.

2.1. Participants

2.1.1. HELP

This study evaluated the effectiveness of a novel multidisciplinary clinic for linking patients in a residential detoxification program to primary medical care (Samet et al., 2003). Patients undergoing inpatient detoxification from alcohol, heroin or cocaine who had no primary care physician were enrolled into this randomized controlled trial (Table 1). The intervention consisted of a clinical evaluation while in a detoxification unit, followed by facilitated referral to an off-site primary care clinic. Follow-up assessments were planned over 2 years at 6-month intervals.

2.1.2. ASAP

This study assessed the effectiveness of a brief intervention for hazardous, harmful, and dependent drinkers among hospitalized patients (Table 1). Among general medical inpatients screened, those drinking at-risk amounts were enrolled. Subjects in the intervention group saw a counselor and received an individualized brief motivational intervention tailored to their alcohol problem severity. Subjects were scheduled for interviews 3 and 12 months after enrollment.

2.1.3. ADHERE

This study assessed the effectiveness of a multi-faceted intervention designed to improve adherence to antiretroviral medications in HIV-infected patients with a history of alcohol problems (Table 1; Samet et al., 2005). The experimental group received multiple visits from a trained nurse, which involved a brief intervention about alcohol use and practical guidance about adherence to medications. Subjects were followed at 3, 6 and 12 months after randomization.

2.2. Measures

2.2.1. Mini-mental state examination (MMSE)

The MMSE enables caregivers to identify patients with cognitive impairment, but it does not provide any specific diagnostic information nor does it assess capacity to perform a specific task or to consent to research or treatment (Folstein et al., 1975). The questions are grouped into seven categories, each one representing a different aspect of cognitive function: orientation to time (5 points); orientation to place (5 points); registration of three words (3 points); attention and calculation (5 points); recall of three words (3 points); language (8 points); and visual construction (1 point) (Tombaugh and McIntyre, 1992). The maximum score is 30. Patients who score in the 24–30 range are considered to be unimpaired, 18–23 indicates mild cognitive impairment and <18 indicates severe cognitive impairment (Anthony et al., 1982; George et al., 1991).

We defined cognitive impairment as an MMSE score of <21, instead of the aforementioned score of 24, consistent with research suggesting that individuals with fewer years of formal education perform less well on the MMSE (Crum et al., 1993; Cummings, 1993) and characteristics of our study populations (Table 1). A large portion of our study subjects did not graduate from high school and many of our subjects were homeless, both indications that they might perform less well on the MMSE.

Table 1
Descriptive statistics of enrolled subjects for three substance abuse clinical trials

Characteristic	HELP ($n=470$)	ASAP ($n=341$)	ADHERE ($n=350$)
Sex %			
Male	76	71	79
Female	24	29	21
Mean age (years)	36	44	41
Race %			
Black	46	45	44
White	37	39	33
Other	11	7	23
Language %			
Spanish	8	5	18
Other	3	1	2
Professional job %	8	15	14
Unskilled job %	25	21	26
Education (≥9 years) %	95	94	87
Physical health (PCS[a]) median (range)	48.9 (14.1 – 74.8)	37.1 (17.8 – 61.9)	45.5 (16.2 – 70.9)
Mental health (MCS[b]) median (range)	28.6 (6.8 – 62.2)	40.2 (10.6 – 68.1)	39.4 (0.2 – 73.0)
Homelessness %	47[c]	25[d]	29[c]
Alcohol use[e] %	85	100	42
Drug use[f] %	83	46	46

[a] Physical Component Summary.
[b] Mental Component Summary.
[c] Homeless = ≥ 1 night in the past 6 months.
[d] Homeless = ≥ 1 night in the past 3 months.
[e] Alcohol use = any alcohol in the past 30 days.
[f] Drug use = cocaine, marijuana or heroin use in past 30 days.

2.2.2. *Proportion of research subjects excluded due to cognitive impairment*

In each study we calculated the proportion of subjects excluded from study participation due to cognitive impairment defined as MMSE score <21 (Fig. 1). In the screening process of all three studies, the MMSE was administered only after all other entry criteria were satisfied. This allowed for the identification of a cohort of subjects eligible for study participation based on

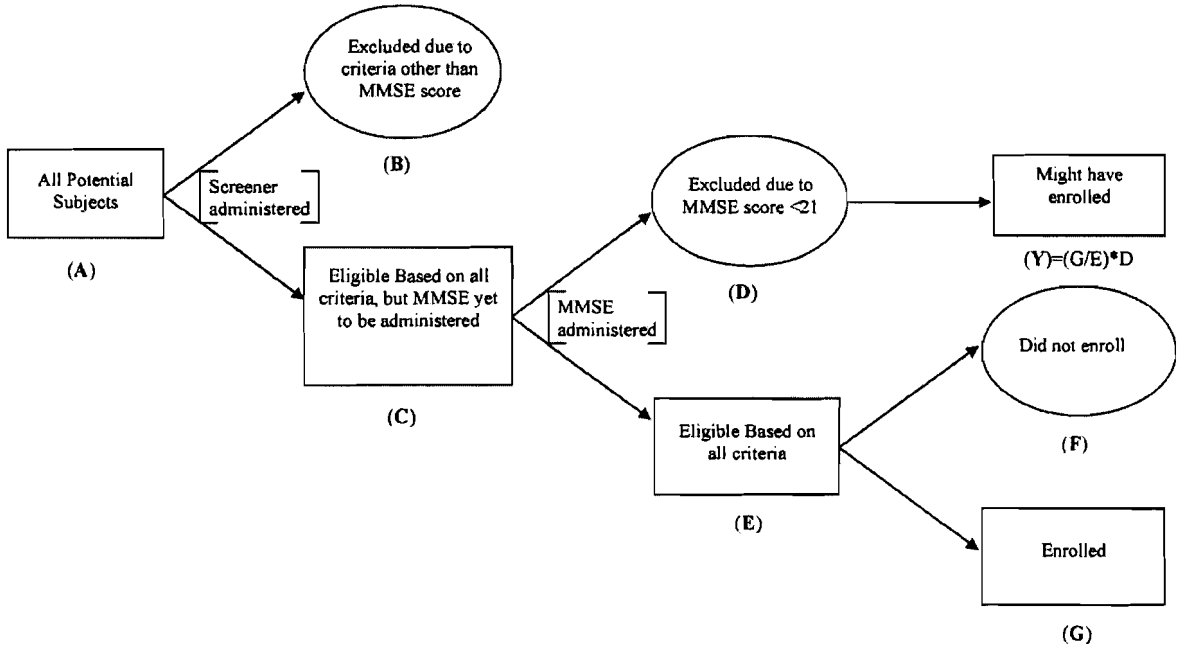


Fig. 1. Proportion of research subjects excluded due to cognitive impairment. This diagram depicts the progression of screened subjects from initial contact, through possible exclusion based on a number of criteria to enrollment. This diagram also illustrates the derivation of the estimate of cognitively impaired subjects who might have enrolled had the mini-mental state examination (MMSE) not been used (Y).

all other entry criteria except for cognitive impairment (Fig. 1D). With this group as the numerator (Fig. 1D) and the denominator being the total number of individuals screened (Fig. 1A), the proportion of subjects excluded only due to cognitive impairment was calculated (Fig. 1D/A).

We also calculated an estimate of the number of cognitively impaired subjects who would have participated in our studies had the MMSE score of <21 not been an exclusion criterion. For each study we calculated the observed enrollment percentage of eligible subjects; the number enrolled divided by the number eligible (Fig. 1G/E). The number of subjects excluded due only to MMSE <21 (Fig. 1D) was multiplied by this enrollment percentage (Fig. 1G/E) to identify a cohort of cognitively impaired individuals who might have enrolled (Fig. 1Y). These cognitively impaired individuals were then added to the number of subjects who did enroll (Fig. 1Y + G). The percentage of enrolled participants that would have had a MMSE score of <21 was calculated by dividing the number of cognitively impaired who might have enrolled (Fig. 1Y) by all estimated enrollees (Fig. 1Y + G).

2.2.3. Completion of-follow-up analysis

A multivariable logistic regression analysis of the association of MMSE score with the completion of a follow-up interview was completed for the HELP and ASAP studies. This regression was not possible in the ADHERE study, as almost half of the enrolled subjects also participated in an intervention study that altered the number and nature of the follow-up sessions. For all analyses, a two-sided $p < 0.05$ was considered to be statistically significant.

Follow-up was coded as a dichotomous variable. Those who participated in at least one follow-up encounter were classified as having follow-up. Those with no follow-up research interview were classified as lost to follow-up. The independent variable, MMSE score, was entered as a continuous variable from 21 to 30. Covariates in this analysis included gender, age, race/ethnicity, language, education level, occupation, randomization group and the physical component summary (PCS) and mental component summary (MCS). These covariates were chosen based on suspected clinical importance or their association with the outcome in bivariate analyses. Race/ethnicity was defined as white, black, Hispanic and other. Language was English, Spanish and other. Education level was defined as 0–8 or 9 years or greater completed. Occupation was defined using Hollingshead occupational categories (Hollingshead, 1976). Randomization group indicated whether or not the subject was assigned to receive the study intervention. The PCS and MCS are 100 point scales derived from the eight scales of the SF-36 (HELP Study) and the SF-12 (ASAP Study), generic multi-dimensional health status measures (Ware and Sherbourne, 1992).

3. Results

3.1. Proportion of research subjects excluded due to cognitive impairment

Across all three studies, 171 (58 for HELP, 86 for ASAP and 27 for ADHERE) of 10,791 screened persons were ineligible based only on a MMSE score of <21. Combining all three studies, the percentage of screened subjects excluded due only to a MMSE score of <21 was 1.7%, (2.8% for HELP, 1.1% for ASAP, 3.0% for ADHERE).

In the three studies a total of 1161 subjects enrolled (470 in HELP, 341 in ASAP and 350 in ADHERE). Overall, the percentage of eligible persons who enrolled was 70% (73% for HELP, 65% for ASAP and 79% for ADHERE). Based on these percentages of eligible subjects who enrolled, 119 of the 171 ineligible persons (42 for HELP, 56 for ASAP and 21 for ADHERE) might have consented and enrolled had the MMSE not been administered. These 119 subjects would have represented 9.3% of all enrolled subjects (8.2% for HELP, 14.1% for ASAP and 5.7% for ADHERE) (Fig. 1, Table 2).

3.2. Completion of follow-up analysis

In the HELP and ASAP studies, 85 and 90% of subjects completed at least one follow-up interview, respectively. In a multivariable logistic regression, cognitive function, as measured by the MMSE, was significantly and positively associated with completion of follow-up in the HELP study (adjusted odds ratio 1.15, 95% confidence interval 1.02–1.30) (Table 3). Additional significant variables included race/ethnicity and Spanish language. However, cognitive function was not significantly associated with completion of follow-up in the ASAP study

Table 2
Impact of the use of mini-mental state examination (MMSE) score as an exclusion criterion on enrollment in three substance abuse clinical trials

	HELP	ASAP	ADHERE
Screened individuals	2062	7824	905
Ineligible persons due only to MMSE <21 (% of screened individuals)	58 (2.8%)	86 (1.1%)	27 (3.0%)
Eligible persons	642	524	444
Enrolled persons (% of eligible)	470 (73%)	341 (65%)	350 (79%)
Number of cognitively impaired persons who might have enrolled had MMSE <21 not been used as exclusion criteria (% who might have been enrolled)[a].	42[b] (8.2%)[c]	56[b] (14.1%)[c]	21[b] (5.7%)[c]

[a] Calculation based on extrapolation from the actual percentages of eligible persons enrolled.

[b] Number of cognitively impaired persons who might have enrolled had MMSE <21 not been used as exclusion criteria = (% eligible patients who enrolled) × (no. of ineligible screeners due only to a MMSE <21).

[c] Percent of cognitively impaired persons who might have enrolled had MMSE <21 not been used as exclusion criteria = (no. of cognitively impaired persons who might have enrolled had MMSE <21 not been used as exclusion criteria)/(no. of cognitively impaired persons who might have enrolled had MMSE <21 not been used as exclusion criteria + no. of enrolled persons).

Table 3
A multivariable logistic regression analysis of the association of mini-mental state examination score with the completion of any follow-up interviews in two substance abuse clinical trials

	HELP study ($n=470$)	ASAP study ($n=330$)[a]
Characteristic	Adjusted odds ratio[b] (95% confidence interval)	Adjusted odds ratio[a] (95% confidence interval)
MMSE score (range 21–30)	1.15* (1.02–1.30)	1.01 (0.86–1.20)
Randomization to the intervention group	1.65 (0.95–2.90)	0.51 (0.23–1.14)
Age (years)	1.03 (0.99–1.07)	1.03 (0.99–1.07)
Race/ethnicity[c]		
White	0.25* (0.13–0.47)	0.66 (0.28–1.57)
Hispanic	0.12* (0.04–0.40)	1.31 (0.14–12.29)
Other	0.74 (0.17–3.27)	1.74 (0.21–14.7)
Language[d]		
Spanish	6.06* (1.39–26.40)	0.26 (0.02–3.19)
Other	0.31 (0.06–1.49)	>999.999 (N/A)
Professional job	0.96 (0.30–3.01)	0.52 (0.19–1.4)
Unskilled job	0.89 (0.46–1.70)	1.18 (0.4–3.45)
Education (≥9 years)	0.87 (0.21–3.63)	<0.001 (N/A)
Mental health-related quality of life	0.98 (0.95–1.00)	0.96* (0.93–0.99)
Physical health-related quality of life	0.99 (0.96–1.01)	1.00 (0.96–1.04)

MMSE: mini-mental state examination.
[a] n = all subjects available for analysis.
[b] Adjusted for all covariates listed in the table.
[c] Relative to African American.
[d] Relative to English.
* $p<0.05$.

(adjusted odds ratio 1.01, confidence interval 0.86–1.20) with the only significant variable being MCS (Table 3).

4. Discussion

By screening for cognitive impairment with the MMSE, a small but substantial number of people were excluded from participation in three large prospective studies that assessed over ten thousand people for eligibility. Among subjects who did enroll in these studies, cognitive impairment was associated with decreased follow-up in one trial.

The impact of using the MMSE to exclude subjects from clinical research studies has not been well described. Our data, however, demonstrate that for three considerably different alcohol and drug-affected populations – residential detoxification, medical inpatients, and HIV-infected adults – the percentage of screened patients excluded from study participation due to MMSE scores <21 is similar and substantial. These results suggest that among persons otherwise eligible to enroll in a wide range of substance abuse studies, screening will identify individuals with cognitive impairment.

An important consideration in using the MMSE as an exclusion criterion is the impact on the time it takes to fully recruit a study cohort. Our work demonstrates differential impact across the three studies. In the ASAP study, the research team might have enrolled 56 individuals excluded based on a low MMSE score; an increase of 14.1% in the number of study subjects. For the ADHERE study, the research team might have enrolled 21 additional individuals, an increase of only 5.7%. Therefore, using a low MMSE score as an exclusion criterion likely slowed reaching full enrollment more in ASAP than ADHERE.

Previous studies have noted that subjects lost to follow-up were younger, heavier drinkers and less educated (Edwards and Rollnick, 1997; Mackenzie et al., 1987). Recent studies have shown decreased treatment retention among more cognitively impaired substance abusers (Teichner et al., 2002; Aharonovich et al., 2003). In our analyses, higher MMSE score was significantly associated with completion of follow-up in one of the two studies examined. These differential findings may be due to the populations' characteristics, suggesting that the effect of cognitive function on follow-up will differ across studies. We found effects in a study of subjects with alcohol, heroin or cocaine dependence but without acute medical illness who had been enrolled during an residential detoxification stay. But in medical inpatients with unhealthy alcohol use, cognitive function did not impact follow-up. Our data demonstrate that for certain populations a brief cognitive function assessment tool can help identify patients more likely to drop out of research studies.

We also speculate that the MMSE could assist in screening research subjects to identify those likely to have impaired capacity to provide informed consent. The MMSE has been shown to reliably identify those without capacity to provide informed consent, though it has not been studied specifically in substance abuse study populations (Pucci et al., 2001). Until further research is completed, use of the MMSE to exclude cognitively impaired individuals may help to protect potential substance abuse study subjects at risk for making uniformed decisions regarding their research participation.

While the examination of the MMSE across three very different substance abuse research populations with a large number of both potential subjects screened and study participants is a strength of this investigation, there are several limitations. A

portion of the analysis rests on the assumption that impaired subjects would enroll in clinical research in similar proportions to unimpaired subjects. There are no published data to indicate that subjects with low MMSE scores would make different decisions regarding enrollment in research studies. It is conceivable, though that impaired subjects, particularly those with insight into their limitations, would choose not to enroll in cognitively demanding interventions, such as in our research trials. Given the lack of data around this issue, our findings could be considered an upper bound calculation of the impact on subject enrollment.

Another limitation of the analysis was that we studied only one approach to identifying cognitive impairment, the MMSE. The MMSE is a well-established and validated brief cognitive screening tool, but there are a number of other similar tools. It is possible that a different cognitive exam would more specifically identify types of patients who would be lost-to-follow-up or identify patients that do not provide meaningful informed consent (Grisso et al., 1997).

The follow-up analysis was limited in that subjects with moderate to severe cognitive impairment, MMSE <21, were excluded. As a result, we were not able to definitively show that follow-up would continue to decline for patients with MMSE scores of less than 21 in the HELP study. However, since the MMSE measures cognitive function on a continuum, it is likely that lower scores would also be associated with loss to follow-up. Another important question not answered by our study was whether people with cognitive dysfunction defined by MMSE score <21 are capable of providing informed consent. The conservative approach, however, with such vulnerable cohorts might be to exclude these patients from participating in human subjects research.

Finally, our study was only able to provide information on all individuals screened. Data for individuals who chose not to enroll, who were screened out at earlier stages and in particular who were excluded due to MMSE score might have yielded further significant findings.

Despite these limitations, our findings are consistent with previous substance abuse treatment studies (Aharonovich et al., 2003; Donovan et al., 1984; Gregson and Taylor, 1977; Teichner et al., 2002; Walker et al., 1983). Whether or not the MMSE should be used to exclude research subjects may depend upon whether one is performing an efficacy or effectiveness study (Tunis et al., 2003). Evaluation of a substance abuse intervention in which some small percentage of patients were cognitively impaired would not allow a true determination of the efficacy of an intervention requiring a certain level of cognitive processing. Including these cognitively impaired patients, however, might more accurately assess an intervention's effectiveness and generalizability to real-world substance abuse populations.

Our data demonstrate the feasibility and argue for the benefit of including a brief cognitive exam as a part of screening protocols for substance abuse research. Without a screening tool, a small minority of cognitively impaired subjects will be included in research studies with possible implications for follow-up and informed consent. While our study does not argue that the MMSE is the best tool for this purpose, our study indicates that it can identify a small cohort of impaired potential research subjects. Widespread use of MMSE scores as exclusion criteria might better guarantee research quality and better protect vulnerable patients from inappropriate study participation. Furthermore, without assessing cognitive status, clinical researchers lose an opportunity to identify subjects at risk for being lost to follow-up.

Acknowledgements

The authors appreciate the contributions of the clinical staff, project managers, and research associates on the HELP, ASAP and ADHERE studies. Data management was provided by both the Data Coordinating Center (DCC), School of Public Health, Boston University, Boston, MA (Christine Lloyd-Travaglini, MPH) and DM-STAT Inc., Medford, MA (Nicole Tibbetts, B.A.). We also acknowledge the contributions of Carly Bridden, MA, MPH to manuscript preparation.

This research was supported by the Boston University School of Medicine, Student Summer Research Program to Kris Smith, as well as the National Institute on Alcohol Abuse and Alcoholism (RO1-AA10870, RO1-AA11785, RO1-AA12617) and the National Institute on Drug Abuse (RO1-DA10019). Research activities were conducted in part in the General Clinical Research Center at Boston University School of Medicine, USPHS grant M01-RR00533.

References

Aharonovich, E., Nunes, E., Hasin, D., 2003. Cognitive impairment, retention and abstinence among cocaine abusers in cognitive-behavioral treatment. Drug Alcohol Depend. 71, 207–211.

Anthony, J.C., LeResche, L., Niaz, U., Von Korff, M.R., Folstein, M.F., 1982. Limits of the 'mini-mental state' as a screening test for dementia and delirium among hospital patients. Psychol. Med. 12, 397–408.

Ardila, A., Roselli, M., Strumwasser, S., 1991. Neuropsychological deficits in chronic cocaine abusers. Int. J. Neurosci. 57, 73–79.

Becker, J.T., Jaffe, J.H., 1984. Impaired memory for treatment relevant information in inpatient men alcoholics. J. Stud. Alcohol 45, 339–343.

Bolla, K.I., Funderburk, F.R., Cadet, J.L., 2000. Differential effects of cocaine and concaine alcohol on neurocogntive performance. Neurology 54, 2285–2292.

Carpenter, W.T., Gold, J.M., Lahti, A.C., Queern, C.A., Conley, R.R., Bartko, J.J., Kovnick, J., Appelbaum, P.S., 2000. Decisional capacity for informed consent in schizophrenia research. Arch. Gen. Psychiatry 57, 533–538.

Crum, R.M., Anthony, J.C., Bassett, S.S., Folstein, M.F., 1993. Population-based norms for the mini-mental state examination by age and educational level. JAMA 269, 2386–2391.

Cummings, J.L., 1993. Mini-mental state examination: norms, normals and numbers. JAMA 269, 2420–2421.

Donovan, D.M., Kivlahan, D.R., Walker, R.D., 1984. Clinical limitations of neuropsychological testing in predicting treatment outcome among alcoholics. Alcohol Clin. Exp. Res. 8, 470–475.

Duka, T., Townshend, J.M., Collier, K., Stephens, D.N., 2003. Impairment in cognitive function after multiple detoxifications in alcoholic inpatients. Alcohol Clin. Exp. Res. 27, 1563–1572.

Edwards, A.G.K., Rollnick, S., 1997. Outcome studies of brief alcohol intervention in general practice: the problem of lost subjects. Addiction 92, 1699–1704.

Etchells, E., Darzins, P., Silberfeld, M., Singer, P.A., McKenny, J., Naglie, G., Katz, M., Guyatt, G.H., Molloy, W., Strang, D., 1999. Assessment of patient capacity to consent for treatment. J. Gen. Intern. Med. 14, 27–34.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psych. Res. 12, 189–198.

George, L.K., Landerman, R., Blazer, D.G., Anthony, J.C., 1991. Cognitive impairment. In: Robins, L.N., Regier, D.A. (Eds.), Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. The Free Press, New York, pp. 291–327.

Gillen, R.W., Kranzler, H.R., Kadden, R.M., Weidenman, M.A., 1991. Utility of a brief cognitive screening instrument in substance abuse patients: initial investigation. J. Subst. Abuse Treat. 8, 247–251.

Glen, S.W., Parsons, O.A., Sinha, R., Stevens, L., 1988. The effects of repeated withdrawals from alcohol on the memory of male and female alcoholics. Alcohol Alcohol. 23, 337–342.

Grant, I., 1987. Alcohol and the brain: neuropsychological correlates. J. Consult. Clin. Psychol. 55, 310–324.

Gregson, R.A.M., Taylor, G.M., 1977. Prediction of relapse in men alcoholics. J. Stud. Alcohol 38, 1749–1760.

Grisso, T., Appelbaum, P.S., Hill-Fotouhi, C., 1997. The MacCAT-T: a clinical tool to assess patients' capacities to make treatment decisions. Psychiatr. Serv. 48, 1415–1419.

Hambridge, D.M., 1990. Intellectual impairment in male alcoholics. Alcohol Alcohol. 25, 555–559.

Hollingshead, A.B., 1976. Hollingshead two factor index of social position, occupational categories. In: Guy, W. (Ed.), ECDEU Assessment Manual for Psychopharmacology. U.S. National Institute of Health, Psychopharmacology Research Branch, Rockville, pp. 516–520.

Kupke, T., O'Brien, W., 1985. Neuropsychological impairment and behavioral limitations exhibited within an alcohol treatment program. J. Clin. Exp. Neuropsychol. 7, 292–304.

Leber, W.R., Parsons, O.R., Nichols, N., 1985. Neuropsychological test results are related to ratings of men alcoholics' therapeutic progress: a replicated study. J. Stud. Alcohol. 46, 116–121.

Loberg, T., Miller, W.R., 1986. Personality, cognitive and neuropsychological correlates of harmful alcohol consumption: a cross-national comparison of clinical samples. Ann. N.Y. Acad. Sci. 472, 75–97.

Mackenzie, A., Funderburk, F.R., Allen, R.P., Stefan, R.L., 1987. The characteristics of alcoholics frequently lost to follow-up. J. Stud. Alcohol 48, 119–123.

Miller, W.R., Saucedo, C.F., 1983. Assessment of neuropsychological impairment and brain damage in problem drinkers. In: Golden, C.J., Moses, J.A., Coffman, J.A., Miller, W.R., Strider, F.D. (Eds.), Clinical Neuropsychology: Interface with Neurologic and Psychiatric Disorders. Grune and Stratton, New York, pp. 141–195.

Moser, D.J., Schultz, S.K., Arndt, S., Benjamin, M.L., Fleming, F.W., Brems, C.S., Paulsen, J.S., Appelbaum, P.S., Andreasen, N.C., 2002. Capacity to provide informed consent for participation in schizophrenia and HIV research. Am. J. Psychiatry 159, 1201–1207.

National Bioethics Advisory Commission, 2001. Ethical and Policy Issues in Research Involving Human Participants, Volume I: Report and Recommendations of the National Bioethics Advisory Commission, 2001 (Full Report). Accessed at http://www.georgetown.edu/research/nrcbl/nbac/human/overvol1.pdf, on August 10, 2005.

O'Farrell, T.T., Langenbucher, J.W., 1985. Alcohol abuse. In: Hersen, M. (Ed.), Practice of Inpatient Behavior Therapy: A Clinical Guide. Grune and Stratton, Orlando, pp. 141–172.

Oldham, J.M., Haimowitz, S., Delano, S.J., 1999. Protection of persons with mental disorders from research risk: a response to the report of the National Bioethics Advisory Commission. Arch. Gen. Psychiatry 56, 688–693.

Pucci, E., Belardinelli, N., Borsetti, G., Rodriguez, D., Signorino, M., 2001. Information and competency for consent to pharmacologic clinical trials in Alzheimer disease: an empirical analysis in patients and family caregivers. Alzheimer Dis. Assoc. Disord. 15, 146–154.

Samet, J.H., Larson, M.J., Horton, N.J., Doyle, K., Winter, M., Saitz, R., 2003. Linking alcohol- and drug-dependent adults to primary medical care: a randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. Addiction 98, 509–516.

Samet, J.H., Horton, N.J., Meli, S., Dukes, K., Tripps, T., Sullivan, L., Freedberg, K.A., 2005. A randomized controlled trail to enhance antiretroviral therapy adherence in patients with a history of alcohol problems. Antivir. Ther. 10, 83–93.

Teichner, G., Horner, M.D., Harvey, R.T., 2001. Neuropsychological predictors of the attainment of treatment objectives in substance abuse patients. Int. J. Neurosci. 106, 253–263.

Teichner, G., Horner, M.D., Roitzsch, J.C., Herron, J., Thevos, A., 2002. Substance abuse treatment outcomes for cognitively impaired and intact outpatients. Addict. Behav. 27, 751–763.

Tombaugh, T.N., McIntyre, N.J., 1992. The mini-mental state examination: a comprehensive review. J. Am. Geriatr. Soc. 40, 922–935.

Tunis, S.R., Stryer, D.B., Clancy, C.M., 2003. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 290, 1624–1632.

Walker, R.D., Donovan, D.M., Kivlahan, D.R., O'Leary, M.R., 1983. Length of stay, neuropsychological performance and aftercare: influences on treatment outcome. J. Consult. Clin. Psychol. 51, 900–911.

Ware, J.J., Sherbourne, C.D., 1992. The MOS 36-item short-form health survey (SF-36). Conceptual framework and item selection. Med. Care 30, 473–483.

Waugh, M., Jackson, M., Fox, G.A., Hawke, S.H., Tuck, R.R., 1989. Effect of social drinking on neuropsychological performance. Br. J. Addict. 84, 659–667.

Journal of Consulting and Clinical Psychology
2006, Vol. 74, No. 6, 1162–1172

0022-006X/06/$12.00 DOI: 10.1037/0022-006X.74.6.1162

Evaluating Nicotine Replacement Therapy and Stage-Based Therapies in a Population-Based Effectiveness Trial

Wayne F. Velicer
University of Rhode Island

Robert H. Friedman
Boston University Medical Center

Joseph L. Fava
University of Rhode Island

Suzy B. Gulliver
Veterans Affairs Medical Center, Boston

Stefan Keller
University of Hawaii

Xiaowu Sun
University of Rhode Island

Harley Ramelson
Boston University Medical Center

James O. Prochaska
University of Rhode Island

Pharmacological interventions for smoking cessation are typically evaluated using volunteer samples (efficacy trials) but should also be evaluated in population-based trials (effectiveness trials). Nicotine replacement therapy (NRT) alone and in combination with behavioral interventions was evaluated on a population of smokers from a New England Veterans Affairs Medical Center. Telephone interviews were completed with 3,239 smokers, and 2,054 agreed to participate (64%). Participants were randomly assigned to one of four conditions: stage-matched manuals (MAN); NRT plus manuals (NRT + MAN); expert system plus NRT and manuals (EXP + NRT + MAN); and automated counseling plus NRT, manuals, and expert system (TEL + EXP + NRT + MAN). Assessments were completed at baseline, 10, 20, and 30 months. The point prevalence cessation rates at final follow-up (30 months) were MAN, 20.3%; NRT + MAN, 19.3%; EXP + NRT + MAN, 17.6%; and TEL + EXP + NRT + MAN, 19.9%. Stage-matched manuals provided cessation rates comparable with previous studies. The addition of NRT, expert system interventions, and automated telephone counseling failed to produce a further increase in intervention effectiveness.

Keywords: effectiveness trial, expert system, nicotine replacement therapy, smoking cessation, telecommunications

Of the people alive in the world today, 500,000,000 are predicted to die from the use of tobacco, with an average loss of 10 years of life (Peto & Lopez, 1990). Consequently, 5 billion years of human life will be lost because of one behavior. A breakthrough in developing an intervention with even a modest impact on populations of smokers could prevent millions of premature deaths and billions of lost years of life.

Wayne F. Velicer, Joseph L. Fava, Xiaowu Sun, and James O. Prochaska, Cancer Prevention Research Center, University of Rhode Island; Robert H. Friedman and Harley Ramelson, Medical Information Systems Unit, Boston University Medical Center; Suzy B. Gulliver, Outpatient Clinic, Veterans Affairs Medical Center, Boston; Stefan Keller, Department of Public Health Sciences and Epidemiology, University of Hawaii.

Portions of this article were presented at the annual meeting of the Society for Research on Nicotine and Tobacco, Arlington, Virginia, February 2000, and at the International Conference for Behavioral Medicine, Brisbane, Australia, November 2000. This work was supported by National Cancer Institute Grant CA71356. We thank John Hughes, University of Vermont, for his comments on a draft of this article.

Correspondence concerning this article should be addressed to Wayne F. Velicer, Cancer Prevention Research Center, 2 Chaffee Road, University of Rhode Island, Kingston, RI 02881. E-mail: velicer@uri.edu

Transdermal nicotine replacement therapy (NRT or the "patch") has been one of the most widely studied and most frequently used smoking cessation interventions. Multiple trials have supported the efficacy of this intervention (Fiore, Smith, Jorenby, & Baker, 1994; Hughes, Shiffman, Callas, & Zhang, 2003; Po, 1993; Silagy, Mant, Fowler, & Lodge, 1994; Tang, Law, & Wald, 1994). However, the vast majority of clinical trials to date have been efficacy trials (i.e., studies that involve highly select volunteer samples). A recent epidemiological study found no evidence supporting NRT in reducing smoking rates at the population level (Pierce & Gilpin, 2002). Other general population studies have not supported this finding (Alberg et al., 2004; Hasford, Fagerstrom, & Haustein, 2003; Miller et al., 2005).

To promote general implementation of an intervention and resolve these conflicting results, an effectiveness trial can provide critical empirical support. A particular strength of effectiveness trials is the attempt to recruit a representative sample from a defined population, allowing an inference of the generalizability of the results to similar populations. The aim of this study was to perform an effectiveness trial of NRT in combination with three low-cost behavioral therapies (manuals, tailored expert system interventions, and an automated counseling intervention). Proactive recruitment was used to recruit a large proportion of a defined

population of smokers. We also conducted an extended follow-up to assess the sustainability of the intervention.

The first goal of the study was to recruit a large proportion (i.e., > 75%) of the identified smokers in a defined population. The second goal was to recruit a representative sample of the population of eligible smokers on the basis of (a) demographic variables, (b) smoking history, and (c) stage of change. The third goal was to prepare the maximum proportion of the population for NRT use through application of the behavioral interventions. In a general population, a large proportion will not be ready to quit smoking (Velicer et al., 1995) and, therefore, not likely to use the NRT intervention. We anticipated that more than 40% of the sample would receive NRT. The fourth goal was to maximize retention at final follow-up (30 months), which we defined as more than 70% of the sample. A fifth goal was to evaluate the differential effectiveness of four combinations of behavioral and pharmacological interventions.

The study was an additive (i.e., each condition adds a component to the previous condition) four-group design with four assessments (baseline, Month 10, Month 20, and Month 30). The first intervention was the comparison condition used of stage-matched manuals (MAN) only. The second intervention condition (NRT + MAN) added NRT to the stage-matched manuals. The third intervention condition (EXP + NRT + MAN) added expert system tailored print reports (EXP) to MAN and NRT. The fourth intervention condition (TEL + EXP + NRT + MAN) added automated telephone counseling (TEL) to MAN, NRT, and EXP. The telecommunications system was developed for this study.

Effectiveness Versus Efficacy

Clinical trials are sometimes classified as efficacy trials or effectiveness trials (Flay, 1986; Glasgow, Lichtenstein, & Marcus, 2003). The efficacy trial relies on a volunteer sample that is randomly assigned to an intervention condition. In contrast, the effectiveness trial attempts to recruit a large representative proportion of a target population, which is similarly randomly assigned to an intervention condition. One of the advantages of an effectiveness trial is that the impact of an intervention (Velicer & DiClemente, 1993; Velicer & Prochaska, 1999) can be estimated. Population impacts of programs are defined as the recruitment rate times the efficacy rate. Producing high impacts begins with recruiting large percentages of eligible populations. For example, if 80% of the population can be recruited for an intervention with an efficacy rate of .10, the impact will be four times larger than an intervention with an efficacy rate of .40 that recruits 5%.

Recruitment and Representativeness

The most common approach for recruitment in efficacy trials has been a reactive approach, that is, subjects are informed about the availability of an intervention program and must initiate contact to participate. This volunteer sample is typically highly motivated to quit smoking and likely to adhere to the treatment protocol. Volunteer samples are also more likely to be female, White, and well educated (Prochaska, Velicer, Fava, Rossi, & Tsoh, 2001). In contrast, effectiveness trials typically rely on a proactive recruitment approach, that is, the subjects are contacted directly and the services are offered to them. The samples should reflect the general population. Two recent smoking cessation effectiveness studies (Prochaska, Velicer, Fava, Ruggiero, et al., 2001; Velicer, Prochaska, Fava, Laforge, & Rossi, 1999) achieved recruitment rates of 82% and 85%, and the samples were demographically similar to the defined population. In contrast, efficacy studies typically recruit 1% to 5% of the population at best (Schmid, Jeffrey, & Hellerstedt, 1989).

Generalizability

Beyond demonstrating the potential impact of an intervention, effectiveness trials increase confidence in the generalizability of the results (Prochaska & Velicer, 2004). Interventions are often less efficacious in effectiveness trials than in efficacy trials. Several reasons might explain these results: (a) Effectiveness trials evaluate treatments in the settings where they will commonly be applied, whereas efficacy trials employ optimal conditions; (b) implementation in a real-world setting must employ available personnel rather than personnel hired especially for the study; (c) some part of the intervention costs may have to be borne by the participants; (d) the intervention may be appropriate only for a small proportion of the population; and (e) effectiveness trials have a lower level of control over the timing of the intervention.

Method

Procedure

As a first step of proactive recruitment, approximately 33,962 letters were sent to potential recruits who were listed as members of a large northeastern U.S. Veterans Affairs Medical Center (VAMC). The letter introduced the study as a collaboration between the University of Rhode Island and the VA and informed the potential recruits about an upcoming telephone survey. Informed consent materials for the phone survey were included in the letter. VA members could return a postcard (postage prepaid) to decline to be contacted for the phone survey. A total of 5,022 returned the refusal card (14.8%). (This was an unusually high number compared with other studies and was attributed to the fact that many VA members had alternative health care providers that were viewed as their primary provider.) Approximately 2 weeks later, all members who did not decline participation (passive consent) were screened for study eligibility via a telephone survey. A total of 4,369 could not be contacted because of nondeliverable mail, a nonworking phone number, currently residing out of the country, or were deceased. Seventy-five were duplicate subjects. We eliminated 2,011 members for health or language issues. A total of 1,429 could not be contacted in 15 attempts (answering machines, not home, etc.), and the attempt to contact was terminated. A total of 2,664 were in the calling queue when recruitment for the study was terminated. Of the 18,392 potential subjects, 3,332 refused to participate in the phone survey when contacted (22.1%).

The screening survey was completed on a total sample of 15,060. Screening continued until the total sample size required for the study was recruited (N = 2,000). Any spouses of VAMC members who smoked were also recruited. The eligibility criteria included self-identification as a smoker who regularly smoked 10 or more cigarettes per day and, therefore, met the requirements for using NRT. Subjects in the action or maintenance stages were excluded.

After completing the survey, all eligible smokers were randomized by computer-based random number generator to one of four intervention conditions. The four intervention conditions were (a) MAN (Velicer, Rossi, Ruggiero, & Prochaska, 1994), (b) NRT + MAN, (c) EXP + NRT + MAN (Velicer et al., 1993), and (d) TEL + EXP + NRT + MAN (Friedman, 1998; Friedman et al., 1996; Ramelson, Friedman, & Ockene,

1999). Subjects were blinded to their treatment condition until they received the first intervention material; thus, awareness of the treatment condition could not influence the readiness for study participation. However, subjects were aware that several of the possible treatment conditions included NRT and that up to four follow-up assessments by telephone were scheduled over the following 30 months. All subjects were assessed at baseline, Month 10, Month 20, and Month 30. The survey center staff was blind to treatment condition. Two groups (EXP + NRT + MAN and TEL + EXP + NRT + MAN) received a limited additional assessment at Month 6 only on those variables needed to generate the expert system progress report.

Recruitment

A total of 3,239 smokers were identified as eligible and completed the full assessment during the telephone survey. Of this group, 324 subjects (10%) declined any further participation in the study after completing the initial phone survey; they are defined as the "survey-only" group. Written informed consent materials were mailed to the 2,915 subjects (90%) who provided verbal informed consent during the telephone survey. Up to 15 telephone contacts were made to participants who did not return the signed informed consent within a 2-month period. Overall, 861 subjects (29.5%) failed to return their written informed consent after having given verbal consent during the telephone interview ("verbal-consent" group). The remaining 2,054 smokers (63.4% of all eligible subjects) who sent back their written consent form constitute the "full-consent" group (see Figure 1). All information for the study was completely confidential.

Measures

The baseline assessment was completed on all subjects in all four groups and included sociodemographic variables, smoking history, and history of NRT use. Also included were the variables of the Transtheoretical Model (TTM) for smoking cessation: stages of change, 10 processes of change (Prochaska, Velicer, DiClemente, & Fava, 1988), pros and cons or decisional balance (Velicer, DiClemente, Prochaska, & Brandenberg, 1985), and situational temptations (Velicer, DiClemente, Rossi, & Prochaska, 1990). These measures were used to generate the expert system progress reports in the EXP + NRT + MAN and TEL + EXP + NRT + MAN groups. The stage variable was needed for the manuals in all four groups.

Three primary outcome measures (24-hr point prevalence, 7-day point prevalence, and 6-month prolonged abstinence) were assessed on the last three occasions (Months 10, 20, and 30). Because self-report is extremely accurate in a low-demand study like this, no biochemical validation was performed (Benowitz et al., 2002; Glasgow et al., 1993; Patrick et al., 1994).

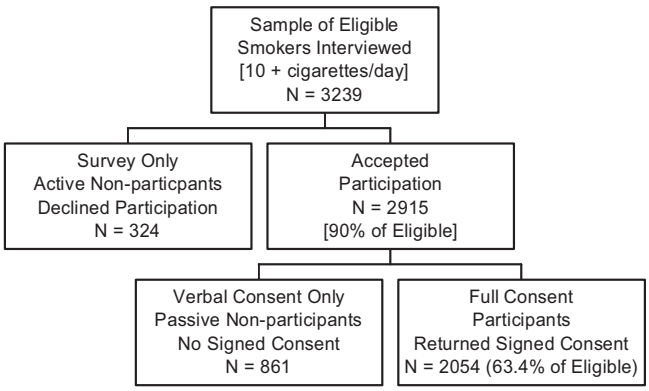


Figure 1. Flowchart for recruitment of eligible smokers classified as one of two nonparticipant groups (survey only and verbal consent only) or the participant group (full consent).

NRT Readiness

The design of the study was to provide NRT only to smokers judged to be ready for use in the immediate future. The decision to provide NRT was based on the smokers' stage of change and their decisional balance at baseline, Month 6, and Month 10 assessments. We projected that the baseline distribution would be 40% in precontemplation, 40% in contemplation, and 20% in preparation. The smokers in preparation and smokers in contemplation who had more pros of quitting than cons of smoking were immediately provided with NRT. On the basis of previous studies, we estimated that 25% would receive NRT at baseline. The remaining smokers were provided with intervention materials designed to promote stage movement and, therefore, eligibility for NRT. We estimated that 10% would receive NRT after the 6-month assessment and an additional 5% after the 10-month assessment, for a total of 40% of the smokers in the three NRT-eligible groups.

Interventions

MAN. The MAN group received the stage-based self-help manuals (Velicer et al., 1994) following baseline contact. The manuals inform users about their particular stage of change and the processes they can use to progress to the next stage. On the basis of their baseline assessment scores, treatment participants were sent the manual matched to their current stage of change and the stage beyond their current stage. Each smoker in each different stage at baseline received the same package of materials; the only difference was the number of manuals received on each occasion. This was viewed as representing a minimal intervention condition. This group served as a comparison group for the other three groups.

NRT. The NRT group received the stage-based self-help manuals following baseline contact. Those subjects for whom NRT was appropriate (preparation stage or contemplation stage with pros > cons) also received NRT. Six months after the initial assessment, the subjects who received only the manuals were recontacted. If they had progressed, NRT was provided. At the Month 10 assessment, subjects who had not received NRT were reevaluated, and those who had progressed received NRT. Subjects who received NRT at one of the early assessments but had relapsed received a second NRT intervention. The patch used was the 16-hr/15-mg patch (Nicotrol) with a 6-week course of treatment. The company provided the NRT replacement therapy at cost.

EXP. The third group (EXP + NRT + MAN) received the stage-based manuals, one expert system feedback report, and NRT when indicated. The expert system report provided feedback on the basis of normative comparisons. The 14 variables of the TTM were assessed as part of the baseline interview, and the responses of each subject were compared with peers in the same stage who were successful in progressing to the next stage. Detailed descriptions of this intervention are available elsewhere (Velicer et al., 1993; Velicer & Prochaska, 1999; Velicer, Prochaska, & Redding, 2006). Participants for whom NRT was not appropriate at baseline were reassessed at Month 6 and again at Month 10. If they had progressed, they were provided with NRT at that time.

TEL. The fourth group (TEL + EXP + NRT + MAN) received the stage-based manuals, the expert system progress report at baseline, and NRT. In addition, they received regular telecommunications contacts via an automated counseling intervention. The interactive telecommunications system was developed for this study and employs a series of prerecorded voice files assembled in the form of a conversation that is tailored to the responses of the smoker. The telecommunications contacts served to both complete the assessment of progress on the 14 TTM variables and provide instant automated feedback. Material similar to that in the written paragraphs of the expert system progress reports was presented during the call

and reproduced verbally. To limit the length of the call to 15 to 20 min, two telecommunications calls were needed to cover the same material as a single progress report. The responses to the assessment questions were entered using the telephone keypad.

There were two different frequency rates for the telecommunication contacts. In each case, the smoker was free to initiate additional contacts at any time and for as many times as desired. However, if a participant did not call in a set period, proactive calls were initiated automatically. The *lean* contact rate was monthly and applied to those smokers who had not received NRT (precontemplators, contemplators with pros > cons). The *dense* contact rate applied to smokers who received NRT and was weekly for the first month, biweekly the second month, and monthly for Months 3–6. The two rates of contact were used because we have observed that early-stage smokers change more slowly and are more likely to develop reactance against the intervention when pressed too hard. In each case, the rate automatically changed from lean to dense for any smoker who progressed and received NRT.

The total length of contact was designed to be 6 months. However, for those smokers who progressed to NRT status later in the study or at the Month 10 assessment, the telecommunications option was available for a 2-month period after receiving NRT.

Results

Sample

A total of 2,054 smokers participated in the study. Table 1 presents the demographic characteristics and stage of change for each group and the total sample. The average age was 50.5 years ($SD = 11.0$), and 77% were men. The mean education level was 13.0 years ($SD = 2.5$), 48.2% were currently married, and the sample was predominantly White (89.4%). Compared with the general population of smokers (Hughes, 2004), the sample included a larger proportion of White non-Hispanics (89.4% vs. 77.9%), men (77.0% vs. 52.6%), and was older (50.5 years vs. 40.8 years). For those who shared a household, 57.6% lived with another smoker. The stage distribution included fewer smokers in precontemplation and more smokers in preparation than other population-based samples (Velicer et al., 1995; Velicer, Prochaska, Fava, et al., 2006): Distributions for five U.S. samples are approximately 40% in precontemplation, 40% in contemplation, and 20% in preparation. As a randomization check, tests of significance ($p < .01$) were performed to determine whether there were any differences between the four groups. All tests were nonsignificant.

Table 1 also presents the comparison of the four groups on smoking history and NRT use. As a randomization check, tests of significance ($p < .01$) were performed to determine whether there were any differences between the four groups. All tests were nonsignificant. The average number of cigarettes smoked per day was 24.5, and the time to first daily cigarette was 35.3 min. The sample demonstrated significant quitting activity ($M = 2.5$ quit attempts during the past year), and 48.9% reported having previously used NRT, many in the past year (45.6%).

Velicer et al. (2005) provided a detailed comparison of the characteristics of the group that participated in this study and two groups of nonparticipants. The participants (full consent) differed significantly from both nonparticipant groups (survey only and verbal consent only). Participants were more likely to be married, young, female, living with others, and to have used NRT previously or considered using NRT. The survey-only group was more likely to be in precontemplation (54%), whereas the full-consent group was more likely to be in contemplation (46%) or preparation (35%). The recruitment procedure resulted in a sample that was not completely representative of the sample of smokers.

Retention

Attrition was classified as lost to follow-up or refused. The overall retention rate at the Month 30 assessment was 61% (1,249/2,054). There were no significant differences between the four groups (see Figure 2). The overall refusal rate was 8.1%.

Preparation for NRT

The proportion of the sample that received NRT on each of the three assessment occasions for each of the groups that were NRT eligible was higher than predicted. We initially estimated that 40% of the sample in the three conditions would receive the intervention. The proportion receiving NRT was 80% in NRT + MAN, 77% in EXP + NRT + MAN, and 79% in TEL + EXP + NRT + MAN. There was no significant difference between the three groups (see Figure 3). There were no adverse events reported.

Comparison of Treatment Conditions

Table 2 presents the results for three outcome measures (24-hr point prevalence, 7-day point prevalence, and 6-month prolonged abstinence) assessed on the final three occasions (Month 10, Month 20, and Month 30). The same pattern of results was observed for all three measures as would be expected given the extremely high correlation between the measures (Velicer & Prochaska, 2004). In this section, we focus on 24-hr point prevalence (see Figure 4) because it is the most sensitive outcome measure (Velicer, Prochaska, Rossi, & Snow, 1992).

The SAS PROC GENMOD (SAS Institute, 1997) procedure was used to perform the GEE analyses for the point prevalence outcome data. This analytic model included parameter estimates for the intercept, treatment effects (MAN, NRT + MAN, EXP + NRT + MAN, TEL + EXP + NRT + MAN), temporal effects at each follow-up assessment (Month 10, Month 20, and Month 30), and a term for the patterns of missing data ("missing"). The intention-to-treat analysis was conducted on the entire sample of 2,054 subjects identified as at risk for smoking and randomized to condition, including individuals with missing data for one or more of the follow-up time points. One parameter beyond the intercept was significant: time ($p < .01$). For the analysis of the time effects, the Month 10 assessment served as the referent because all respondents were smoking at baseline. The significant time effect indicates that there were small treatment effects over time between Month 10 and Month 20, $\chi^2(1) = 3.28$, $p < .10$, and big effects between Month 10 and Month 30, $\chi^2(1) = 20.21$, $p < .0001$. Overall smoking cessation rates increased from Month 10 (13.25%) to Month 20 (15.67%) and again to Month 30 (19.30%). Different patterns of missing data were modeled. The missing data parameter and the interactions of the missing data parameter and intervention parameters were not significant.

We used all available data at each assessment ("available"). Given the results of the GEE missing data analysis, this approach is appropriate for this study. Table 3 presents the 24-hr point

Table 1
Comparison of Four Intervention Groups on Demographic, Nicotine Replacement Therapy (NRT), and Smoking History Variables

	Intervention group														
	MAN (n = 523)			NRT + MAN (n = 522)			EXP + NRT + MAN (n = 509)			TEL + EXP + NRT + MAN (n = 500)			Total (N = 2,054)		
Variable	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*
Stage															
Precontemplation	19.12			17.62			19.65			19.00			18.84		
Contemplation	46.85			43.49			47.15			46.40			45.96		
Preparation	34.03			38.89			33.20			34.60			35.20		
Gender															
Men	78.78			77.59			75.64			75.80			76.97		
Race/ethnicity															
White	89.87			91.35			88.58			87.53			89.36		
Black	4.59			4.23			6.30			5.23			5.08		
Asian	0.00			0.00			0.39			0.20			0.15		
Native American	1.15			1.73			1.77			1.81			1.61		
Other	4.40			2.69			2.95			5.23			3.81		
Hispanic															
Yes	3.08			0.96			2.55			2.40			2.25		
Marital status															
Married	50.10			49.33			47.05			46.09			48.17		
Living with partner	9.56			11.95			10.04			9.62			10.30		
Not married	12.43			12.33			14.37			16.23			13.81		
Separated	5.93			5.39			5.51			4.21			5.27		
Divorced	18.16			18.11			18.70			20.64			18.89		
Widowed	3.82			2.89			4.33			3.21			3.56		
No. in household															
1	21.31			22.20			23.61			21.13			22.06		
≥2	78.69			77.80			76.39			78.87			77.94		
Other smoker in household															
No	44.01			43.18			44.42			38.11			42.44		
Yes	55.99			56.82			55.58			61.89			57.56		
Age (years)		50.52	10.55		50.89	10.11		50.6	10.62		49.93	10.7		50.49	10.5
Education (years)		12.98	2.72		12.98	2.36		13.16	2.49		13.1	2.42		13.05	2.5
Ever used NRT?															
Yes	50.19			51.44			46.17			47.60			48.88		
No	49.81			48.56			53.83			52.40			51.12		
Used for recommended time?															
Yes	42.86			40.84			38.70			39.15			40.45		
No	57.14			59.16			61.30			60.85			59.55		
Used NRT in past year?															
Yes	45.00			45.52			45.53			46.64			45.65		
No	55.00			54.48			54.47			53.36			54.35		
Ever considered using NRT?															
Yes	66.28			68.11			67.65			72.52			68.64		
No	33.72			31.89			32.35			27.48			31.36		
Quit attempts in past 3 months (n)		1.19	2.08		1.45	2.36		1.21	2.03		1.15	2.05		1.25	2.14
Quits in past 12 months (n)		2.46	3.01		2.75	3.16		2.49	3.03		2.25	2.87		2.49	3.02
No. cigarettes smoked per day		25.18	12.87		24.55	12.96		24.31	11.39		23.85	12.74		24.48	12.51
Length of last quit attempt (days)		356.32	693.13		323.71	654.7		335.93	649.72		323.24	618.73		334.94	654.68
Days without smoking in past year (n)		19.38	54.00		15.80	39.51		12.93	35.60		15.92	40.79		16.02	43.13
Time until first daily cigarette (minutes)		31.43	49.85		36.84	66.05		33.35	58.54		39.71	85.27		35.29	66.04

Note. MAN = stage-matched manuals; EXP = expert system intervention; TEL = telecommunications.

prevalence estimates for four different missing data mechanisms (complete case, available, expectation maximization [EM], and intention to treat). The method labeled "intention to treat" represents the widely employed ad hoc procedure in which the status of smoker is assigned for all missing observations. This procedure makes the unreasonable assumption that smoking status is the only reason that an observation is missing and leads to extreme distortions of the data when an extended follow-up is employed (Hall et

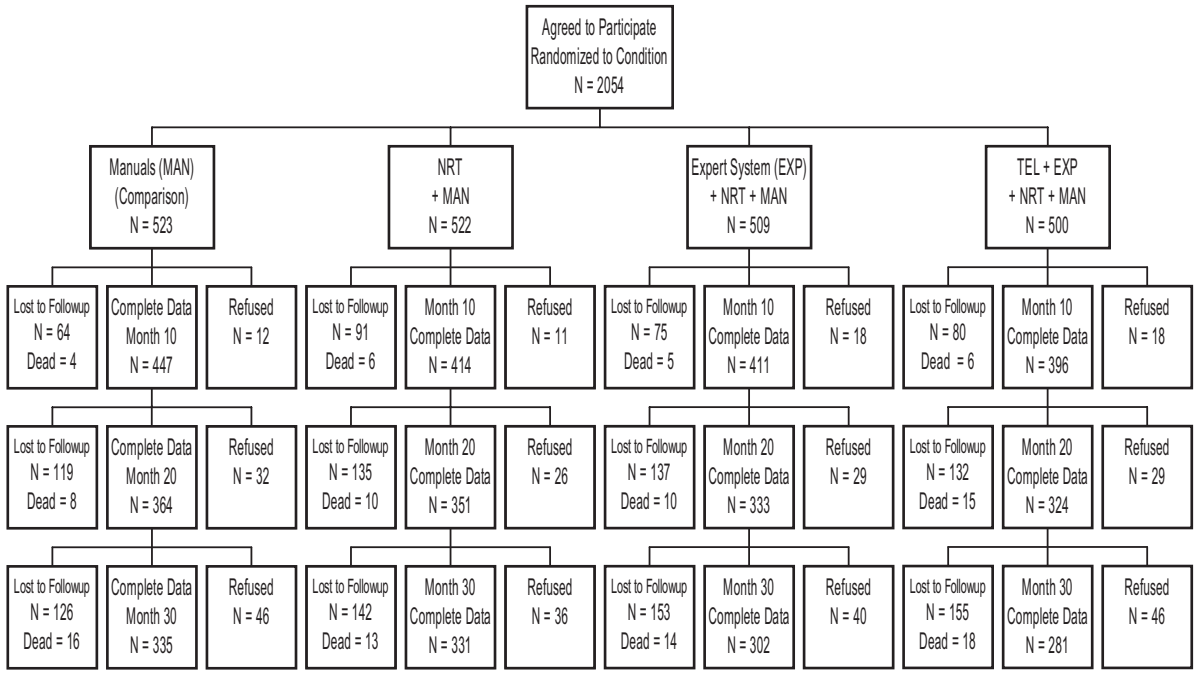


Figure 2. Retention rates for the four groups with dropouts classified as lost to contact or refusals. MAN = stage-matched manuals; NRT = nicotine replacement therapy; EXP = expert system intervention; TEL = telecommunications.

al., 2001). The complete case procedure (i.e., only participants whose data are available on each occasion) is another widely employed ad hoc procedure that is reasonable only when the observations can be assumed to be missing at random. The EM algorithm is one recommended statistical procedure for accurately estimating parameters in the presence of missing data and provides a maximum likelihood estimate of the parameters using all available data to estimate the missing values (Schafer & Graham, 2002). The same pattern of results was observed for all four methods.

Discussion

The overall results of the study demonstrate that a large percentage of a defined sample can be recruited to participate in a smoking cessation study. The overall recruitment rate (63.4%) was lower than comparable studies using the same method. The sample was less representative of the defined population than expected, with an underrepresentation of early-stage smokers. The retention rate of 61% at Month 30 was lower than expected. The proportion of the sample that received NRT was very large (80%), perhaps as a result of the failure to recruit early-stage smokers. All four interventions were equally effective, resulting in an almost 20% reduction in smoking at the final assessment. However, there were no differences between the four intervention conditions, and the absence of a control group limits our ability to conclude that any of the interventions were effective.

Manuals

Manuals are sometimes used to represent a minimal or no-intervention condition. In this case, stage-based manuals represent an active treatment condition. In previous research (Prochaska, DiClemente, Velicer, & Rossi, 1993), stage-matched manuals outperformed standard manuals (18.5% cessation rate at 24 months vs. 11%). A similar cessation rate (16.5%) was reported by Velicer et al. (1999). In this study, the MAN condition resulted in a slightly higher cessation rate than in previous studies (20.3% current vs. 18.5% in Prochaska et al., 1993, and 16.5% in Velicer et al., 1999). The higher rate may be due to the longer follow-up in this study (30 vs. 18 or 24 months).

NRT

The point prevalence cessation rate in the NRT condition produced cessation rates below the rates reported in efficacy studies. The most direct comparison is the 10-month rate because most previous efficacy studies fail to conduct an extended follow-up. The rate at 10 months was 11.4% compared with 19.5% at Month 6 in a meta-analysis. There are no comparable extended outcome data available for other NRT studies.

There are several potential explanations for the difference reported in efficacy trials and our current estimate and the results of over-the-counter (OTC or nonprescription) trials. First, only a proportion of our sample received NRT (80%), and all smokers receive NRT in efficacy studies. Second, there was no selectivity and only limited contacts; therefore, the compliance rate was much lower than expected in typical efficacy trials. Third, the VA sample represents a unique sample that may represent a more difficult challenge than the typical volunteer sample in an efficacy trial. For example, subjects were not screened for comorbid conditions, and the sample was older than most samples. However, a more appropriate comparison is the recent meta-analysis on seven

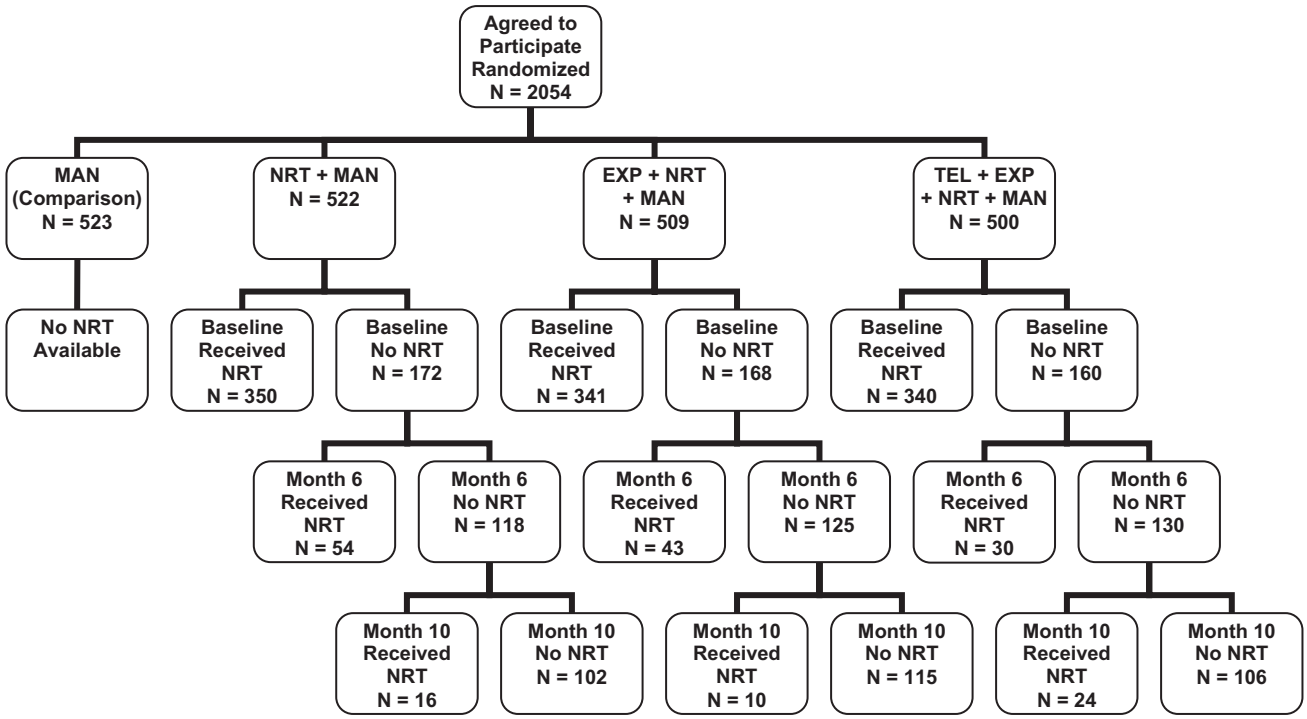


Figure 3. Proportion of each group receiving nicotine replacement therapy (NRT) on each assessment occasion. MAN = stage-matched manuals; EXP = expert system intervention; TEL = telecommunications.

studies of OTC NRT that reported a rate of 7% at 6-month follow-up (Hughes et al., 2003). The reported rate of 7.3% is very similar.

Expert System Intervention

The point prevalence cessation rate (17.6%) at Month 30 was below what would be expected on the basis of previous studies. The expert system intervention alone, typically combined with stage-tailored manuals, has resulted in cessation rates of 22–26% at 18 or 24 months across seven studies (Velicer, Prochaska, & Redding, 2006). The expectation was that the combination of NRT and EXP would outperform EXP alone by about 7%. Instead, the combination underperformed EXP alone from previous studies by about 7%.

There are several potential explanations. The expert system intervention used here departed from the usual three-report protocol, providing only a single expert system report. Velicer et al. (1999) compared three reports and a single report and found no significant difference. A second explanation is that the inclusion of

Table 2
Point Prevalence and Prolonged Abstinence Measures for the Four Intervention Groups at Months 10, 20, and 30

	Intervention group				
Assessment	MAN (n = 523)	NRT + MAN (n = 522)	EXP + NRT + MAN (n = 509)	TEL + EXP + NRT + MAN (n = 500)	Total (N = 2,054)
Month 10					
24-hr point prevalence	12.1	11.4	15.3	14.4	13.3
7-day point prevalence	6.7	7.3	9.7	11.1	8.6
6-month prolonged abstinence	3.6	4.1	5.4	6.6	4.9
Month 20					
24-hr point prevalence	15.1	12.8	17.7	17.3	15.7
7-day point prevalence	11.8	10.8	12.3	13.0	11.9
6-month prolonged abstinence	8.5	8.3	11.1	9.3	9.3
Month 30					
24-hr point prevalence	20.3	19.3	17.6	19.9	19.3
7-day point prevalence	14.9	15.1	15.2	15.0	15.1
6-month prolonged abstinence	12.5	10.0	13.6	15.0	12.7

Note. All data entries represent percentage of abstinent participants. MAN = stage-matched manuals; NRT = nicotine replacement therapy; EXP = expert system intervention; TEL = telecommunications.

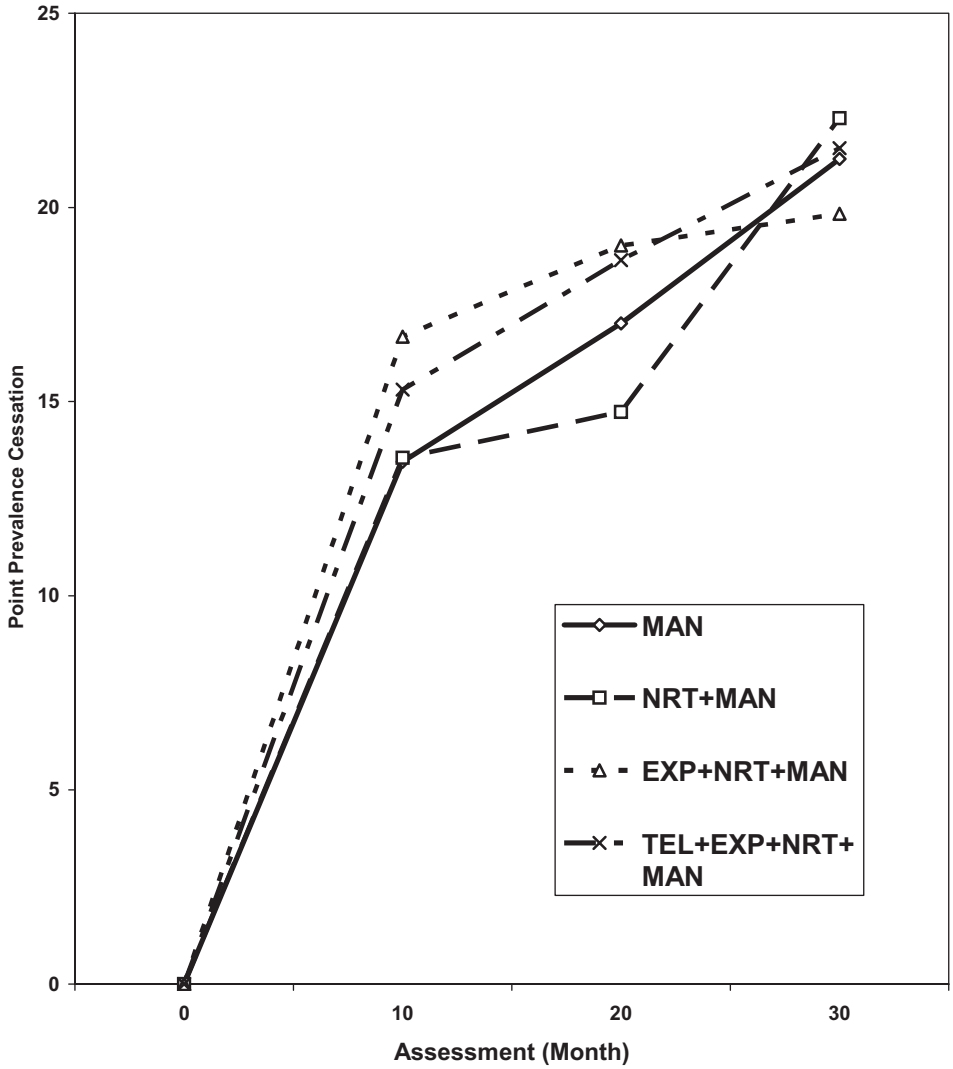


Figure 4. Point prevalence smoking cessation for the four intervention conditions at baseline, Month 10, Month 20, and Month 30. MAN = stage-matched manuals; NRT = nicotine replacement therapy; EXP = expert system intervention; TEL = telecommunications.

NRT resulted in overreliance on the pharmacological component, resulting in a failure to engage in the necessary behavior changes for successful cessation. This explanation gains some support from the fact that the expert system did not demonstrate the delayed treatment effects, that is, an increasing proportion quitting after treatment is completed, which have been observed with other stage-matched interventions. There was a slight increase from Month 10 (15.3%), the end of intervention, to Month 20 (17.7%), but no further increase occurred at Month 30 (17.6%). In contrast, the MAN condition went from 12.1% at Month 10 to 15.1% at Month 20 to 20.3% at Month 30. Previous reviews have reported that behavioral therapy and pharmacotherapy are likely to be more effective than pharmacotherapy alone (Hughes, 1995). A third possible explanation is that this was a unique sample that presents a more difficult challenge for the intervention.

Telecommunications

The inclusion of the telecommunication intervention represents a unique intervention that has not been evaluated previously in this context. The study failed to provide evidence of added value for this intervention. Telecommunication interventions have been effective with other behaviors, including adherence to medical procedures (Friedman et al., 1996), increasing exercise (King et al., 2003; Pinto et al., 2002), and improving diet (Delichatsios et al., 2001).

The effects of the telecommunication intervention are difficult to evaluate in the context of the other interventions. As with the EXP intervention, the inclusion of NRT could have resulted in an overreliance on the pharmacological intervention. Some qualitative utilization data indicate that subjects may have viewed the TEL intervention as an option rather than an expectation. The

Table 3
Point Prevalence Abstinence Estimates (24 hr) for Four Treatment Groups at Months 10, 20, and 30 for Four Missing Data Procedures

		Time					
		Month 10		Month 20		Month 30	
Group	Procedure	*n*	%	*n*	%	*n*	%
MAN	Complete case	297	12.79	297	16.50	297	19.87
	Available	447	12.08	364	15.11	335	20.30
	EM algorithm	523	13.45	523	17.01	523	21.25
	Intention to treat	523	10.33	523	10.52	523	13.00
NRT + MAN	Complete case	273	11.36	273	12.45	273	16.12
	Available	414	11.35	351	12.82	331	19.34
	EM algorithm	522	13.55	522	14.73	522	22.30
	Intention to treat	522	9.00	522	8.62	522	12.26
EXP + NRT + MAN	Complete case	250	16.40	250	18.80	250	19.20
	Available	411	15.33	333	17.72	302	17.55
	EM algorithm	509	16.67	509	19.02	509	19.84
	Intention to treat	509	12.38	509	11.59	509	10.41
TEL + EXP + NRT + MAN	Complete case	241	15.77	241	18.67	241	21.16
	Available	396	14.39	324	17.28	281	19.93
	EM algorithm	500	15.31	500	18.64	500	21.53
	Intention to treat	500	11.40	500	11.20	500	11.20

Note. Baseline predictors used were as follows: 24-hr quit in the past year, time to first daily cigarette, cigarettes per day in past week, gender, weight, education, age, and stage membership. MAN = stage-matched manuals; NRT = nicotine replacement therapy; EXP = expert system intervention; TEL = telecommunications; EM algorithm = expectation-maximization algorithm.

utilization rates were lower than in other telecommunications trials. Approximately 30% used the telecommunication on multiple occasions, 30% used it on a single occasion, and the remaining 40% did not use it at all. A study evaluating TEL alone for smoking is needed.

Limitations

The study has two important limitations. First, the inclusion of a pure control group would have served to exclude the interpretation that none of the interventions was effective and that the observed differences were the result of secular trends. The observed quit rate was far in excess of the 2–5% that has been observed in monitoring studies, but the population was unusual; therefore, this explanation cannot be confidently excluded. Second, the use of the VA sample represents both a strength and a weakness. The sample is clearly in need of services. The VA organization provided the opportunity to perform the proactive recruitment. However, it is an unusual sample with a much higher average age and higher rates of psychiatric problems and is predominantly male, which make generalization to other samples of smokers difficult.

Future Directions

In this effectiveness trial with a defined population, NRT was not more effective than stage-based manuals alone. Furthermore, NRT was not more effective than manuals alone, regardless of which behavioral interventions were combined with NRT.

We cannot determine the effects of NRT on EXP because EXP alone was not tested. In previous trials, adding counselors to EXP made no difference, even when the EXP alone was effective (Prochaska et al., 1993; Prochaska, Velicer, Fava, Ruggiero, et al., 2001). On the other hand, adding nicotine-fading computers (Lifesign) to EXP was significantly less effective than EXP alone (Prochaska, Velicer, Fava, Ruggiero, et al., 2001). We would predict that EXP + NRT < EXP alone.

It is tempting to conclude that NRT is not effective with populations of smokers. However, recent innovative applications of NRT have been designed to use NRT to reduce the number of cigarettes smoked in unmotivated smokers. These interventions have produced significantly more cessation than did a control condition (Carpenter, Hughes, Solomon, & Callas, 2004). This was true for NRT reduction counseling and motivational interviewing plus NRT. These results suggest that NRT may be effective with populations of smokers when used innovatively to reduce smoking rather than as historically applied to produce more immediate cessation.

Conclusions

The study had five primary goals. The first goal was to recruit a large proportion (75%) of the sample of identified smokers. The study recruited 63.4% of the sample of identified smokers, which is smaller than has been reported in other studies. However, the initial rate of 90% was higher than in other studies. The focus in the informed consent process on the potential use of NRT may explain the lower and highly selective recruitment rate (Velicer et al., 2005). Another potential explanation is the unique nature of the sample of VA members.

The second goal was to recruit a representative sample of the identified smokers. Several segments of the population were underrepresented in the final sample, particularly smokers in the precontemplation stage.

The third goal was to prepare early-stage smokers for NRT and deliver NRT to the maximum proportion of the sample possible. This goal was exceeded. The NRT utilization rate was projected to be 40%. The actual rate in the study was 80%. One explanation is that the self-selection that occurred during recruitment produced a sample that was more ready to use NRT.

The fourth goal was to retain 70% of the sample at long-term follow-up. This study had one of the longest follow-up rates in any NRT study and provides evidence of the extended effects of the intervention. The retention rate at 30 months was 61%. This is below the 70% at 24 months reported in comparable studies. The lower retention rate is probably due to the additional 6 months involved in this study.

The fifth goal was to examine the differential effectiveness of NRT alone and in combination with three behavioral interventions. No significant differences were observed between the four conditions. Stage-matched manuals provided point prevalence cessation rates comparable to two previous studies. The addition of NRT, EXP, and TEL failed to produce a further increase in intervention effectiveness.

References

Alberg, A. J., Margalit, R. S., Burke, A., Rasch, K. A., Stewart, N., Kline, J. A., et al. (2004). The influence of offering free transdermal nicotine patches on quit rates in a local health department's smoking cessation program. *Addictive Behaviors, 29,* 1763–1768.

Benowitz, N. L., Jacob, P., III, Ahijevych, K., Jarvis, M. J., Hall, S., LeHouezec, J., et al. (2002). Biochemical verification of tobacco use and cessation. *Nicotine and Tobacco Research, 4,* 149–159.

Carpenter, M. J., Hughes, J. R., Solomon, L. J., & Callas, P. W. (2004). Both smoking reduction with nicotine replacement therapy and motivational advice increase future cessation among smokers unmotivated to quit. *Journal of Consulting and Clinical Psychology, 72,* 371–381.

Delichatsios, H. K., Friedman, R. H., Glanz, K., Tennstedt, S., Smigelski, C., Pinto, B. M., et al. (2001). Randomized trial of a "talking computer" to improve adult's eating habits. *American Journal of Health Promotion, 15,* 215–224.

Fiore, M. C., Smith, S. S., Jorenby, D. E., & Baker, T. B. (1994). The effectiveness of the nicotine patch for smoking cessation: A meta-analysis. *Journal of the American Medical Association, 266,* 3139–3144.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15,* 451–474.

Friedman, R. H. (1998). Automated telephone conversations to assess health behavior and deliver behavioral interventions. *Journal of Medical Systems, 22,* 95–102.

Friedman, R. H., Kazis, L. E., Jette, A., Smith, M. B., Stollerman, J., Torgerson, J., & Carey, K. (1996). A telecommunications system for monitoring and counseling patients with hypertension: Impact on medication adherence and blood pressure control. *American Journal of Hypertension, 9,* 285–292.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93,* 1261–1267.

Glasgow, R. E., Mullooly, J. P., Vogt, T. M., Stevens, V. J., Lichtenstein, E., Hollis, J. F., et al. (1993). Biochemical validation of smoking status in public health settings: Pros, cons and data from four low-intensity intervention trials. *Addictive Behaviors, 18,* 504–527.

Hall, S. M., Delucchi, K. L., Velicer, W. F., Kahler, C. W., Ranger-Moore, J., Hedeker, D., et al. (2001). Statistical analysis of randomized trials in tobacco treatment: Longitudinal designs with dichotomous outcome. *Nicotine and Tobacco Research, 3,* 193–202.

Hasford, J., Fagerstrom, K. O., & Haustein, K.-O. (2003). A naturalistic cohort study on effectiveness, safety and usage pattern of an over-the-counter nicotine patch. *European Journal of Clinical Pharmacology, 59,* 443–447.

Hughes, J. R. (1995). Combining behavioral therapy and pharmacotherapy for smoking cessation: An update. In L. S. Onken, J. D. Blaine, & J. J. Boren (Eds.), *Integrating behavior therapies with medication in the treatment of drug dependence* (NIDA Research Monograph; pp. 92–109). Washington, DC: U.S. Government Printing Office.

Hughes, J. R. (2004). Data to estimate the similarity of tobacco research samples to intended populations. *Nicotine and Tobacco Research, 6,* 177–179.

Hughes, J. R., Shiffman, S., Callas, P., & Zhang, J. (2003). A meta-analysis of the efficacy of over-the-counter nicotine replacement. *Tobacco Control, 12,* 21–27.

King, A. C., Friedman, R. H., Marcus, B., Napolitano, M., Castro, C., & Forsyth, L. (2003). Increasing regular activity via humans or automated technology: The CHAT trial. *Annals of Behavioral Medicine, 25*(Suppl.), S111.

Miller, N., Frieden, T. R., Liu, S. Y., Matte, T. D., Mostashari, F., Deitcher, D. R., et al. (2005). Effectiveness of a large scale distribution programme of free nicotine patches: A prospective evaluation. *Lancet, 365,* 1849–1854.

Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P. M. Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health, 84,* 1086–1093.

Peto, R., & Lopez, A. (1990). World-wide mortality from current smoking patterns. In B. Durstone & K. Jamrogik (Eds.), *The global war: Proceedings of the Seventh World Conference on Tobacco and Health* (pp. 62–68). East Perth, Western Australia: Organizing Committee of Seventh World Conference on Tobacco and Health.

Pierce, J. P., & Gilpin, E. A. (2002). Impact of over-the-counter sales on effectiveness of pharmaceutical aids for smoking cessation. *Journal of the American Medical Association, 288,* 1260–1264.

Pinto, B. M., Friedman, R. H., Marcus, B. H., Kelly, H., Tennstedt, S., & Gillman, N. W. (2002). Effects of computer-based telephone counseling systems on physical activity. *American Journal of Preventive Medicine, 23,* 113–120.

Po, A. L. W. (1993). Transdermal nicotine in smoking: A meta-analysis. *European Journal of Clinical Pharmacology, 45,* 519–528.

Prochaska, J. O., DiClemente, C. C. Velicer, W. F., & Rossi, J. S. (1993). Standardized, individualized, interactive and personalized self-help programs for smoking cessation. *Health Psychology, 12,* 399–405.

Prochaska, J. O., & Velicer, W. F. (2004). Integrating population smoking cessation policies and programs. *Public Health Reports, 119,* 244–252.

Prochaska, J. O., Velicer, W. F., DiClemente, C. C., & Fava, J. L. (1988). Measuring the processes of change: Applications to the cessation of smoking. *Journal of Consulting and Clinical Psychology, 56,* 520–528.

Prochaska, J. O., Velicer, W. F., Fava, J. L., Rossi, J. S., & Tsoh, J. Y. (2001). Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. *Addictive Behaviors, 26,* 583–602.

Prochaska, J. O., Velicer, W. F., Fava, J. L., Ruggiero, L., Laforge, R. G., & Rossi, J. S. (2001). Counselor and stimulus control enhancements of a stage-matched expert system intervention for smokers in a managed care setting. *Preventive Medicine, 32,* 23–32.

Ramelson, H. R., Friedman, R. H., & Ockene, J. K. (1999). An automated telephone-based smoking cessation education and counseling system. *Patient Education and Counseling, 36,* 131–144.

SAS Institute. (1997). *SAS/STAT software: Changes and enhancements through release 6.12.* Cary, NC: Author.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schmid, T. L., Jeffrey, R. W., & Hellerstedt, W. L. (1989). Direct mail recruitment to house-based smoking and weight control programs: A comparison of strengths. *Preventive Medicine, 18,* 503–517.

Silagy, C., Mant, D., Fowler, G., & Lodge, M. (1994). Meta-analysis on efficacy of nicotine replacement therapies in smoking cessation. *Lancet, 343,* 139–142.

Tang, J. L., Law, M., & Wald, N. (1994). How effective is nicotine replacement therapy in helping people to stop smoking? *British Medical Journal, 308,* 21–26.

Velicer, W. F., & DiClemente, C. C. (1993). Understanding and intervening with the total population of smokers. *Tobacco Control, 2,* 95–96.

Velicer, W. F., DiClemente, C. C., Prochaska, J. O., & Brandenberg, N. (1985). A decisional balance measure for assessing and predicting smoking status. *Journal of Personal and Social Psychology, 48,* 1279–1289.

Velicer, W. F., DiClemente, C. C., Rossi, J. S., & Prochaska, J. O. (1990). Relapse situations and self-efficacy: An integrative model. *Addictive Behaviors, 15,* 271–283.

Velicer, W. F., Fava, J. L., Prochaska, J. O., Abrams, D. B., Emmons, K. M., & Pierce, J. (1995). Distribution of smokers by stage in three representative samples. *Preventive Medicine, 24,* 401–411.

Velicer, W. F., Keller, S., Friedman, R. H., Fava, J. L., Gulliver, S. B., Ward, R. M., et al. (2005). Comparing participants and nonparticipants recruited for an effectiveness study of nicotine replacement therapy. *Annals of Behavioral Medicine, 21,* 181–191.

Velicer, W. F., & Prochaska, J. O. (1999). An expert system intervention for smoking cessation. *Patient Education and Counseling, 36,* 119–129.

Velicer, W. F., & Prochaska, J. O. (2004). A comparison of four self-report smoking cessation outcome measures. *Addictive Behaviors, 29,* 51–60.

Velicer, W. F., Prochaska, J. O., Bellis, J. M., DiClemente, C. C., Rossi, J. S., Fava, J. L., & Steiger, J. H. (1993). An expert system intervention for smoking cessation. *Addictive Behaviors, 18,* 269–290.

Velicer, W. F., Prochaska, J. O., Fava, J. L., Laforge, R. G., & Rossi, J. S. (1999). Interactive versus non-interactive interventions and dose-response relationships for stage-matched smoking cessation programs in a managed care setting. *Health Psychology, 18,* 1–8.

Velicer, W. F., Prochaska, J. O., Fava, J. L., Redding, C. A., Rossi, J. S., Rossi, S. R., et al. (2006). *Stage distributions for three risk factors in three large USA samples*. Manuscript submitted for publication.

Velicer, W. F., Prochaska, J. O., & Redding, C. A. (2006). Tailored communications for smoking cessation: Past successes and future directions. *Drug and Alcohol Review, 25,* 47–55.

Velicer, W. F., Prochaska, J. O., Rossi, J. S., & Snow, M. (1992). Assessing outcome in smoking cessation studies. *Psychological Bulletin, 111,* 23–41.

Velicer, W. F., Rossi, J. S., Ruggiero, L., & Prochaska, J. O. (1994). Minimal interventions appropriate for an entire population of smokers. In R. Richmond (Ed.), *Interventions for smokers: An international perspective* (pp. 69–92). Baltimore: Williams & Wilkins.

Received August 25, 2005
Revision received July 13, 2006
Accepted July 14, 2006 ■

Readiness to Change in Primary Care Patients Who Screened Positive for Alcohol Misuse

Emily C. Williams, MPH[1–3]
Daniel R. Kivlahan, PhD[2,4]
Richard Saitz, MD, MPH[5,6]
Joseph O. Merrill, MD, MPH[3,7]
Carol E. Achtmeyer, MN[1,2,8]
Kinsey A. McCormick, BA[1,3]
Katharine A. Bradley, MD, MPH[1–3,8,9]

[1]Health Services Research & Development, VA Puget Sound Health Care System, Seattle, Wash

[2]Center of Excellence in Substance Abuse Treatment and Education, VA Puget Sound Health Care System, Seattle, Wash

[3]Department of Medicine, School of Medicine, University of Washington, Seattle, Wash

[4]Departments of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Wash

[5]Clinical Addiction Research and Education Unit, Department of Medicine, Boston University School of Medicine, and Boston Medical Center, Boston, Mass

[6]Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, Mass

[7]Harborview Medical Center, Seattle, Wash

[8]Primary and Specialty Medical Care Service, VA Puget Sound Health Care System, Seattle, Wash

[9]Department of Health Services, University of Washington, Seattle, Wash

Conflicts of interest: none reported

CORRESPONDING AUTHOR

Emily C. Williams, MPH
VA Puget Sound Health Care System
1100 Olive Way, Suite 1400
Seattle, WA 98101
emwilli@u.washington.edu

ABSTRACT

PURPOSE Readiness to change drinking may influence the content or effectiveness of brief alcohol counseling. This study was designed to assess readiness to change and its relationship to alcohol misuse severity among primary care patients whose screening questionnaire was positive for alcohol misuse.

METHODS This study was a cross-sectional analysis of data collected from 2 consecutive mailed questionnaires. Male outpatients at 7 Veterans Affairs (VA) general medicine clinics were eligible if they returned both questionnaires, screened positive for alcohol misuse (augmented CAGE Questionnaire ≥1 point), responded to 3 readiness-to-change questions, and completed the Alcohol Use Disorders Identification Test (AUDIT). A validated algorithm based on 3 standardized questions categorized participants into 3 readiness groups (precontemplation, contemplation, action). Measures of alcohol misuse severity included AUDIT, CAGE, and the 3 consumption questions from the AUDIT (AUDIT-C). Analyses were descriptive; linear-by-linear associations between alcohol misuse severity and readiness were tested with χ^2 statistics.

RESULTS Response rates to the first and second surveys were 59% and 55%, respectively. Of the 6,419 eligible outpatients who screened positive for alcohol misuse, 4,797 (75%) reported any readiness to change (contemplation 24%, action 51%). Among patients with AUDIT scores >8, more than 90% indicated that they drank more than they should and/or had contemplated drinking less. Greater readiness was significantly associated with greater alcohol misuse severity (P <.001 for all measures).

CONCLUSIONS Most primary care patients who screen positive for alcohol misuse indicate some readiness to change. Contrary to stereotypes of denial, those with greater alcohol misuse severity are more likely to report readiness to change.

Ann Fam Med 2006;4:213-220. DOI: 10.1370/afm.542.

INTRODUCTION

Alcohol misuse is common and among the most costly public health problems in the United States.[1,2] Brief counseling interventions reduce drinking in patients who screen positive for alcohol misuse.[3] The US Preventive Services Task Force recommends routine alcohol screening in primary care settings and brief alcohol counseling for patients who screen positive for alcohol misuse.[4] Though the exact content of brief alcohol counseling interventions evaluated in randomized trials has varied, these interventions typically include patient-centered assessment, individualized feedback and advice, agreement on a drinking goal, and follow-up.[5] These interventions last approximately 5 to 15 minutes when delivered in primary care settings.[6]

As with counseling about other health behaviors, such as tobacco use, when clinicians counsel patients about drinking, the appropriate focus of the discussion and patients' receptivity can depend on patients' readiness to change.[7-9] For example, the content of brief counseling should reflect whether the patient recognizes his drinking is a problem and whether he

has contemplated or tried to change.[10-12] One study suggested that the efficacy of brief alcohol counseling differed for patients with differing readiness to change.[11] Further, clinicians have reported concern that primary care patients who screen positive for alcohol misuse will deny that they misuse alcohol or will not be interested in discussing or changing their drinking when the issue is raised.[13-16] Denial of alcohol misuse is sometimes assumed to be a characteristic of alcohol dependence.[14]

Despite its clinical utility, little research has addressed the prevalence of denial or readiness to change among primary care patients who misuse alcohol.[11,17,18] No study has described readiness to change in a large sample of primary care patients who screened positive for alcohol misuse but were not recruited into a study focused exclusively on alcohol misuse. This study describes readiness to change and evaluates its relationship to alcohol misuse severity in more than 6,000 outpatients who screened positive for alcohol misuse as part of a quality improvement study addressing multiple common outpatient conditions.

METHODS

Setting and Population

General internal medicine outpatients from 7 Veterans Affairs (VA) sites (Birmingham, Ala; Little Rock, Ark; San Francisco, Calif; West Los Angeles, Calif; White River Junction, Vt; Richmond, Va; and Seattle, Wash) were surveyed by mail as part of the VA Ambulatory Care Quality Improvement Project (ACQUIP).[19] The ACQUIP trial was a quality improvement study evaluating the effectiveness of giving clinicians individualized feedback based on patient questionnaires regarding 6 medical conditions (coronary artery disease, diabetes, chronic obstructive pulmonary disease, hypertension, depression, and alcohol misuse). A representative sample of 62,487 patients who visited a participating general internal medicine clinic from 1997 to 2000 received a mailed Health Checklist, which included a validated 8-item screen consisting of the CAGE Questionnaire (*c*ut down, *a*nnoyed, *g*uilty, and *e*ye opener), combined with 4 additional questions about alcohol consequences and previous problem drinking.[20] Potential ACQUIP participants were identified through medical records review. A waiver of written informed consent was obtained along with study approval from the University of Washington Human Subjects Committee and the institutional review boards at each site.

Patients who returned this Health Checklist and who had a screening questionnaire that was positive for alcohol misuse (≥1 point)[20] on the augmented CAGE Questionnaire were mailed a subsequent survey instrument, the Drinking Practices Questionnaire, to further assess alcohol use. This instrument included the 10-item Alcohol Use Disorders Identification Test (AUDIT) and 3 readiness-to-change questions (both measures described below). Patients were also mailed up to 5 other condition-specific questionnaires if they had indicated they had these conditions. A generic cover letter was enclosed with the condition-specific questionnaires indicating that patient responses might be shared with their primary care clinicians as part of a study designed to determine whether such information improved care. There was no alcohol-specific information in the cover letter.

Male respondents to the Drinking Practices Questionnaire who reported drinking in the past year and completed the readiness-to-change, AUDIT, and CAGE questions were included in this study. The few female respondents were excluded because of sex differences in the performance of alcohol-screening questionnaires[21] and insufficient variation in alcohol misuse severity among participating women to conduct sex-specific analyses.

Measures

Readiness to Change

Readiness to change has often been categorized into 3 stages (precontemplation, contemplation, and action) based on Prochaska and DiClemente's transtheoretical model.[11,18,22-27] Although contemplation was later broken down into component stages, the original transtheoretical model divided patients into 3 groups: those who had no recognition that they drank more than they should and who were not trying to change, those who had some recognition of drinking excessively and who were often contemplating change, and those taking steps to change.[23] In our study, readiness to change was measured using a brief algorithm based on 3 standardized questions (Figure 1).[28] This algorithm categorizes patients who misuse alcohol into precontemplation, contemplation, or action groups based on the transtheoretical model.[29] The questions are designed to guide clinician interventions with patients who screen positive for alcohol misuse and to provide information about changes in drinking at follow-up visits. The questions address any recent changes in the patient's drinking (past 3 months), patient self-recognition of excessive drinking, and whether the patient had considered changing his drinking. In addition to face validity, the readiness-to-change algorithm had good concurrent validity when compared with the longer Readiness to Change Questionnaire by Rollnick et al,[30] which was validated in a population of female VA patients who screened positive for alcohol misuse.[28]

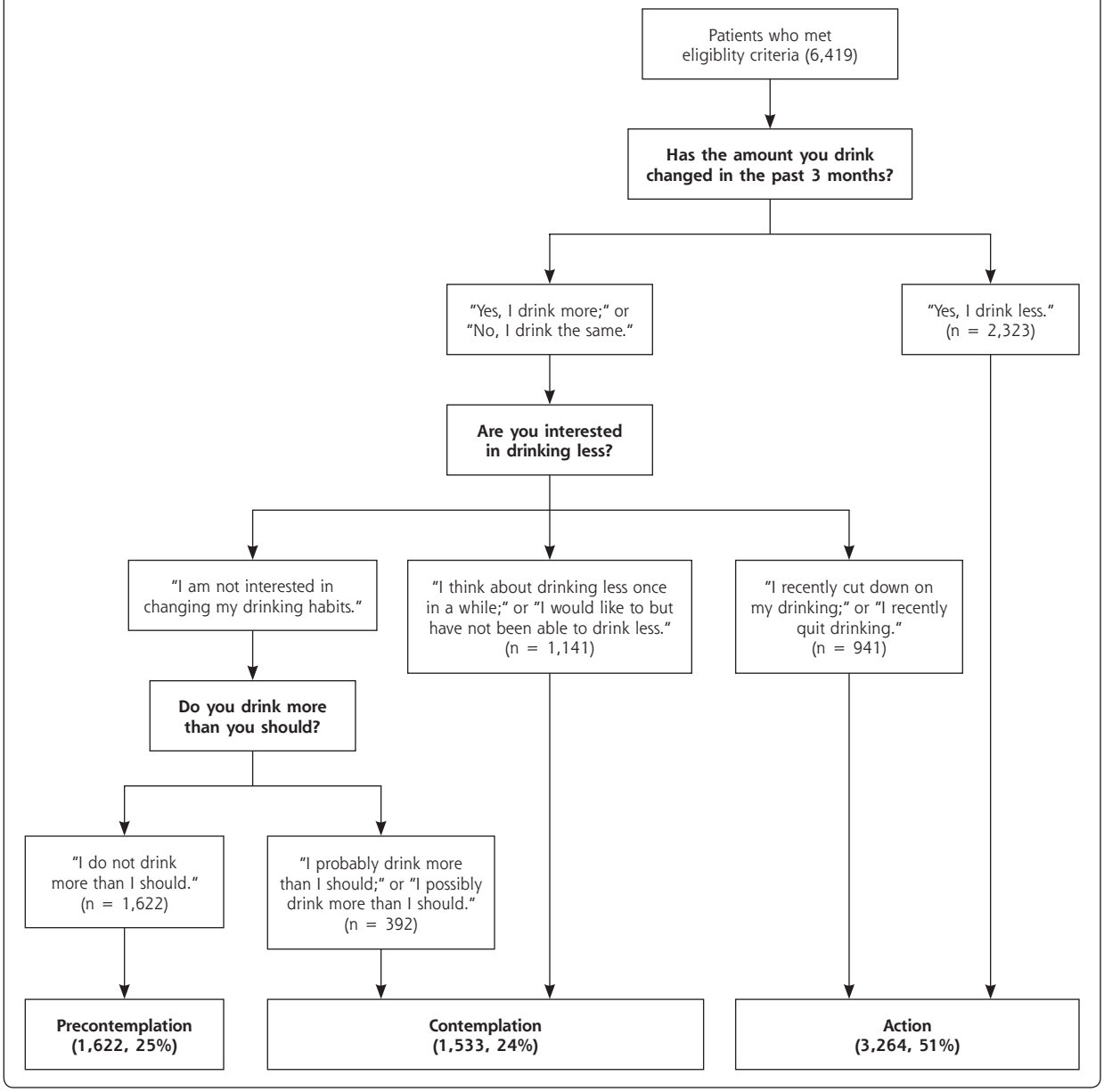
Measures of Alcohol Misuse Severity

Our primary measure of alcohol misuse severity was the AUDIT, a validated alcohol-screening questionnaire specifically developed to identify patients across the spectrum of alcohol misuse[31] and validated in a male VA outpatient population.[32] The AUDIT scores range from 1 to 40 for drinkers, with higher scores indicating greater severity.[33] The World Health Organization recommends the use of 4 severity zones on the AUDIT (<8, 8-15, 16-19, ≥20). Brief alcohol counseling is recommended for those with AUDIT scores ≥8 but <20.[33]

Patients with high AUDIT scores are those willing to report past-year problems caused by drinking, which may, in itself, reflect increased recognition or willingness to consider changing drinking. We therefore evaluated the association between readiness to change and 2 scores of other brief alcohol-screening questionnaires that do not ask explicitly about problems caused by drinking in the past year: the 3 AUDIT consumption questions (AUDIT-C) and the CAGE Questionnaire.

The first 3 questions of the 10-item AUDIT, the AUDIT-C, address alcohol consumption patterns in the past year and have been validated as a brief alcohol-screening test. AUDIT-C scores range from 1 to 12 for drinkers, and a score of 4 or more is effective for screening for alcohol misuse in men.[34] Typical con-

Figure 1. Readiness to change of patients enrolled in the study.



sumption reported on these questions is imprecise and on average underestimates actual consumption,[35] but increasing scores are related to increasing severity of self-reported alcohol-related problems in the past year.[36]

The 4-item CAGE is an effective screening test for active alcohol abuse and/or dependence (possible scores 0 to 4) but is less sensitive for detecting patients with milder alcohol misuse.[32,37] A "yes" response to any CAGE question (score ≥1) indicates a positive finding for active alcohol abuse and/or dependence in this population,[32] although ≥2 has often been used.[38] The CAGE score is associated with increasing severity of problems caused by drinking in the past year reported on the AUDIT, but the relationship is weaker than that with AUDIT-C scores.[36]

Demographic Characteristics

Patient age and race were obtained from VA electronic medical records. Other demographic characteristics including income, education, and marital status were obtained from the ACQUIP questionnaires.

Analyses

Descriptive analyses assessed demographic and clinical characteristics, alcohol-screening results, and the prevalence of precontemplation, contemplation, and action in the study sample of ACQUIP participants who screened positive for alcohol misuse on the augmented CAGE Questionnaire and completed the readiness-to-change questions. To evaluate nonresponse bias, demographic characteristics and alcohol misuse severity measures from the Health Checklist were cross-tabulated for study participants and nonparticipants, and Pearson χ^2 statistics were obtained to identify differences.

Analyses first evaluated the association of readiness to change with alcohol misuse severity as measured by the AUDIT. The proportion of patients with each AUDIT score who were in the contemplation or action group was evaluated and depicted graphically. Linear-by-linear χ^2 tests were then completed to test the association between readiness groups and all alcohol misuse severity measures ($\alpha = .05$). To explore further the association of readiness to change with the severity of alcohol misuse based on the AUDIT, we evaluated responses to each of the 3 readiness-to-change questions within individual readiness groups (precontemplation, contemplation, and action). All analyses were carried out using SPSS Version 12.0.[39]

RESULTS

More than one half the patients returned the Health Checklist (32,821; 59% of those eligible) and 11,889 (36% of respondents) screened positive for alcohol misuse (≥1 point)[20] on the augmented CAGE; of these, 6,551 patients (55% of those eligible for this survey, 10% of the initial study population) completed the Drinking Practices Questionnaire. There were 6,419 male respondents to the Drinking Practices Questionnaire who reported drinking in the past year and who completed the readiness-to-change, AUDIT, and CAGE questions.

The characteristics of the 6,419 participants are displayed in Table 1. These respondents represented 55% of the men who screened positive for alcohol misuse on the initial Health Checklist. When these participants were compared with male patients who screened positive for alcohol misuse was positive but did not respond to the Drinking Practices Questionnaire (n = 4,815) or

Table 1. Demographic Characteristics of Participants (N = 6,419)

Characteristic	Value
Age in years, mean (SD)	60 (11.5)
Race,* No. (%)	
African American	1,199 (19)
White	4,081 (64)
Other	981 (15)
Marital status,† No. (%)	
Never married	597 (9)
Currently married	3,163 (49)
Divorced/separated/widowed	2,553 (40)
Education, some college, No. (%)	3,294 (51)
Annual income, No. (%)	
<$20,000	3,769 (58)
$20,000-$50,000	1,946 (30)
>$50,000	704 (11)
10-item AUDIT score, No. (%)	
1-7 (zone 1)	4,083 (64)
8-15 (zone 2)	1,543 (24)
16-19 (zone 3)	290 (5)
20-40 (zone 4)	503 (8)
AUDIT-C score, No. (%)	
1-3	2,427 (38)
4-5	1,749 (27)
6-7	1,031 (16)
8-9	667 (10)
10-12	545 (9)
CAGE score, No. (%)	
0	1,595 (25)
1	1,424 (22)
2	1,598 (25)
3	1,095 (17)
4	707 (11)

AUDIT = Alcohol Use Disorders Identification Test; AUDIT-C = 3 AUDIT consumption questions; CAGE = CAGE Questionnaire (cut down, annoyed, guilty, and eye opener).

* n = 6,261, missing data for 158 subjects.
† n = 6,269, missing data for 150 subjects.

indicated they no longer drank alcohol (n = 389), study participants were slightly older (59.5 vs 58.1 years); were more likely to be white (64% vs 53%), married (49% vs 43%), retired (39% vs 35%), and to have attended college (51% vs 46%); and reported incomes greater than $20,000 per year (41% vs 36%) (P <.001 for all measures). Participants were also more likely than eligible nonrespondents to the Drinking Practices Questionnaire to have a positive score on the AUDIT-C (≥4 points) (67% vs 58%; P <.001) and less likely to score 2 or more points on the CAGE (53% vs 60%; P <.001) on the initial Health Checklist questionnaire.

The 6,419 respondents reflected a broad spectrum of alcohol misuse severity based on responses to all 3 measures, with 36% of patients having AUDIT scores ≥8. According to their responses to the readiness-to-change questions, 25% were categorized into precontemplation, 24% into contemplation, and 51% into action. When readiness groups were cross-tabulated with alcohol misuse severity measures, there were significant, positive linear-by-linear associations between readiness groups and alcohol misuse severity (P <.001) (Table 2).

Because of the algorithm used to define the precontemplation and action groups (Figure 1), analyses of individual responses to readiness-to-change questions among patients categorized into these groups showed relatively homogenous responses (data not presented). Patients who were categorized into the contemplation group, however, had potentially important variation in responses to the 3 individual readiness-to-change questions. Specifically, the proportion of patients in the contemplation group who reported "I am sure I drink more than I should" increased as the severity of alcohol misuse increased: 4% in AUDIT zone 1 (AUDIT scores <8); 16% in zone 2 (AUDIT scores 8 to 15); 39% in AUDIT zone 3 (AUDIT scores 16 to 19); and 69% in zone 4 (AUDIT scores >20). Similarly, higher severity of alcohol misuse was associated with increasing percentages of patients in the contemplation group reporting wanting but not having been able to reduce drinking: 3% in AUDIT zone 1; 10% in AUDIT zone 2; 27% in AUDIT zone 3; and 60% in AUDIT zone 4.

Table 2. Participants Categorized Into Readiness to Change Groups Stratified by Alcohol Screening Scores

Alcohol Misuse Severity Score	Precontemplation n (%)	Contemplation n (%)	Action n (%)	Total	*P* Value
AUDIT					<.001
<8	1,518 (37)	631 (15)	1,934 (47)	4,083	
8-15	96 (6)	535 (35)	912 (59)	1,543	
16-19	5 (2)	122 (42)	163 (56)	290	
>20	3 (1)	245 (49)	255 (51)	503	
AUDIT-C					<.001
1-3	956 (39)	167 (7)	1,304 (54)	2,427	
4-5	497 (28)	444 (25)	808 (46)	1,749	
6-7	113 (11)	362 (35)	556 (54)	1,031	
8-9	36 (5)	278 (42)	353 (53)	667	
10-12	20 (4)	282 (52)	243 (45)	545	
CAGE					<.001
0	672 (42)	297 (19)	626 (39)	1,595	
1	283 (20)	386 (27)	755 (53)	1,424	
2	357 (22)	418 (26)	823 (52)	1,598	
3	197 (18)	272 (25)	626 (57)	1,095	
4	113 (16)	160 (23)	434 (61)	707	

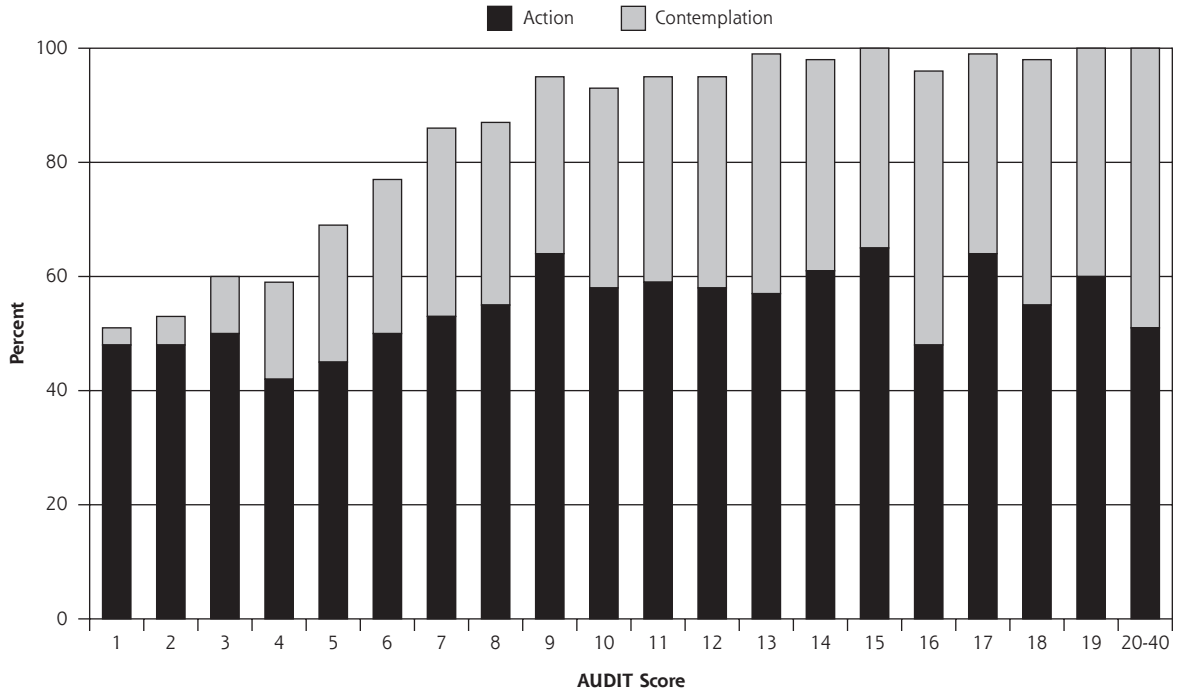
AUDIT = Alcohol Use Disorders Identification Test; AUDIT-C = 3 AUDIT consumption questions ; CAGE = CAGE Questionnaire (cut down, annoyed, guilty, and eye opener).

DISCUSSION

In this population of male VA primary care patients who screened positive for alcohol misuse in the past year, 75% indicated some readiness to change. Contrary to a stereotype that patients with alcohol misuse will deny concerns about their drinking, most patients whose tests were positive for alcohol misuse in this population indicated they drank more than they should or had considered or tried decreasing drinking. Moreover, readiness to change increased steadily as the severity of alcohol misuse increased. Among patients who scored 8 or more on the AUDIT, indicative of a high likelihood of active alcohol use disorders in this population,[34,35] readiness to change was relatively stable, with more than 90% of patients indicating that they drank excessively or had considered change (Figure 2).

Findings of the present study were consistent with those of 2 previous primary care studies.[11,17,18] Both required that patients provide written informed consent to participate in a study of their drinking. One study was small, and 56% of the 184 participating patients who had positive CAGE scores no longer drank alcohol.[18] The other study used a sample of patients recruited for a randomized controlled trial of brief alcohol counseling interventions.[11] In that study, 78% of patients who had positive test results did not participate in the

Figure 2. Percentage of male patients in contemplation or action groups by scores on the full 10-item AUDIT (n = 6,419).



AUDIT = Alcohol Use Disorders Identification Test.

trial, potentially limiting the generalizability of their findings. Our study is the first to describe readiness to change in a large sample of primary care patients who screened positive for alcohol misuse but were not being recruited into a study focusing exclusively on their drinking.

This study has several noteworthy limitations. First, our population consisted of male VA patients who were predominantly white and older, which may limit the generalizability of our findings to women and other primary care populations. Second, nonresponse bias could have influenced our findings. Forty-five percent of patients who screened positive for alcohol misuse on the initial ACQUIP survey were not included in these analyses. Nonparticipants were more likely to screen positive for a lifetime history of alcohol abuse or dependence on the CAGE, but were less likely to report high levels of current consumption on the AUDIT-C. Lower AUDIT-C scores among nonrespondents could reflect underreporting of alcohol use. Even in the unlikely event that all nonrespondents to the Drinking Practices Questionnaire were in the precontemplation group, however, almost one half of the resulting sample would still have been classified in contemplation or action. Social desirability could also have accounted for some of our findings. Patients who were not interested in changing might minimize their alcohol misuse and related problems or overreport recent decreases in their drinking so that their primary care clinicians would not bother them about their drinking. Finally, the algorithm used to measure readiness to change in this study has been validated only in a female VA population.[28]

The study also has several important strengths. First, the size of the sample is a strength, as is that more than 10% of the entire eligible study sample screened positive for alcohol misuse and returned the Drinking Practices Questionnaire. Additionally, the prevalences of self-reported alcohol misuse and related problems were high, decreasing the likelihood that high rates of readiness to change reflected social desirability bias and providing rich variation in alcohol misuse among participants. Finally, although the readiness-to-change questions have been validated in only women veterans, the brevity and clinical accessibility of this instrument makes it one of our study's unique strengths.

This study indicates that most primary care patients who screened positive for alcohol misuse and who returned a questionnaire that assesses alcohol misuse had some recognition that they drink more than they should and/or have at least contemplated drinking less. Moreover, as screening scores increased,

patients were more likely to report wanting but having been unable to decrease drinking. These findings have several implications for clinicians. First, primary care clinicians sometimes expect patients to deny their alcohol misuse when the issue is raised.[13-16] Such denial is often thought to be especially common among patients with the most-severe problems, indicative of alcohol dependence.[14]

Our findings support findings of previous studies in more-select populations that suggest the opposite is true. The more a patient drinks and the more severe his problems caused by drinking, the more likely he will report recognition of or interest in changing his drinking. A recent study of clinicians' attitudes and their association with smoking cessation counseling reported that clinicians' perceptions that patients are not willing to quit smoking were associated with a lower counseling proclivity.[40] Similar clinicians' attitudes may be a barrier to conducting brief alcohol counseling interventions among patients with alcohol misuse. Our finding that most primary care patients with alcohol misuse report some level of readiness to change could help correct such attitudes. Second, these findings suggest that primary care clinicians could use scores obtained from brief alcohol-screening questionnaires as an indicator of readiness to change. Patients whose screening tests are positive but who have low scores are least likely to recognize they drink more than they should or to considered changing; brief advice aimed at assisting with problem recognition and building motivation may be most appropriate for these patients.

Maisto et al found that low levels of readiness to change at baseline were associated with improved drinking outcomes at follow-up for patients who received brief advice.[11] Taken with our findings, we hypothesize that patients with the least severe alcohol misuse, who are least likely to recognize that they drink excessively or to have contemplated change, may benefit most from brief alcohol-related advice in primary care settings.[11] This stance is consistent with current evidence-based guidelines that recommend screening for the entire spectrum of alcohol misuse in primary care settings.[4,41] Appropriate and brief screening tests for this purpose include the AUDIT-C or single-item questions about binge drinking.[34,42,43]

That patients in this study with higher alcohol-screening scores were more likely to indicate recognition of their alcohol misuse or interest in changing should not be confused with readiness to enter specialized alcohol treatment. Many patients in this study's contemplation group reported wanting but having been unable to decrease drinking, which was positively associated with the severity of alcohol misuse. Patients with more-severe alcohol misuse may need more-intense or repeated primary care interventions to support them toward specialized addictions treatment or abstinence.[44,45] Similarly, it is unknown whether indication of readiness to change by primary care patients will result in subsequent changes in drinking, though patient readiness to change has been associated with decreased drinking in hospitalized patients.[27] Further research is needed in both of these areas.

In conclusion, this study shows that most primary care patients who screen positive for alcohol misuse indicate concern about or are considering changing their drinking. Further, simple questions can elicit statements reflecting readiness to change, and patients with the highest alcohol-screening scores and greater alcohol misuse severity are most likely to indicate some readiness to change drinking. These findings should help counter clinician attitudes that patients with alcohol misuse deny excessive drinking or will not be interested in changing.

To read or post commentaries in response to this article, see it online at http://www.annfammed.org/cgi/content/full/4/3/213.

Key words: Alcohol drinking; alcoholism/diagnosis; patient acceptance of health care

Submitted June 29, 2005; submitted, revised, October 28, 2005; accepted November 28, 2005.

A preliminary version of this study was presented at the Society of General Internal Medicine annual meeting, Chicago, Ill, May 14, 2004.

Funding support: The Ambulatory Care Quality Improvement Project (ACQUIP) was funded by Veterans Affairs Health Services Research and Development grants: SDR 96-002 and IIR 99-376. Dr Bradley is supported by a National Institute on Alcohol Abuse and Alcoholism award (K23AA00313) and was a Robert Wood Johnson Generalist Physician Faculty Scholar at the time this work was completed.

Disclaimer: Views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs, the University of Washington, the National Institute on Alcohol Abuse and Alcoholism, or the Robert Wood Johnson Foundation.

References

1. Grant BF, Dawson DA, Stinson FS, et al. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. *Drug Alcohol Depend.* 2004;74:223-234.
2. *Tenth Special Report to the US Congress on Alcohol and Health.* US Department of Health and Human Services. National Institute on Alcohol Abuse and Alcoholism; 2000.
3. Whitlock EP, Polen MR, Green CA, Orleans T, Klein J. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2004;140:557-568.
4. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. *Ann Intern Med.* 2004;140:554-556.
5. Saitz R. Clinical practice. Unhealthy alcohol use. *N Engl J Med.* 2005;352:596-607.

6. Babor TF, Higgins-Biddle JC. Alcohol screening and brief intervention: dissemination strategies for medical practice and public health. *Addiction.* 2000;95:677-686.
7. US Department of Health and Human Services. National Institute on Alcohol Abuse and Alcoholism. *Helping Patients With Alcohol Problems: A Practitioner's Guide.* Bethesda, Md: National Institute of Health; 2003.
8. Rustin TA. Assessing nicotine dependence. *Am Fam Physician.* 2000;62:579-584, 591-572.
9. Dijkstra A, De Vries H, Roijackers J. Targeting smokers with low readiness to change with tailored and nontailored self-help materials. *Prev Med.* 1999;28:203-211.
10. Samet JH, Rollnick S, Barnes H. Beyond CAGE. A brief clinical approach after detection of substance abuse. *Arch Intern Med.* 1996;156:2287-2293.
11. Maisto SA, Conigliaro J, McNeil M, et al. Effects of two types of brief intervention and readiness to change on alcohol use in hazardous drinkers. *J Stud Alcohol.* 2001;62:605-614.
12. Heather N, Rollnick S, Bell A, Richmond R. Effects of brief counseling among male heavy drinkers identified on general hospital wards. *Drug Alcohol Rev.* 1996;15:29-38.
13. Beich A, Gannik D, Malterud K. Screening and brief intervention for excessive alcohol use: qualitative interview study of the experiences of general practitioners. *BMJ.* 2002;325:870.
14. Arborelius E, Damstrom Thakker K. Why is it so difficult for general practitioners to discuss alcohol with patients? *Fam Pract.* 1995;12:419-422.
15. Thom B, Tellez C. A difficult business: detecting and managing alcohol problems in general practice. *Br J Addict.* 1986;81:405-418.
16. Aira M, Kauhanen J, Larivaara P, Rautio P. Factors influencing inquiry about patients' alcohol consumption by primary health care physicians: qualitative semi-structured interview study. *Fam Pract.* 2003;20:270-275.
17. Maisto SA, Conigliaro J, McNeil M, et al. Factor structure of the SOCRATES in a sample of primary care patients. *Addict Behav.* 1999;24:879-892.
18. Samet JH, O'Connor PG. Alcohol abusers in primary care: readiness to change behavior. *Am J Med.* 1998;105:302-306.
19. Fihn SD, McDonell MB, Diehr P, et al. Effects of sustained audit/feedback on self-reported health status of primary care patients. *Am J Med.* 2004;116:241-248.
20. Bradley KA, Kivlahan DR, Bush KR, et al. Variations on the CAGE alcohol screening questionnaire: strengths and limitations in VA general medical patients. *Alcohol Clin Exp Res.* 2001;25:1472-1478.
21. Bradley KA, Boyd-Wickizer J, Powell S, Burman ML. Alcohol screening questionnaires in women: a critical review. *JAMA.* 1998:280:166-171.
22. DiClemente CC, Schlundt D, Gemmell L. Readiness and stages of change in addiction treatment. *Am J Addict.* 2004;13:103-119.
23. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. *Psychother Theory Res Prac.* 1982;19:276-288.
24. Vik PW, Culbertson KA, Sellers K. Readiness to change drinking among heavy-drinking college students. *J Stud Alcohol.* 2000;61:674-680.
25. Isenhart CE. Pretreatment readiness for change in male alcohol dependent subjects: predictors of one-year follow-up status. *J Stud Alcohol.* 1997;58:351-357.
26. Dent TH, Shepherd RM, Alexander GJ, London M. Do CAGE scores predict readiness to reduce alcohol consumption in medical inpatients? *Alcohol Alcohol.* 1995;30:577-580.
27. Heather N, Rollnick S, Bell A. Predictive validity of the Readiness to Change Questionnaire. *Addiction.* 1993;88:1667-1677.
28. Epler AJ, Kivlahan DR, Bush KR, Dobie DJ, Bradley KA. A brief readiness to change drinking algorithm: concurrent validity in female VA primary care patients. *Addict Behav.* 2005;30:389-395.
29. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol.* 1983;51:390-395.
30. Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. *Br J Addict.* 1992;87:743-754.
31. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction.* 1993;88:791-804.
32. Bradley KA, Bush KR, McDonell MB, Malone T, Fihn SD. Screening for problem drinking: comparison of CAGE and AUDIT. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *J Gen Intern Med.* 1998;13:379-388.
33. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro M. *AUDIT - The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care.* 2nd ed. World Health Organization, Department of Mental Health and Substance Abuse; 2001. Report No.: WHO/MSD/MSB/01.6a.
34. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med.* 1998;158:1789-1795.
35. Bradley KA, McDonell MB, Bush K, et al. The AUDIT alcohol consumption questions: reliability, validity, and responsiveness to change in older male primary care patients. *Alcohol Clin Exp Res.* 1998;22:1842-1849.
36. Bradley KA, Kivlahan DR, Zhou XH, et al. Using alcohol screening results and treatment history to assess the severity of at-risk drinking in Veterans Affairs primary care patients. *Alcohol Clin Exp Res.* 2004;28:448-455.
37. Adams WL, Barry KL, Fleming MF. Screening for problem drinking in older primary care patients. *JAMA.* 1996;276:1964-1967.
38. Fiellin DA, Reid MC, O'Connor PG. Screening for alcohol problems in primary care: a systematic review. *Arch Intern Med.* 2000;160:1977-1989.
39. *SPSS for Windows* [computer program]. Version 10.0.7. Chicago, Ill: SPSS Inc; 1999.
40. Meredith LS, Yano EM, Hickey SC, Sherman SE. Primary care provider attitudes are associated with smoking cessation counseling and referral. *Med Care.* 2005;43:929-934.
41. US Department of Health and Human Services. National Institute on Alcohol Abuse and Alcoholism. *Helping Patients Who Drink Too Much: A Clinician's Guide.* Bethesda, Md: National Institute of Health; 2005.
42. Taj N, Devera-Sales A, Vinson DC. Screening for problem drinking: does a single question work? *J Fam Pract.* 1998;46:328-335.
43. Williams R, Vinson DC. Validation of a single screening question for problem drinking. *J Fam Pract.* 2001;50:307-312.
44. Lieber CS, Weiss DG, Groszmann R, Paronetto F, Schenker S. I. Veterans Affairs Cooperative Study of polyenylphosphatidylcholine in alcoholic liver disease: effects on drinking behavior by nurse/physician teams. *Alcohol Clin Exp Res.* 2003;27:1757-1764.
45. Willenbring ML, Olson DH. A randomized trial of integrated outpatient treatment for medically ill alcoholic men. *Arch Intern Med.* 1999;159:1946-1952.

PERSPECTIVE

Toward a VA Women's Health Research Agenda: Setting Evidence-based Priorities to Improve the Health and Health Care of Women Veterans

Elizabeth M. Yano, PhD,[1,2] Lori A. Bastian, MD,[3,4] Susan M. Frayne, MD, MPH,[5,6] Alexandra L. Howell, PhD,[7,8] Linda R. Lipson, MA,[9] Geraldine McGlynn, MEd,[10] Paula P. Schnurr, PhD,[8,11] Margaret R. Seaver, MD, MPH,[12,13] Ann M. Spungen, PhD,[14,15] Stephan D. Fihn, MD, MPH[16,17]

[1]VA Greater Los Angeles HSR&D Center of Excellence, VA Greater Los Angeles Healthcare System, Sepulveda, CA, USA; [2]Department of Health Services, UCLA School of Public Health, Los Angeles, CA, USA; [3]Center for Health Services Research in Primary Care, VA HSR&D Center of Excellence, VA Medical Center, Durham, NC, USA; [4]Duke University, Durham, NC, USA; [5]Center for Healthcare Evaluation, VA HSR&D Center of Excellence, VA Palo Alto Healthcare System, Palo Alto, CA, USA; [6]Stanford University, Palo Alto, CA, USA; [7]VA Medical Center, White River Junction, VT, USA; [8]Dartmouth Medical School, Hanover, NH, USA; [9]Veterans Health Administration, VA HSR&D Service, Washington, DC, USA; [10]Center for Information Dissemination and Education Resources (CIDER), VA HSR&D Service, VA Boston Healthcare System, Boston, MA, USA; [11]VA National Center for Posttraumatic Stress Disorder, VA Medical Center, White River Junction, VT, USA; [12]Primary Care Service Line, VA Boston Healthcare System, Boston, MA, USA; [13]Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [14]Center of Excellence for the Medical Consequences of Spinal Cord Injury, VA Rehabilitation Research and Development (RR&D), James J. Peter VA Medical Center, Bronx, NY, USA; [15]Mount Sinai School of Medicine, New York, NY, USA; [16]Northwest Center for Outcomes Research in Older Adults, VA HSR&D Center of Excellence, VA Puget Sound Healthcare System, Seattle, WA, USA; [17]Department of Medicine, University of Washington, Seattle, WA, USA.

The expansion of women in the military is reshaping the veteran population, with women now constituting the fastest growing segment of eligible VA health care users. In recognition of the changing demographics and special health care needs of women, the VA Office of Research & Development recently sponsored the first national VA Women's Health Research Agenda-setting conference to map research priorities to the needs of women veterans and position VA as a national leader in Women's Health Research. This paper summarizes the process and outcomes of this effort, outlining VA's research priorities for biomedical, clinical, rehabilitation, and health services research.

KEY WORDS: women's health; research and development; research priorities; veterans; health care quality; access and evaluation.
DOI: 10.1111/j.1525-1497.2006.00381.x
J GEN INTERN MED 2006; 21:S93–101.

Consistent with strategic planning processes led by the Department of Veterans Affairs (VA) to ensure that increasingly scarce resources are invested in areas of highest priority, the VA Office of Research & Development (ORD) has initiated a process of analyzing and evaluating its research portfolio. In recognition of the changing demographics and the special health care needs of women, the ORD has assigned research on women's health a high priority and it is one of the first topics to undergo such review. Over the last decade, the VA has built an increasingly productive portfolio of research in all 4 of its Research and Development Services (Biomedical Laboratory, Clinical Science, Rehabilitation, and Health Services), with significant potential to improve the health of women veterans. The purpose of this paper is to summarize the VA's current research efforts related to women's health, describe the agenda-setting process, and present the resulting national VA Women's Health Research Agenda.

VA WOMEN'S HEALTH RESEARCH AGENDA-SETTING PROCESS

In early 2004, the VA Office of Research & Development tasked VA HSR&D Service with oversight of the development of the first national VA Women's Health Research Agenda that would span all 4 Research and Development Services. Representatives from across the country with demonstrated track records in VA Women's Health Research were invited to join a national planning group, create an agenda-setting plan, and enact it. The Planning Group developed a 4-step action plan, designed to meet the health care needs of women veterans and position VA as a national leader in Women's Health Research (Table 1).[1]

Appraisal of the VA's Research Portfolio

As of fiscal year 2003, funding of Women's Health Research to VA-based investigators totaled $27.9 million for 273 studies (National Institutes of Health [NIH], foundation, private, other federal and government, and VA funding combined), constituting 2.6% of all funding reported by VA investigators ($1.08 billion). Although the absolute amount of funding increased from 2000 to 2003, there was a decline in the overall proportion of Women's Health Research funding in relation to total

There are no conflicts of interest to declare.

Address correspondence and requests for reprints to Elizabeth M. Yano, PhD: Deputy Director, VA Greater Los Angeles HSR&D Center of Excellence, VA Sepulveda Ambulatory Care Center, 16111 Plummer Street (Mailstop 152), Sepulveda, CA 91343 (e-mail: elizabeth.yano@va.gov).

Table 1. Four-Step Action Plan Toward a VA Women's Health Research Agenda

Action Plan	Approach	Outcome
Step #1: Critically appraise the VA research portfolio	Obtain and review history of funding among VA researchers	Conducted portfolio analysis of VA women's health research funding
	Analyze data by types of funding (e.g., VA, other federal, private, foundation)	
	Analyze data by research area (e.g., aging, mental health, reproductive health)	Established methods for ongoing monitoring of VA women's health research portfolio
Step #2: Obtain systematic information about the health and health care of women veterans to provide an evidence base for the research agenda	Conduct gender-specific analyses of an array of VA secondary data	Obtained analyses from 15+ centers in support of agenda-setting process
	Conduct a systematic women veterans' literature review	Completed a systematic literature review
		Developed a web-based bibliography
Step #3: Based upon gaps between the current VA research portfolio (Step #1) and the assessment of the evidence base (Step #2), identify strategic priorities for the VA women's health research agenda	Adapt priority-setting strategies used by other agencies (e.g., NIH Office of Research on Women's Health, AHRQ)	Created a compendium of women's health research initiatives
	Review VA strategic plans (e.g., Women Veterans Health Program, Advisory Committee for Women Veterans)	Conducted the first national VA women's health agenda-setting conference (Nov 2004)
	Conduct gap analysis, priority-setting, and consensus development in an agenda-setting conference	Disseminated web-based products from all steps
Step #4: Foster the conduct of VA women's health research	Build research capacity through improved collaboration, networking, and mentoring	Created a VA women's health research website*
	Solve methodologic challenges (e.g., small sample sizes)	Created a Listserv for use by VA women's health researchers
	Increase awareness and visibility of VA women's health research	Develop web- and cyber-based educational modules on key methodologic issues

**VA women's health research website (http://www.va.gov/resdev/programs/womens_health/) contains background information on women veterans, summaries, and online presentations from the VA women's health research conference, and other useful announcements and links.*

funding (Table 2). The majority of funding for Women's Health Research among VA investigators was from NIH sources (Fig. 1). The VA-funded portion, $6.9 million, amounted to about 25% of each dollar spent on Women's Health Research in 2003, or 1.9% of the $366.9 million total VA research funding. VA increased its investment by almost $1 million in Women's Health Research between 2002 and 2003, while total funding for Women's Health Research by all funding sources and total VA funding declined in the same period (Table 2). The categories with the highest funding included chronic diseases, aging,

Table 2. Women's Health Research Funding Portfolio Among VA Investigators

Year	Total Funding	Women's Health Funding	Percent
Women's Health Research as percent of total research funding (all funders)			
2000	$821,032,693	$25,680,259	3.1
2001	$928,540,420	$28,896,057	3.1
2002	$1,010,795,393	$30,788,988	3.0
2003	$1,079,979,025	$27,933,800	2.6
Women's Health Research as percent of total research funding (VA only)			
2000	$313,967,151	$6,074,963	1.9
2001	$333,564,215	$6,097,515	1.8
2002	$370,290,877	$6,018,916	1.6
2003	$366,908,456	$6,931,449	1.9
Year	**Total Number of Projects**	**Number of New Projects**	**Number of Investigators**
Numbers of Women's Health Research total projects, new projects and investigators (all funders)			
2000	309	117	212
2001	319	109	213
2002	301	91	204
2003	273	65	192

Source: U.S. Department of Veterans Affairs, Office of Research and Development (ORD), 2004.

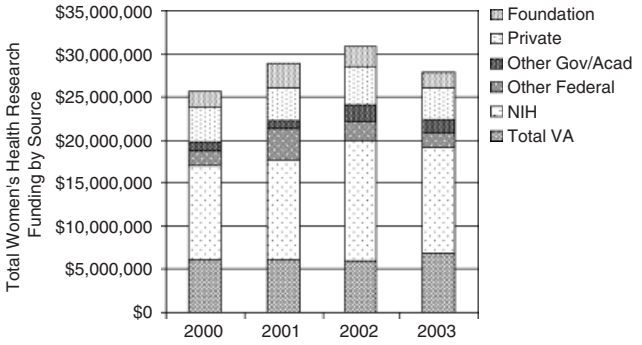


FIGURE 1. Total women's health research funding by source.

breast cancer, and osteoporosis (Fig. 2). VA's research investment was highest in mental health, where it exceeded non-VA funding (67% of total). Little Women's Health Research has been funded on substance abuse, cancers other than breast, or Alzheimer's disease.

We also examined VA-funded Women's Health Research. Some of the VA's hallmark studies include the National Vietnam Veterans Readjustment Study, which included women veterans[2–6]; studies of the impact of military environmental exposure on reproductive outcomes among U.S. women Vietnam veterans[7,8]; the first national assessment of the health status and effects of military service on self-reported health among women veterans who use VA ambulatory care[9–13]; analyses of the large survey of veterans, which included over 30,000 women veterans[14]; and an evaluation of the surgical risks and outcomes of women treated in VA hospitals.[15–17] Table 3 presents highlights of recent VA Women's Health Research.

Establishing the Evidence Base for Agenda Development

One of the goals of the VA research agenda-setting process was to build a systematic evidence base that supported the alignment of VA research priorities with the health-related needs of women veterans. We used 2 strategies to accomplish this goal, which included (1) capitalizing on VA's extensive clinical and administrative data repositories to conduct gender-specific analyses[18] and (2) conducting a systematic literature review and synthesis through a partnership with the Southern California Evidence-Based Practice Center.

Secondary Analyses of VA Data. Our objective was to identify high-prevalence, high-cost, high-impact conditions among women veterans, as well as conditions with disproportionate burden among women (e.g., obesity, incontinence, osteoporosis) or with distinct clinical presentations in women (e.g., coronary artery disease). We began by listing the available data sources for conducting queries by gender (Table 4). Over 15 research centers responded to our requests for gender-specific analyses of existing data, demonstrating both the capacity and commitment to furthering the VA Women's Health Research Agenda. While the results of these secondary analyses are too numerous to cover here, subsequent priority-setting was informed by the most prevalent diagnoses (e.g., post-traumatic stress disorder [PTSD], arthritis, chronic low back pain, hypertension, chronic lung disease, depression), most commonly prescribed drugs (e.g., simvastatin, levothyroxine, lisinopril), and gender comparisons in patient satisfaction, quality, and costs of care. This process highlighted that these data sources had been underutilized in the past, demonstrating substantial opportunities for additional analyses.

Systematic Literature Review. The Office of Research & Development commissioned the conduct of a systematic literature review to develop a synthesis of what is known about women veterans' research.[19] The resulting review pointed to gaps in knowledge about specific health risks among women veterans, quality of care, and treatments for PTSD and other conditions of high prevalence among women veterans. A full bibliography is available on request.

Achieving Consensus on Research Priorities

Several governmental agencies and private organizations are committed to the advancement of Women's Health Research

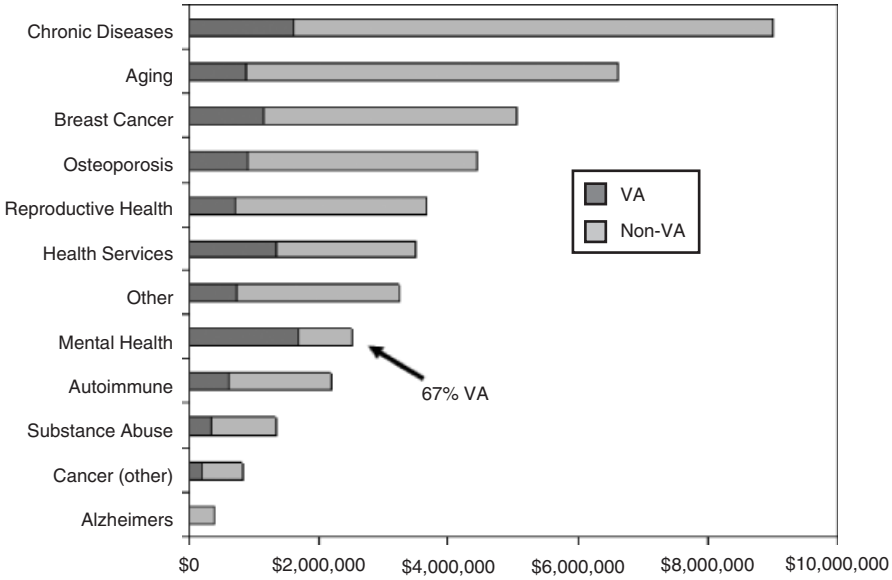


FIGURE 2. Total fiscal year 2003 women's health research funding by disease area and funding source (VA and non-VA).

Table 3. Highlights of Recent VA Women's Health Research*

VA Research Service	Highlights of Recent Research
VA Biomedical Laboratory Research and Development	Identification of a new synthetic estrogen-like compound that reverses bone loss in mice without the reproductive side effects of conventional hormone replacement therapy Prevention of a disease resembling multiple sclerosis in female mice through a combined therapy of estrogen and a T-cell receptor vaccine Association of systemic lupus erythematosus (SLE) with prolactin, a pituitary hormone that increases during pregnancy, leading to bromocriptine, a prolactin suppressant, as a potential treatment for SLE
VA Clinical Science Research and Development (including the Cooperative Studies Program)	Randomized clinical trial of treatment for PTSD in women veterans (jointly funded by Department of Defense): First multi-site VA clinical trial focused only on women 12 sites, 284 women veterans and active duty military enrolled Evaluating efficacy of a type of cognitive behavioral therapy for treating PTSD
VA Rehabilitation Research and Development	Animal model of stress urinary incontinence being developed and tested to develop new strategies for treatment and prevention Estrogen treatment at time of initial injury (rather than prior to or after injury) may facilitate functional recovery and regeneration of injured nerves A functional virtual reality model of the pelvic floor and organs being developed for education, simulation of surgical outcomes, and planning complex surgical procedures
VA Health Services Research and Development	Evaluation of the prevalence and risks of problem drinking among women veterans Influence of PTSD, depression, and military sexual assault on physical health/function Comparison of patient satisfaction in different VA women's health care models Assessment tool and intervention to enhance gender-aware VA health care Identification of women veterans' ambulatory care use, barriers, and influences

All VA-funded studies are searchable online at http://www1.va.gov/resdev/.

both within and outside the VA. To assure that our approach and priorities were set within the context of the substantial work accomplished by others, the Planning Group adapted themes and strategies used by other agencies and organizations to develop VA's research priorities in combination with empirical evidence regarding patterns of disease burden among women veterans. These included, for example, the NIH Office of Research on Women's Health,[20] the Defense Women's Health Research Program,[21] the TriService Nursing Research Program,[22] and the Society for Research on Women's Health.[23] We also reviewed strategic planning and advisory documents from the Women Veterans Health Program,[24] and obtained research recommendations from the Advisory Committee for Women Veterans,[25] and the Defense Advisory Committee on Women in the Services[26] through the Center for Women Veterans.

We combined results from the work of these other groups, our own appraisal of the gaps between the current VA research portfolio (Step #1), and the assessment of the evidence base (Step #2) and presented them as a foundation for an agenda-setting conference held in November 2004. Over 50 VA and non-VA Women's Health Researchers were invited to participate in the consensus development effort to synthesize information from overviews presented by VA leaders, followed by a series of presentations summarizing advance work completed by planning group members.[27] Participants were subsequently divided into 5 workgroups (biomedical, clinical, rehabilitation, health services, and infrastructure), each with a planning group moderator to help them cull the presented information and generate research priorities and solicitation topics. Workgroups then reconvened as a whole, presented their recommendations, and received expert input from a panel of senior VA leaders in operations and research, and Women's Health Research experts at the NIH Office of Research on Women's Health[28–29] and the Agency for Healthcare Research and Quality (AHRQ).[30]

VA'S WOMEN'S HEALTH RESEARCH AGENDA

Biomedical Laboratory Research Priorities

The Biomedical Workgroup established research on sex-based influences on prevention, induction, and progression of diseases relevant to women veterans as their overarching focus. Based on current evidence of the prevalence of conditions among women veterans, the Biomedical research priorities focused on (1) *mental health* (especially PTSD, stress, addiction, sexual trauma, and depression), (2) *military occupational hazards* (focused on injury and rehabilitation, wound healing, tissue remodeling, vaccine development, and biological and chemical exposures), (3) *chronic diseases* (with emphasis on diabetes, infections, autoimmunity, osteoporosis, arthritis, and chronic pain), (4) *cancer* (focused on etiology and response to treatment for exposure-related cancers), and (5) *reproductive health* (including fertility, contraception, and menopausal issues).

Because many of these priorities overlap with programmatic themes of the NIH Office of Research on Women's Health, VA researchers will need to remain apprised of advances and opportunities that cross agency lines. Nonetheless, VA has unique strengths that will facilitate the advancement of novel biomedical research.

Clinical Science Research Priorities

The Clinical Science Workgroup focused on the relative paucity of reliable epidemiologic data on women veterans, spanning from risks and exposures before entry into the military, through military experience and exposures, to status after military discharge regardless of their ultimate choice of care provider (VA or not VA). While the Department of Defense (DoD) has established inception cohorts of female veterans, access to these data for the purposes of linking past exposures forward through their veteran years has been problematic. Moreover,

Table 4. Selected Data Sources Available for the Assessment of Women Veterans' Health and Health Care

Data Source	Description
National Survey of Veterans (NSV)*	Approximately decennial survey (2001, 1992, etc.) conducted among random digit dial (RDD) veteran samples augmented by VA administrative lists of VA users to provide national estimates for veterans overall and for key subgroups (n=20,048, 2001 NSV) Contains sociodemographic characteristics, period of service, combat exposure, insurance coverage, VA and non-VA health care utilization, health status, functional limitations, health conditions, eligibility, and more
Survey of Healthcare Experiences of Patients (SHEPs)†	Adapted from earlier annual patient satisfaction surveys launched in the mid 1990s Random samples of veteran users of VA services Contains data on health status, quality of life, health care utilization, satisfaction, etc.
External Peer Review Program (EPRP)†	Part of VHA's performance measurement system composed of externally abstracted medical record data from randomly sampled records of VA users at each VA facility Patient- and facility-level data on chronic disease quality (e.g., foot sensation exams among diabetics) and preventive practice (e.g., flu shots)
Large Survey of Veterans†	National survey sample of veteran users of VA health care (includes about 33,000 women) (1999) Includes health status, conditions, satisfaction, utilization, quality of life, etc.
VHA Medical SAS Datasets (utilization data)‡	National administrative data for VA-provided health care used primarily by veterans, but also by some non-veterans (e.g., employees) (housed in Austin Automation Center) Includes medical inpatient data (acute, extended, observation, non-VA) organized by stay, bedsection, procedures and surgeries; outpatient data (visits and events); long-term care data (representing an array of services in VA nursing homes, community nursing homes, domiciliaries, home-based primary care, home health care, etc.)
VA-Medicare Data‡	Linked VA and Medicare health care utilization data, including Part A and B claims, and patient and provider information files
Pharmacy Benefits Management (PBM) Program‡	Prescription information for all VA patients who obtain their prescriptions within the VA system (available from FY 1999 through present) Useful for studying prescribing habits, drug utilization trends, and pharmacoeconomics
Decision Support System (cost data)‡,§	Contains data on the cost of care of every individual patient care encounter in VA Starts from fiscal year 1998 in national extracts organized by inpatient discharges, inpatient treating specialty files, and outpatient files
National Surgical Quality Improvement Program (NSQIP)‖	National, validated, outcomes-based, risk-adjusted program for measurement and enhancement of surgical care (begun in 1991) Currently incorporates all VAMCs and 14 private hospitals with extensive surgical data
Quality Enhancement Research Initiative (QUERI) Centers¶	National initiative to translate research into practice organized around specific conditions (diabetes, mental health, ischemic heart disease, spinal cord injury, HIV/AIDS, colorectal cancer, stroke, substance abuse) Selected QUERIs have developed patient registries allowing for disease-based analyses
Women Veterans Health Program (WVHP) Evaluation#	Organizational data at the VISN, VA medical center, and practice levels, with multiple perspectives available at the practice-level Includes structure, staffing, leadership, authority, resource sufficiency, etc.

Full survey final report and national frequencies online at http://www.va.gov/vetdata/SurveyResults/final.htm.
†*Available through formal Data Use Agreements with the VA Office of Quality and Performance (OQP).*
‡*VA Information Resource and Education Center (VIREC) (http://www.virec.research.med.va.gov).*
§*VA Health Economics Resource Center (HERC) (http://www.herc.research.med.va.gov).*
‖*VA National Surgical Quality Improvement Program (http://www.nsqip.org).*
¶*VA Quality Enhancement Research Initiative (QUERI) (http://www.hsrd.research.va.gov/queri/) (links to individual QUERI Centers also available through this web address).*
#*Women Veterans Health Program (WVHP) Office (http://www1/va/gov/wvhp).*

few VA clinical studies have been conducted among women veterans, hindered mainly by the small numbers of women at individual facilities. Priority recommendations included creating data use agreements that facilitate VA researchers' access to DoD databases on military women. Barring that, creation of a prospective cohort of women upon discharge from the military (i.e., when they become veterans) should be pursued to build the necessary foundation for future VA research.

In the interim, the Clinical Sciences Workgroup identified special conditions and populations on whom VA clinical research should be focused, including (1) pregnancy and fertility issues, (2) returning military and reservists, (3) long-term care, (4) substance abuse and mental health, (5) homelessness, (6) PTSD and military sexual trauma, and (7) recent amputees.

Rehabilitation Research Priorities

The VA's Rehabilitation Research and Development (RR&D) Service spans biomedical, clinical, and health services research in service of maximizing function and quality of life (including vocational outcomes), preventing and treating secondary complications, and addressing psychosocial issues associated with disability and recovery. The Rehabilitation Workgroup established 6 priority conditions/diseases, focused on the rehabilitative aspects associated with (1) arthritis, (2) chronic pain, (3) obesity, (4) osteoporosis/fall-related injuries, (5) amputation (specifically, socket-fit technology), and (6) reproductive challenges for disabled women veterans. While some of these priorities are shared by NIH, VA's unique contributions include prosthetics (e.g., menstrual cycle/limb volume variability and socket-fit for amputees) and rehabilitation engineering (e.g., assistive technologies among women with disabilities; gender-specific technologies for urinary incontinence). Because of VA's investment in centralized administrative and clinical databases, VA researchers are also well-positioned to explore gender differences in chronic pain and obesity in relation to rehabilitation outcomes. Given the rehabilitation demands of the injuries incurred by women veterans who have served in

Table 5. Improving the Infrastructure for Enhancing VA Women's Health Research

Research Barriers	Strategic Problems	Solutions
Small number of women in the VA system	Lack of network or infrastructure to facilitate research	Develop VA practice-based research networks among VA facilities with adequate women veteran samples (e.g., recruit sites with a VA Comprehensive Women's Health Center or other large caseload sites)
	Inadequate knowledge and familiarity with small-sample study designs and statistics	Improve familiarity and use of VA Cooperative Studies Program (CSP) for multi-site trials (invite CSP staff to present to VA women's health research audiences in different venues)
	Lack of reviewer knowledge of small-sample issues	Address skill and knowledge deficits through educational programs (e.g., reviewer training, research briefs to investigators, seminars linked to VA meetings, web- and cyber-education) Create methodologic briefs for distribution (similar to VA HSR&D Management Briefs)
Identifying women veterans who do not use the VA	Lack of reliable, valid, updated women veterans registry	Coordinate recruitment and enrollment of recently discharged veterans through the Transitional Assistance Program (TAP) (i.e., advertise VA women's health research)
	HIPAA restrictions on accessing data enabling research across settings	Develop and maintain an updated women veterans' registry from military discharge forward Identify non-VA databases that identify veteran status and foster inclusion of veteran status in those without such indicators Forge VA-DoD research partnerships enabling VA researchers to build on active duty research and offering DoD researchers opportunities to conduct longitudinal research
Problems with secondary databases	Lack of coordination with other agencies (including difficulty obtaining non-VA funds to study veterans, inadequate outside understanding of VA relevance)	Forge VA-DoD research partnerships and data sharing agreements to improve VA investigator access
	Lack of relevant variables in centralized data sources	Develop mechanisms to link VA data to other non-VA databases (model after VA/Medicare data merge)
	Lack of knowledge on available data	Incorporate more gender-specific measures in centralized data collection efforts and database composition
	Need for information about outside VA use (including contract care)	Increase degree of over-sampling of women veterans in ongoing data collection efforts (e.g., Office of Quality & Performance chart-based or survey data) Enhance use of gender-specific data by routinely distributing aggregated data by gender Distribute information about data sources that may be used for assessment of women veterans' health and health care issues (e.g., Listserv, weblinks from the VA R&D Women's Health Research site to VA datasets or resource centers)
Perceived barriers to conducting and publishing VA women's health research	Negative attitudes and misperceptions about women veterans' research	Produce regular women veterans' research updates to larger VA research community
	Pressures on clinician investigators (need for protected time and methodologic supports)	Assess barriers faced by clinician investigators and evaluate options for leveraging time
	Reviewers within and outside VA with limited knowledge of women veterans' health issues	Partner clinician investigators with doctorally trained researchers where possible
	Competing research investment demands and constrained budgets	Add women's health researchers to VA scientific review committees
		Provide all VA research reviewers with training on issues relevant to women veterans' research, including briefings on small sample size research designs, statistics, and analysis (or consider separate review groups with needed expertise) Leverage existing research funding by providing access to administrative supplements for studies that will add women or female specimens or animals to do gender comparisons Provide planning funds to researchers to determine feasibility and strategies to recruit adequate sample size or specimen quality Foster creation of VISN pilot funds for research projects that evaluate women veterans' health and health care

Iraq and Afghanistan, opportunities for using merged DoD-VA data in service of research capable of improving their quality of care are being missed. They also recommended joint agency requests-for-applications (RFAs), for example, between the VA and the National Institute of Disability and Rehabilitation Research or within-VA initiatives, for example, between RR&D and the VA's Quality Enhancement Research Initiative (QUERI), leveraging resources and expertise to improve women veterans' health and health care related to disabling stages of QUERI conditions (e.g., stroke).

Health Services Research Priorities

The Health Services Workgroup focused on development of 2 targeted RFAs, 1 on evaluation of models for delivery of women veterans' health care, and another fostering needs assessment projects. The core goals for delivery model studies focused on the need to measure the quality associated with different care models serving women veterans, including, for example, evaluations by setting (e.g., large VA medical centers vs. small community-based outpatient clinics); by type of provider (e.g., among fee-basis or contract providers, same-gender providers) and to evaluate the quality, costs, access, and continuity tradeoffs women veterans face in different care settings and for different health conditions (e.g., mental health, specialty care, gender-specific services). Benchmarking VA-based access and quality to services outside the VA is also a priority to ensure equitable care provision. The Workgroup called for needs assessment for high-impact conditions, including psychiatric/emotional disorders and military-specific exposures, assessments of women veterans' needs and preferences for health services and their care environment, gender-specific barriers to access (including issues related to service connection), and better epidemiologic data on their disease burden and utilization patterns. Selected on the basis of their likely impact on health-related quality of life, high-priority conditions included the following:

- Psychiatric/emotional health
- Reproductive health/infertility/pregnancy
- Military-specific exposures
- Bone and musculoskeletal diseases
- Chronic pain
- Behavioral health (e.g., drugs, alcohol, tobacco, stress-related)
- Obesity/metabolic syndrome/diabetes
- Thyroid disorders
- Urinary incontinence
- Menstrual disorders/menopausal symptoms
- Oral health
- Eye/vision problems

Building an Infrastructure for Fostering the Conduct of VA Women's Health Research

At all stages, the need to build an effective infrastructure for fostering the conduct of VA Women's Health Research was deemed central to the success of the resulting agenda. In particular, while several pioneering VA researchers interested in exploring women veterans' health research have made significant inroads in contributing to our knowledge base over the past decade, anecdotal stories about perceived barriers to conducting, and publishing research about women veterans challenged us to ascertain their prevalence.

Conference participants were therefore asked to complete a brief barriers survey before the conference to permit time for analysis and feedback (85% response rate, $n=28$). VA-based Women's Health Research was roughly split between the study of nonveteran women (61%) and veteran women who used the VA (57%). (Note: Conferees could report more than 1 type of research, resulting in sums over 100%.) Over a quarter (28%) conducted research involving women veterans who do not use VA health care; only 18% had conducted research on women in the military. Only 18% had done research on biomedical samples taken from women and 14% on animal studies related to gender issues, although these figures also reflect the distribution of survey respondents (18% were biomedical researchers).

The top 5 perceived barriers to conducting VA Women's Health Research were cited as: (1) the lack of a network of VA facilities to recruit women veterans for research studies, (2) difficulty in identifying women veterans who do not use the VA, (3) lack of coordination with other agencies (e.g., DoD), (4) lack of availability of needed variables in centralized databases, and (5) low numbers of women veterans overall. These results were reported to all conference participants and provided to the Infrastructure Workgroup for discussion and suggestions for resolution.

Details for resolving each identified barrier are listed in Table 5. Central to building the needed infrastructure is the development of VA practice-based research networks akin to those cultivated by AHRQ for primary care research, but among sites with larger caseloads of women veterans to facilitate recruitment efforts. Considerable education of the field (i.e., reviewers, investigators, non-VA research partners) is also needed to publicize the opportunities and demand for more VA Women's Health Research, as well as solutions to some of the methodologic challenges, such as the Institute of Medicine's brief on small sample size methods and their role in advancing research. The value of and potential role for interagency collaborations is substantial, for example, with DoD to conduct longitudinal research that builds on military cohorts, and with the National Center for Health Statistics to integrate veteran status into national surveys, as AHRQ does in the Medical Expenditure Panel Survey. Finally, backing the agenda with new funding is key. VA HSR&D Service has already published a new Women's Health solicitation, while planning grants, pilot funds, and administrative supplements to add women (or female specimens) to existing studies were proposed to accelerate and promote greater inclusion of women.

Building a consortium of researchers committed to women veterans' health research within VA and through university and other partnerships is a crucial next step. The agenda-setting conference was an important first step in this regard, building on existing ties across VA and non-VA organizations and creating new ones. The VA research website has already fostered new collaborations and mentoring relationships, while providing access to a searchable database of VA investigators, funded studies and publications. Access to VA datasets has been enhanced through data use agreements and technical consultation with 1 or more VA resource centers, such as the VA Information Resource and Education Center. While leading VA-funded research still requires a 5/8th VA appointment, non-VA researchers commonly collaborate with VA-based researchers, capitalizing on special expertise and common interests to pursue a broad range of research studies, whereas other agencies (e.g., National Cancer Institute) also fund women veterans' research, providing additional venues for non-VA researchers to directly contribute to this growing field.

CONCLUSIONS

Using a systematic evidence base and consensus development process among stakeholders within and outside the VA, we report on the first national VA Women's Health Research Agenda. While the VA made women's health a research priority in the

early 1990s, enabling the development and funding of an important array of studies that spanned the research spectrum from bench-to-bedside, we anticipate that the level of commitment and strategic planning of this agenda-setting effort has the potential to serve as a strong foundation for the next decade of women veterans' health research. The processes used to set research priorities also have important implications for improving research management across diverse programs, particularly in reference to special populations.

To effectively foster the conduct and expansion of Women's Health Research in VA, the consensus was that the VA Office of Research & Development needs to build research capacity, solve methodologic issues that limit participation of women in research, and increase the awareness and visibility of VA Women's Health Research. Building bridges to research partners at agencies with longstanding commitments to advancing women's health and improving gender equity will continue to invigorate the VA research process.[31,32]

While the VA mandated inclusion of women in all VA studies in 1983, our assessment of funded research suggests that compliance has been less than optimal but may not be correctable without assistance to researchers to recruit greater numbers of women veterans into their studies. Given fiscal realities of constrained federal budgets at the same time new veterans are entering the system, we offer some innovative solutions to leverage system resources and talents of the VA's many investigators and their partners in other systems.

This work was funded by the Department of Veterans Affairs' Office of Research & Development (Project # CSF 04-376) and the VA HSR&D Center for the Study of Healthcare Provider Behavior (Project # HFP 94-028) and overseen by the VA Health Services Research & Development (HSR&D) Service. The project was approved by the IRB at the VA Greater Los Angeles Healthcare System. We are indebted to John Demakis, MD, Shirley Meehan, PhD, MBA, Martha Bryan, EdD, and Serena Chu, PhD, in VA HSR&D Service for planning support; Ismelda Canelo, Sam Garcia, Lisa Tarr, Lorena Barrios, Allyson Szabo, and Vera Snyder-Schwartz at the VA Greater Los Angeles HSR&D Center of Excellence for project support; Brigadier General Wilma L. Vaught (USAF, Ret.), former Department of Veterans Affairs' Chief of Staff Nora Egan, Deputy Under Secretary for Health Michael Kussman, MD, and Deputy Under Secretary for Health for Operations and Management Laura Miller, CHE, MBA, for their leadership support; and Rosaly Correa-de-Araujo, MD, PhD, of the Agency for Healthcare Research and Quality and Loretta Finnegan, MD, of the NIH Office of Research on Women's Health, for serving as conference panelists. We would also like to acknowledge the input of all conference participants. This work was partially presented and generated at the National VA Women's Health Research Agenda-Setting Conference held in Arlington, Virginia, November 8–9, 2004. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

REFERENCES

1. **Yano EM.** Toward a VA women's health research agenda. SGIM Forum. 2004, p. 4, 9.
2. **Jordan BK, Schlenger WE, Hough R, et al.** Lifetime and current prevalence of specific psychiatric disorders among Vietnam veterans and controls. Arch Gen Psychiatry. 1991;48:206–15.
3. **Fontana A, Schwartz LS, Rosenheck R.** Posttraumatic stress disorder among female Vietnam veterans: a causal model of etiology. Am J Public Health. 1997;87:169–75.
4. **Zatzick DF, Weiss DS, Marmar CR, et al.** Post-traumatic stress disorder and functioning and quality of life outcomes in female Vietnam veterans. Mil Med. 1997;162:661–5.
5. **King DW, King LA, Foy DW, Gudanowski DM.** Prewar factors in combat-related posttraumatic stress disorder: structural equation modeling with a national sample of female and male Vietnam veterans. J Consult Clin Psychol. 1996;64:520–31.
6. **King LA, King DW, Fairbank JA, Keane TM, Adams GA.** Resilience-recovery factors in post-traumatic stress disorder among female and male Vietnam veterans: hardiness, postwar social support, and additional stressful life events. J Pers Soc Psychol. 1998;74:420–34.
7. **Kang HK, Mahan CM, Lee KY, Magee CA, Mather SH, Matanoski G.** Pregnancy outcomes among U.S. women Vietnam veterans. Am J Ind Med. 2000;38:447–54.
8. **Kang HK, Mahan CM, Lee KY, Magee CA, Selvin S.** Prevalence of gynecologic cancers among female Vietnam veterans. J Occup Environ Med. 2000;42:1121–7.
9. **Hankin CS, Skinner KM, Sullivan LM, Miller DR, Frayne S, Tripp TJ.** Prevalence of depressive and alcohol abuse symptoms among women VA outpatients who report experiencing sexual assault while in the military. J Trauma Stress. 1999;12:601–12.
10. **Frayne SM, Skinner KM, Sullivan LM, et al.** Medical profile of women Veterans Administration outpatients who report a history of sexual assault occurring while in the military. J Womens Health Gend Based Med. 1999;8:835–4.
11. **Skinner KM, Furey J.** The focus on women veterans who use Veterans Administration health care: the Veterans Administration women's health project. Mil Med. 1998;163:761–6.
12. **Skinner KM, Kressin N, Frayne SM, et al.** Prevalence of military sexual assault among female Veterans' Administration outpatients. J Interpers Violence. 2000;15:291–310.
13. **Barnard K, Frayne SM, Skinner KM, Sullivan LM.** Health status among women with menstrual symptoms. J Womens Health (Larchmt). 2003; 12:911–9.
14. **Frayne SM, Seaver MR, Loveland S, et al.** Burden of medical illness in women with depression and posttraumatic stress disorder. Arch Intern Med. 2004;164:1306–12.
15. **Weaver F, Hynes DM, Goldberg JM, Khuri S, Daley J, Henderson W.** Hysterectomy in Veterans Affairs Medical Centers. Obstet Gynecol. 2001;97:880–4.
16. **Weaver F, Hynes DM, Hopkinson W, et al.** Preoperative risks and outcomes of hip and knee arthroplasty in the Veterans Health Administration. J Arthroplasty. 2003;18:693–708.
17. **Hynes DM, Weaver F, Morrow M, et al.** Breast cancer surgery trends and outcomes: results from a national Department of Veterans Affairs study. J Am Coll Surg. 2004;198:707–16.
18. **Lamoreaux J.** The organizational structure for medical information management in the Department of Veterans Affairs: an overview of major health care databases. Med Care. 1996;34(suppl 3):MS31–44.
19. **Goldzweig CL, Balekian T, Rolon C, Yano EM, Shekelle PG.** The state of women veterans' health research: results of a systematic review. J Gen Intern Med. 2006;21(3):S82–S92.
20. **U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health.** Agenda for Research on Women's Health for the 21st Century: A Report of the Task Force on the NIH Women's Health Research Agenda for the 21st Century, Volume 1. Executive Summary. Bethesda, MD: NIH Publication No. 99-4385, 1999
21. **U.S. Department of Defense, Congressionally Directed Medical Research Programs.** Defense Women's Health Research Program. Available at: http://cdmrp.army.mil/annreports/1999annrep/section10.htm. Accessed January 28, 2005.
22. **Institute of Medicine.** The Program for Research in Military Nursing: Progress and Future Directions. Washington, DC: National Academy Press; 1996.
23. **Haseltine FP, Jacobson BG.** Women's Health Research: A Medical and Policy Primer. Washington, DC: American Psychiatric Press; 1997.
24. **U.S. Department of Veterans Affairs, Women Veterans Health Program.** Final Report of the National Women Veterans Health Strategic Work Group. Washington, DC: Women Veterans Health Program; 2002.
25. **U.S. Department of Veterans Affairs.** Advisory Committee for Women Veterans. Available at: http://www.va.gov/womenvet/. Accessed February 12, 2004.

26. **U.S. Department of Defense.** Defense Advisory Committee on Women in the Services (DACOWITS). Available at: http://www.dtic.mil/dacowits/. Accessed January 25, 2005.
27. **U.S. Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development (ORD).** VA R&D Women's Health Research Conference. Available at: http://www1.va.gov/resdev/programs/womens_health/conference/default.cfm. Accessed December 18, 2004.
28. **Finnegan LP.** The NIH Women's Health Initiative: its evolution and expected contributions to women's health. Am J Prev Med. 1996;12:292–3.
29. **Riley PL, Finnegan LP.** Observations from the CDC: the prevention research centers program: collaboration in women's health. J Womens Health. 1997;6:281–3.
30. **Correa-de-Araujo R.** A wake-up call to advance women's health. Women's Health Issues. 2004;14:31–4.
31. **Bierman AS.** Climbing out of our boxes. Advancing women's health for the twenty-first century. Women's Health Issues. 2003;13:201–3.
32. **Satcher D.** American women and health disparities. JAMWA. 2001;56:131–3.

Voices of Women Veterans (continued)

VA HEALTH CARE EXPERIENCES

``In a real time of need, the VA has by the Grace of God come to my aid. After a 30 year marriage, my husband wanted a divorce - I lost my home and my business - my retirement and all insurance I had. Unbeknownst to me, I was in fact eligible for health care. . ..''

``When I was leaving active duty, I went to a VA counselor who told me that the highest possible disability rating I could get would be 10%, and that was improbable. So for several years, I forgot about the VA. My sister encouraged me to try again, and a DAV rep helped me get a 70% rating. Ever since then, I have been thrilled with the medical care I have received. I have never been made to feel rushed or unimportant.''

``As far as health care through the VA system, I could not have asked for better. The doctors and all other staff are well-trained, knowledgeable, and most of all caring.''

``Prior to finding out about the women's clinic within the VA, I did not use the VA because getting appointments was a hassle.''

``I never knew until the middle 90s that I could get health care at the VA. Two and a half years ago, I re-injured my back and had a lot of trouble getting help at the VA. I was sent to and finally paid for outside care on my own.''