

Teaching Residents Evidence-based Medicine Skills

A Controlled Trial of Effectiveness and Assessment of Durability

Christopher A. Smith, MD, Pamela S. Ganschow, MD, Brendan M. Reilly, MD, Arthur T. Evans, MD, MPH, Robert A. McNutt, MD, Albert Osei, MD, Muhammad Saquib, MD, Satish Surabhi, MD, Sunil Yadav, MD

OBJECTIVES: To measure the effectiveness of an educational intervention designed to teach residents four essential evidence-based medicine (EBM) skills: question formulation, literature searching, understanding quantitative outcomes, and critical appraisal.

DESIGN: Firm-based, controlled trial.

SETTING: Urban public hospital.

PARTICIPANTS: Fifty-five first-year internal medicine residents: 18 in the experimental group and 37 in the control group.

INTERVENTION: An EBM course, taught 2 hours per week for 7 consecutive weeks by senior faculty and chief residents focusing on the four essential EBM skills.

MEASUREMENTS AND MAIN RESULTS: The main outcome measure was performance on an EBM skills test that was administered four times over 11 months: at baseline and at three time points postcourse. Postcourse test 1 assessed the effectiveness of the intervention in the experimental group (primary outcome); postcourse test 2 assessed the control group after it crossed over to receive the intervention; and postcourse test 3 assessed durability. Baseline EBM skills were similar in the two groups. After receiving the EBM course, the experimental group achieved significantly higher postcourse test scores (adjusted mean difference, 21%; 95% confidence interval, 13% to 28%; $P < .001$). Postcourse improvements were noted in three of the four EBM skill domains (formulating questions, searching, and quantitative understanding [$P < .005$ for all], but not in critical appraisal skills [$P = .4$]). After crossing over to receive the educational intervention, the control group achieved similar improvements. Both groups sustained these improvements over 6 to 9 months of follow-up.

CONCLUSIONS: A brief structured educational intervention produced substantial and durable improvements in residents' cognitive and technical EBM skills.

KEY WORDS: evidence-based medicine; clinical trial; graduate medical education; internship and residency.

J GEN INTERN MED 2000;15:710-715.

Clinicians recognize the potential value of evidence-based medicine (EBM) but many feel ambivalent about its use in patient care. Major barriers to implementation include a perception that the appeal of EBM is more academic than practical, and clinicians cite their own lack of EBM skills as a contributing factor.¹

Many postgraduate training programs currently attempt to teach EBM skills, primarily in journal clubs and through didactic lectures. However, published outcomes of these efforts have shown only limited effectiveness.²⁻⁸ Most of the studies have focused on critical appraisal while excluding or inadequately measuring other EBM skills.⁹ Consequently, there are no published studies that assess how best to teach more fundamental EBM skills: the ability to create searchable questions that arise from knowledge gaps in patient care and the ability to perform efficient literature searches to answer those questions.¹⁰ The acquisition of these skills, in addition to those of critical appraisal, is essential if EBM is to favorably influence patient outcomes.¹¹⁻¹³

We conducted a trial to test whether an educational program could improve four essential EBM skills among first-year internal medicine residents: (1) posing well-structured, searchable questions arising from clinical cases; (2) performing efficient electronic literature searches to answer these questions; (3) understanding quantitative outcomes from published studies about diagnosis and treatment; and (4) evaluating the methodological quality of published studies and their clinical relevance to specific patients. We also carried out repeat assessments to see if these acquired skills were durable over time. These four skills are similar, but not identical, to those espoused by proponents of EBM—namely, question formulation, literature searching, critical appraisal skills, and application of the results to the patient.¹⁴

METHODS

Study Design

We performed a firm-based, controlled trial of an EBM educational course for first-year residents in the Cook County Hospital internal medicine training program. All residents and faculty in the program are assigned (without formal randomization) to one of three firms in the Department of Medicine.

There were 55 eligible first-year residents in our program at the start of the study. First-year residents in one firm ($n = 18$) were assigned to the experimental group;

Received from the Department of Medicine, Cook County Hospital and Rush Medical College, Chicago, Ill.

Presented in part at the annual meeting of the Society of General Internal Medicine, San Francisco, Calif, April 30, 1999.

Address correspondence and reprint requests to Dr. Smith: Department of Medicine, Cook County Hospital, 1900 W. Polk St., 9th Floor, Chicago, IL 60612 (e-mail: csmith8@rush.edu).

those in the remaining two firms ($n = 37$) were assigned to the control group. The experimental firm received a 7-week course taught twice weekly by three faculty members (BMR, ATE, and RAM), and six chief medical residents. During this time, the control firms participated in equally intense educational curricula of identical duration, usually case-based clinical presentations in small interactive groups emphasizing clinical practice, but without formal EBM teaching.

Evidence-based medicine skills were assessed before the intervention in January 1998 (precourse test); after the experimental firm completed the EBM course in March 1998 (postcourse test 1); after the control group received the EBM course in June 1998 (postcourse test 2); and a final time in November 1998 (postcourse test 3). The primary outcome was the difference in test scores at postcourse test 1. The study design is shown in Figure 1.

Educational Intervention

The EBM course consisted of 1-hour noon conferences twice a week for 7 consecutive weeks and a 1.5-hour session in a computer lab. The course curriculum highlighted diagnosis and treatment issues—not risk, prognosis, or other content areas. The first session of each week was a didactic seminar which focused on se-

lected components of the EBM curriculum. In the second session, these skills were practiced in an interactive format involving actual patient cases (Table 1).

In teaching question formulation, we focused on four components: (1) a defined patient population; (2) the intervention of interest (a test or a treatment); (3) clinically meaningful outcome measures; and (4) a comparison group.¹⁵ Literature searching was taught in 1 didactic and 1 computer lab session; the searching process was also reinforced weekly during follow-up discussions. Diagnosis seminars focused on understanding 2×2 tables and core criteria necessary to appraise studies about diagnostic tests rather than likelihood ratios and odds.¹⁶ Treatment seminars focused on the numerical measures of treatment benefits and harms (risk reductions and number needed to treat [NNT]),¹⁷ understanding confidence intervals (CIs), and core criteria to appraise the validity of treatment trials.¹⁶

Outcome Measures

The primary outcome measure was a written test of EBM skills. We also surveyed residents' self-reported attitudes and behaviors. The EBM skills test included 5 sets of questions; each set was prefaced by a clinical case scenario and designed to evaluate one of five content areas: (1) posing well-structured clinical questions; (2) performing computerized literature searches (using OVID MEDLINE—the only resource available at our institution) for both diagnostic and treatment questions; (3) understanding quantitative aspects of published diagnostic studies (2×2 tables, predictive value, pretest and posttest probabilities, etc.); (4) understanding quantitative aspects of published treatment studies (relative risk reduction, absolute risk reduction, number needed to treat, confidence intervals); and (5) assessing the quality of treatment studies and their clinical relevance to particular patients. For purposes of scoring the tests, we collapsed content areas 3 and 4 above into one skill domain (quantitative understanding), yielding four tested skill domains. The maximum total score (100%) was a weighted average of scores in the four domains: formulating questions (16%), literature searches (36%), quantitative understanding (24%), and appraisal of study quality and clinical relevance (24%).

The overall structure of the test, the weighting of each skill domain, and the specific content areas remained unchanged for the four tests given between January and November. However, the clinical cases were changed and the wording of test questions varied somewhat to minimize recall bias on test performance. Two raters graded each exam independently; disagreements were resolved by consensus or after consultation with a third or fourth coinvestigator; however, raters were not blinded to study group assignment.

In addition, we collected basic demographic data from the residents and a survey of self-reported EBM skills and behavior. This survey was repeated each time

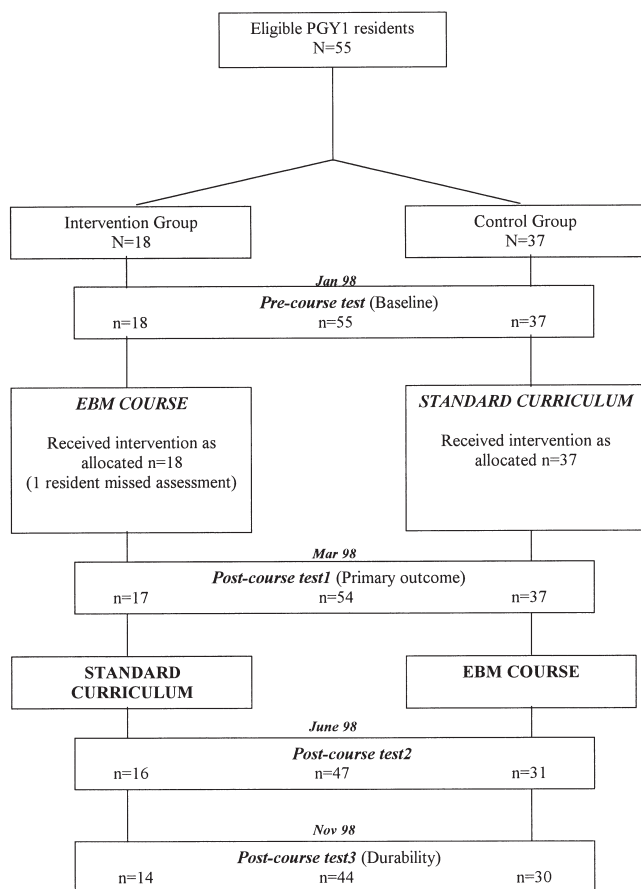


FIGURE 1. Study design and timetable.

Table 1. Evidence-Based Medicine Educational Curriculum and Course Design

Week	Session 1: Didactic Seminar	Session 2: Practice
1	Introduction and question formulation	Case presentations Asking clinical questions
2	Literature searching Computerized searching for diagnosis and treatment articles	Case presentation Searches assigned for diagnosis questions
3 & 4	Diagnosis Quantitative concepts: 2×2 tables, sensitivity, specificity, pretest and posttest probabilities Core validity criteria for diagnosis studies Review assigned searches (diagnosis)	Case presentations Searches assigned for diagnosis questions (Week 3) Searches assigned for treatment questions (Week 4)
5 & 6	Treatment Quantitative concepts: ARR, RRR, NTT, confidence intervals Core validity criteria for treatment studies Review assigned searches (treatment)	Case presentations Searches assigned for treatment questions (Week 5) Searches assigned for both diagnosis and treatment questions (Week 6)
7	Review Major topics and assigned searches	Postcourse test 1

ARR indicates absolute risk reduction; RRR, relative risk reduction; NTT, number needed to treat.

the written test was administered. The residents were not given any feedback on test performance or any answers until after the completion of the study to minimize the impact of testing four times in 11 months.

Statistical Analysis

Our primary analysis was a comparison of the mean total test scores and scores in the four domains between control and intervention groups after controlling for baseline differences. The study had greater than 90% power to detect a 15-percentage point difference in total test scores assuming a standard deviation of 15 and an α of 5%.

We tested the differences in postcourse test scores with an analysis of covariance while controlling for pre-course test performance. We also examined whether other baseline characteristics of study participants were possibly confounding or modifying the effect of the educational intervention.

The demographic data and residents' self-reported EBM attitudes, skills, and behavior were analyzed using Mann-Whitney *U* tests for ordinal data and χ^2 tests for dichotomous data. All analyses were by intention-to-treat using the statistical software SPSS 8.0 (Chicago, Ill).

RESULTS

Study Subjects

Baseline characteristics of the experimental and control groups are described in Table 2. Mean age was higher in the experimental group, but age was not associated with either precourse ($P = .7$) or postcourse test scores ($P = 1.0$). Among the 55 first-year residents enrolled in the study, all completed the precourse test, 54 (98%) com-

pleted postcourse test 1, 47 (85%) completed postcourse test 2, and 44 (80%) completed all four tests over the 10-month evaluation period. The overall course attendance for the initial intervention was 77%. Precourse total test scores demonstrated equivalent baseline EBM skills in the 2 groups: 40% and 42% for the experimental and control groups, respectively ($P = .6$).

Primary Outcomes

Following the educational intervention, the experimental group achieved significantly higher mean total postcourse test 1 scores (64%) than the control group (45%). After adjusting for baseline differences, the mean adjusted difference was 21 percentage points (95% CI, 13% to 28%; $P < .001$).

The experimental group's improved test scores reflected statistically significant improvements in three of the four EBM skill domains (Table 3).

Secondary Outcomes

After the intervention, we found improvements in several of the residents' self-reported EBM skills and behavior. The number of residents performing three or more literature searches per week was 10 (59%) of 17 in the experimental group and only 6 (16%) of 37 in the control group ($P = .003$). In addition, over half of the experimental group (9 residents, 53%) reported their ability to interpret studies as "good or very good" compared with 11 (30%) in the control group ($P = .04$). There was no difference between groups in their self-perceived ability to apply the results of studies to patients ($P = .4$). None of the remaining self-reported behaviors or attitudes toward EBM was different between groups.

Table 2. Subject Characteristics at Baseline

	Experimental Group n = 18	Control Group n = 37	P Value
Background characteristics			
Mean age, y (range)	31 (25–48)	29 (25–37)	.04
Male, % (n)	61 (11)	65 (24)	.8
Previous research experience, median, y (range)	1 (0–2)	1 (0–4)	1.0
Previous critical appraisal training, % yes (n)	67 (12)	68 (25)	.7
Own home computer, % yes (n)	39 (7)	43 (16)	.9
Self-reported behavior			
Hours reading medicine/wk, median (IQ range*)	6 (3–7)	5 (4–7)	.7
Journal articles read/wk, median (IQ range*)	2 (2–2)	2 (1–2)	.3
Perform ≥3 literature searches/wk, % (n)	17 (3)	14 (5)	1.0
Seek an original study article, % frequently or almost always (n)	22 (4)	19 (7)	.6
Self-reported attitudes and skills			
Importance of evaluating the medical literature, % very important or essential (n)	78 (14)	65 (24)	.7
Ability to interpret study results, % good or very good (n)	28 (5)	21 (8)	.7
Ability to apply study results, % good or very good (n)	39 (7)	38 (14)	.7
Precourse test			
Question formulation, mean % correct	52	46	.6
Literature searching, mean % correct	33	33	.9
Quantitative understanding, mean % correct	33	38	.5
Study quality and clinical relevance, mean % correct	49	55	.5
Total score, mean % correct	40	42	.6

*IQ range indicates interquartile range.

After the control group residents were crossed-over and received the educational intervention, their scores improved by 14 percentage points to a total of 58% (Fig. 2), similar to the score (62%) achieved by the experimental group at postcourse test 2 ($P = .4$). Thus, the EBM course produced similar improvements in test scores for both the control group (postcourse test 2 – postcourse test 1) and intervention group (postcourse test 1 – precourse test) ($P = .1$). When the four skill domains were examined separately for postcourse test 2, no differences between groups were found ($P = .2$ to $.9$).

The durability of the educational intervention effect was demonstrated by scores on postcourse test 3, which was administered 9 and 6 months, respectively, after the experimental and control groups had received the EBM course (Fig. 2). At that time, both groups sustained their

previously improved performance, and there was no difference between them ($P = .2$).

DISCUSSION

These results show that our educational intervention significantly enhanced residents' EBM skills. Compared with the control group, residents in the intervention group improved substantially their ability to pose well-structured clinical questions, to search the literature for studies pertinent to their questions, and to understand studies' quantitative results. These educational outcomes were durable over a 6 to 9 month period. We did not demonstrate an effect of the EBM course on residents' abilities to evaluate studies' quality or relevance to individual patients. Taken together, these objective findings were

Table 3. Primary Outcomes: Total and Domain Scores in Postcourse Test 1

	Experimental Group n = 17	Control Group n = 37	Adjusted Difference* (95% Confidence Interval)	P Value
	Mean % Correct			
Question formulation	82	56	26% (16 to 36)	<.001
Literature searching	74	43	31% (20 to 42)	<.001
Quantitative understanding	59	46	15% (5 to 25)	.004
Study quality and clinical relevance	44	38	7% (–9 to 23)	.4
Total score	64	45	21% (13 to 28)	<.001

*Adjusted for precourse test score.

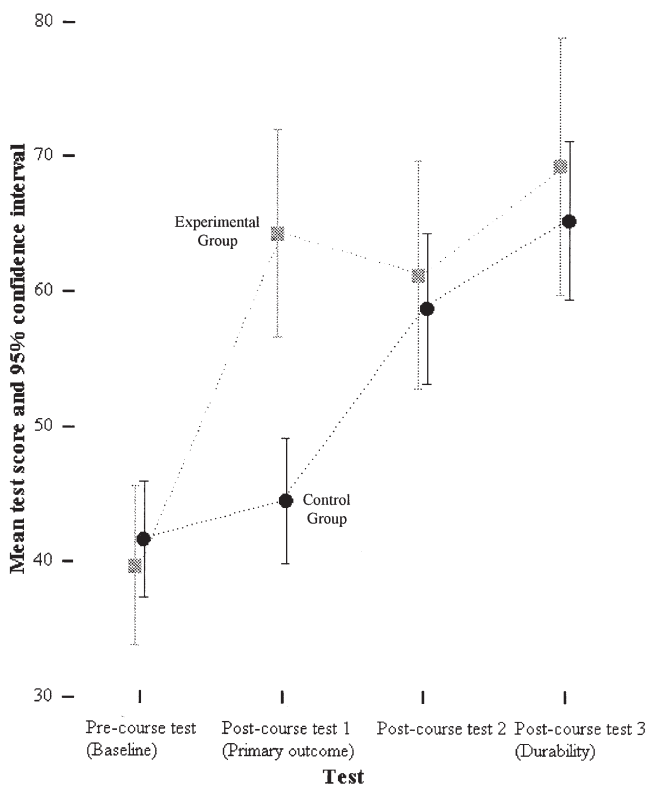


FIGURE 2. Test scores for the precourse test and the three post-course tests by experimental group. The gray squares represent the mean test scores in the experimental group for each test with 95% confidence intervals. The black circles represent the mean scores for the control group with 95% confidence intervals.

consistent with the residents' subjective self perceptions: after taking the EBM course, they believed they were better able to interpret study results but not better able to apply those results to patients' care.

We are encouraged by these findings. Many observers have noted the difficulties inherent in applying published clinical evidence to the care of individual patients.¹⁸⁻²¹ Thus, we were not surprised to find that a brief intervention for relatively inexperienced clinicians failed to improve measures of that outcome. We suspect that acquiring expertise in this "ultimate" EBM skill may require a career-long, conscientious effort. In contrast, as shown in this study, the more fundamental EBM skills can be taught successfully even to junior physicians.

There are several potential limitations to our study. First, residents were not formally randomized. However, the two groups had similar baseline characteristics, and adjustment for baseline differences, including performance on the precourse test, did not alter the results. Second, contamination between groups was possible, but this bias would minimize, not increase, differences between groups. Third, nondifferential measurement bias might have occurred, since our test to measure EBM skills had not been previously assessed. However, the impressive initial improvement by the intervention group and the consistency

of that improvement by the "crossed over" control group support the instrument's validity and reliability. Finally, the two raters grading the test were not blinded to group assignment, but each scored the tests independently using strict criteria for free-text answers, and most test questions were multiple choice, minimizing the opportunity for measurement bias. In fact, the greatest magnitude of improvement was seen for literature searching (multiple choice questions), and no improvement was seen in study quality assessment (free-text answers), supporting our belief that bias was unlikely to explain our findings.

Despite these limitations, we believe that we have shown that physicians in training can achieve proficiency in seeking, finding and understanding the "best evidence." Clearly, further efforts are required to produce greater improvements in critical appraisal skills²²⁻²⁴ with the goal being for learners to achieve all-round competency in EBM.²⁵ We think our results propel the EBM movement toward new more challenging terrain, such as assessing whether residents will employ these new skills in taking care of patients. The objective of teaching EBM should be to promote its adoption into everyday clinical practice. As technology and the availability of high quality evidence improve, our hope is that this can be accomplished during physicians' formative training years.

Our expectation is that research and innovation in medical education will lead the way. In our own educational program, where advocacy of the principles of EBM continues to grow,²⁶ we plan to study next the interface between "best evidence" and "best practice." Assessing whether the application of EBM skills by practicing physicians actually produces improvements in patient care, and evaluating why physicians choose to use, or not use, published evidence relevant to the care of their patients are important areas for future research.²⁷⁻²⁹ What can medical educators learn from those choices? These and other questions need answers before EBM can realize its full potential.

We thank Alice Furumoto-Dawson, PhD, for her help with data analysis.

The Department of Medicine of Cook County Hospital provided all funding and support for this project.

REFERENCES

- McAlister FA, Graham I, Karr GW, Laupacis A. Evidence-based medicine and the practicing clinician. *J Gen Intern Med.* 1999;14:236-42.
- Green ML, Ellis PJ. Impact of an evidence-based medicine curriculum based on adult learning theory. *J Gen Intern Med.* 1997;2:742-50.
- Williams BC, Stern DT. Exploratory study of residents' conceptual framework for critical appraisal of the literature. *Teaching and Learning in Medicine.* 1997;9:270-75.
- Stern DT, Linzer M, O'Sullivan PS, Weld L. Evaluating medical residents' literature appraisal skills. *Acad Med.* 1995;70:152-4.
- Kitchens JM, Pfeifer MP. Teaching residents to read the medical literature: a controlled trial of curriculum in critical appraisal/clinical epidemiology. *J Gen Intern Med.* 1989;4:384-7.

6. Linzer M, Brown JT, Frazier LM, DeLong ER, Siegel WC. Impact of a medical journal club on housestaff reading habits, knowledge and critical appraisal: a randomized controlled trial. *JAMA*. 1988;260:2537-41.
7. Seelig CB. Affecting residents' literature reading attitudes, behaviors, and knowledge through a journal club intervention. *J Gen Intern Med*. 1991;6:330-4.
8. Linzer M, DeLong ER, Hupart KH. A comparison of two formats for teaching critical reading skills in a medical journal club. *J Med Educ*. 1987;62:690-2.
9. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal and evidence-based medicine: a critical review of curricula. *Acad Med*. 1999;74:686-94.
10. Datta D, Adler J, Sullivan J, Leipzig RM. Test instruments for assessing evidence-based medicine competency. *J Gen Intern Med*. 1999;14(suppl 2):132. Abstract.
11. Sackett DL, Rosenberg WMC. On the need for evidence-based medicine. *J R Soc Med*. 1995;88:620-4.
12. Sackett DL, Rosenberg WMC. On the need for evidence-based medicine. *Health Econ*. 1995;4:249-54.
13. Sackett DL, Rosenberg WMC. On the need for evidence-based medicine. *J Public Health Med*. 1995;17:330-4.
14. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based Medicine: How to Practice and Teach EBM*. Edinburgh, UK: Churchill Livingstone; 1997:3.
15. Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123:A12-3. Editorial.
16. Haynes RB, Sackett DL, Muir-Grey JA, Cook DL, Guyatt GH. Transferring evidence from research in to practice: 2. Getting the evidence straight. *ACP J Club*. 1997;126:A14-6. Editorial.
17. Guyatt GH, Cook DJ, Jaeschke R. How should clinicians use the results of randomized trials? *ACP J Club*. 1995;122:A12-3. Editorial.
18. VanWeel C, Knottnerus JA. Evidence-based interventions and comprehensive treatment. *Lancet*. 1999;353:916-8.
19. Culpepper L, Gilbert TT. Evidence and ethics. *Lancet*. 1999;353:829-31.
20. Mant D. Can randomized trials inform clinical decisions about individual patients? *Lancet*. 1999;353:743-6.
21. Rosser WW. Application of evidence from randomized controlled trials to general practice. *Lancet*. 1999;353:661-4.
22. Linzer M. Critical appraisal: more work to be done. *J Gen Intern Med*. 1990;5:457-9.
23. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ*. 1998;158:177-81.
24. Sackett DL. Teaching critical appraisal: no quick fixes. *CMAJ*. 1998;158:203-4. Editorial.
25. Ghali WA, Lesky LG, Hershmann WY. The missing curriculum. *Acad Med*. 1998;737:734-6.
26. Reilly B, Lemon M. Evidence-based morning report: a popular new format in a large teaching hospital. *Am J Med*. 1997;103:419-26.
27. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient medicine is evidence based. *Lancet*. 1995;346:407-10.
28. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized controlled trials and recommendations of clinical experts. Treatments for myocardial infarctions. *JAMA*. 1992;268:240-8.
29. Weingarten S, Stone E, Hayward R, et al. The adoption of preventive care guidelines by primary care physicians: do actions match intentions? *J Gen Intern Med*. 1995;10:138-44.