

Educational Epidemiology

Applying Population-Based Design and Analytic Approaches to Study Medical Education

Patricia A. Carney, PhD

David W. Nierenberg, MD

Catherine F. Pipas, MD

W. Blair Brooks, MD

Therese A. Stukel, PhD

Adam M. Keller, MPH

CLINICIANS ARE COMMITTED TO providing the best patient care possible and, over time, the medical profession has increased its standards for assessment of clinical research in evidence-based medicine.¹ Observational epidemiologic studies often generate testable hypotheses or support hypotheses subsequently tested in randomized controlled trials (RCTs). An example is the number of epidemiologic studies on the risks and benefits associated with hormone therapies that have suggested that these therapies reduce risk of cardiovascular disease and osteoporosis. This extensive body of research provided the basis for the Women's Health Initiative, in which women were randomly assigned to receive hormone therapy or placebo.^{2,3} Results indicated that estrogen plus progestin reduced risk of osteoporosis but not of cardiovascular disease or invasive breast cancer,² and unopposed estrogen was found to increase risk of stroke.³ While the results of the Women's Health Initiative study proved some hypotheses (eg, use of hormone therapy reduced the risk of osteoporosis), other results contradicted both existing literature and treatment guidelines (eg, use of hor-

Conducting educational research in medical schools is challenging partly because interventional controlled research designs are difficult to apply. In addition, strict accreditation requirements and student/faculty concerns about educational inequality reduce the flexibility needed to plan and execute educational experiments. Consequently, there is a paucity of rigorous and generalizable educational research to provide an evidence-guided foundation to support educational effectiveness. "Educational epidemiology," ie, the application across the physician education continuum of observational designs (eg, cross-sectional, longitudinal, cohort, and case-control studies) and randomized experimental designs (eg, randomized controlled trials, randomized crossover designs), could revolutionize the conduct of research in medical education. Furthermore, the creation of a comprehensive national network of educational epidemiologists could enhance collaboration and the development of a strong educational research foundation.

JAMA. 2004;292:1044-1050

www.jama.com

mone therapy increased the risk of cardiovascular disease and breast cancer). This is an important illustration of an evolving body of evidence leading to an RCT, the findings of which were clinically relevant.

In contrast, medical education is less evidence-based, despite its increasingly precise national requirement. For example, what is the basis for the Liaison Committee for Medical Education (LCME) and Accreditation Council for Graduate Medical Education (ACGME) accreditation requirements? If medicine has a high threshold for evidence of clinical care, why is there no corresponding threshold for educational effectiveness? Most research in medical education is qualitative as opposed to quantitative and is composed of anecdotal reports, opinion pieces, and other descriptive reviews or position papers.⁴ Furthermore, difficulties exist that are associated with conducting high-quality educational research, such as lack of time and money for educational interventions and measurement.⁴ Most medical school faculty considering educational research are limited to their

dotal reports, opinion pieces, and other descriptive reviews or position papers.⁴ Furthermore, difficulties exist that are associated with conducting high-quality educational research, such as lack of time and money for educational interventions and measurement.⁴ Most medical school faculty considering educational research are limited to their

Author Affiliations: Departments of Community & Family Medicine (Drs Carney, Pipas, and Stukel), Medicine (Drs Nierenberg and Brooks), and Pharmacology & Toxicology (Dr Nierenberg), Dartmouth Medical School, Hanover and Lebanon, NH; Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, Ontario (Dr Stukel); Institute for Clinical Evaluative Sciences, Toronto (Dr Stukel); Office of the Vice President and Treasurer, Dartmouth College, Hanover, NH (Mr Keller).
Corresponding Author: Patricia A. Carney, PhD, Dartmouth Medical School, 1 Medical Center Dr HB 7925, Lebanon, NH 03756 (patricia.a.carney@dartmouth.edu).

own educational environments and learners, reducing generalizability and statistical power. Some of these challenges can be overcome with multi-institutional studies⁵; however, such studies are often restricted by the costs and complexities involved in collaborating.

Conducting educational research is especially challenging because widely accepted study designs are difficult to apply in curriculum-based programs. Strict accreditation requirements regarding both teaching content and methods reduce the flexibility needed to plan, execute, and evaluate educational experiments or innovations. In addition, an unwillingness of students and faculty to be involved in randomized studies because of perceived educational inequality can reduce participation. Despite these challenges, cost-effective opportunities to study medical education do exist, and validated knowledge- and clinical skills-based outcome measures are available to every medical school in the United States and Canada.

THE CONCEPT OF EDUCATIONAL EPIDEMIOLOGY

Educational epidemiology applies existing scientific methods to educational settings. Although most epidemiologic connotations refer to the study of risk factors that determine occurrence of disease or death in a population, these principles can be applied to the study of educational outcomes.⁶ For example, in medical education, students and residents constitute "populations" that make independent choices about events that potentially influence eventual competence as physicians. These choices include which medical schools to attend, clerkship sequence and site placements, electives, career paths, and ranking of residency programs. In addition, several short- and long-term outcomes can be identified. Short-term outcomes might include passing United States Medical Licensing Examination (USMLE) Steps 1-3, or attainment of first choice in residency

program. Longer-term outcomes, which would require collection of additional data and consideration of graduate medical education and continuing medical education events, might include patient satisfaction, quality of care, likelihood of being sued for medical malpractice, and experiences with medical errors.

The time to use epidemiologic approaches in educational research could not be better. The admissions office of each US medical school collects core characteristics, such as age, sex, grade point average, and Medical College Admission Test score, for each of its students. Although these variables could confound results of analyses related to the relationships between programmatic exposures and licensing outcomes, such as scores on USMLE Steps 1-3, they could be handled as covariates in any such analysis. Every US medical school also collects and classifies specific information about its educational programs, including programmatic and teaching methods, for LCME accreditation. In addition, several new requirements exist for US medical schools scheduled for accreditation in 2005⁷: (1) there must be comparable educational experiences and equivalent methods of evaluation across all alternative instructional sites, within a given discipline (LCME Educational Objective 8 [ED-8]); (2) medical school faculty must establish a system for the evaluation of student achievement throughout medical school that uses a variety of measures of knowledge, skills, behaviors, and attitudes (ED-26); (3) directors of all courses and clerkships must design and implement a system of formative and summative evaluation of student achievement in each course and clerkship (ED-30); and (4) curricula must clearly list competencies and how they are evaluated (ED-1A).

Medical schools have uniform validated data on learners, including results on USMLE Steps 1 and 2. Although these examinations primarily have assessed biomedical science and clinical knowledge, expansion to competency-based assessment of clinical

skills will begin in July 2004,⁸ with the USMLE Step 2—Clinical Skills Examination. All these variables would be readily available for use in statistical analyses. Although passing rates on these national examinations appear to be high, they have not, to our knowledge, ever been linked to longer-term competency-based outcomes. Creating these linkages through educational research can serve to strengthen the clinical care provided by the physician population.

The Association of American Medical Colleges' Medical Education Objectives Project⁹ and the ACGME Core Competencies¹⁰ further demonstrate that educational structures will align with standardized goals, allowing for assessments that are more robust than those previously possible. Identification and requirements of general ACGME competencies are the first step in a long-term effort to emphasize educational outcomes in the accreditation process. Required competencies include: medical knowledge, clinical skills for patient care, interpersonal communication, professionalism, systems-based practice, and practice-based learning and improvement. These competency requirements demand different educational outcome measures than previously have been applied in educational research.

What do these policy changes mean for US medical schools and residency programs? Specific mechanisms for data collection and use must be designed, tested, and implemented, not only for end-of-course, clerkship, or rotation assessment, but also to provide formative feedback to ensure that learners have an opportunity to improve. Creating effective educational mechanisms to address these requirements presents an opportunity for developing relational databases for multipurpose evaluation and research, such as those considered in the following examples.

EDUCATIONAL EPIDEMIOLOGIC RESEARCH DESIGNS AND EXAMPLES

The function of epidemiologic research designs is to conduct an unbi-

ased assessment of factors associated with an outcome in 2 or more groups. All study designs can both generate and test hypotheses, depending on the study question. For example, a surprise finding within a subgroup analysis in an RCT can generate new hypotheses. Alternatively, a longitudinal cohort study is one of the best designs for demonstrating cause and effect.

The value of hypothesis generation should not be underestimated. It requires rigorous thought about all possible explanations of findings, followed by more-discriminating study designs. In follow-up studies, relevant aims guided by appropriate conceptual frameworks must be delineated. The study design applied to test hypotheses must satisfy thresholds for scientific acceptance. Below we outline epidemiologic study designs and examples found in both clinical medicine and medical education. We use this comparative approach to encourage clinician teachers to think differently about educational research.

Observational Designs

Cross-Sectional Studies. Cross-sectional assessments are usually performed using survey methods at a single point in time. An example in clinical medicine is a study that assessed differences in the prevalence of panic attacks among adults in the US population between 1980 and 1995.¹¹ Survey responses from 1980 (n=20291), gathered using data from the Epidemiologic Catchment Area Program,¹² were compared with survey responses from 1995 (n=3032), gathered using the Midlife Development in the United States survey.¹³ A greater than 2-fold increase in the prevalence of panic attacks occurred between the 2 time periods (from 5.3% in 1980 to 12.7% in 1995), a finding with clinical relevance for psychiatrists and primary care clinicians who could expect to see an increase in this disorder in patient panels.

One cross-sectional educational study assessed computer connectivity and use for clinical and educational purposes among community-based pri-

mary care preceptors.¹⁴ Because the hypothesis being tested was that younger physician preceptors would have more computer connectivity and use than their older counterparts, analyses were stratified by age. The hypothesis was not supported by the results of this study because the oldest group (aged ≥ 60 years) of community-based physician preceptors used the Internet more often for both patient care decisions and trainee educational activities than did their younger counterparts. This example illustrates how cross-sectional research can identify inaccuracies in our assumptions about the use of educational resources.

Longitudinal Studies. Longitudinal studies use either ongoing surveillance or frequent cross-sectional methods to allow for assessments of change over time. An example in clinical research compared cancer trends in the United States and Europe¹⁵ using population-based sample sizes in the hundreds of thousands. In this analysis, mortality rates for breast cancer were currently lower in the United States compared with 20 years prior and also were currently lower in the United States compared with the current rate in Europe. These findings may be due to improvements made by the Mammography Quality Standards Act (enacted in 1996)¹⁶ or to adjuvant therapy. This type of research indicates the potential impact longitudinal studies can have when evaluating changes in health policy or initiation of new treatment modalities.

An example in medical education is the Medical Education Assessment Project¹⁷ being conducted in 10 US medical schools to assess how the attitudes and beliefs of medical students about medicine change throughout their 4-year program. Preliminary findings from 4 schools illustrate important differences in the development of attitudes toward medicine that cannot be explained by admissions criteria. A strength of longitudinal studies is that important causal associations can be identified that would not be identifiable without repeated measures.

Cohort Studies. Cohort studies involve assembling 1 or more groups based on exposure to environmental or behavioral factors or an intervention and then following the cohort over long periods of time. Generally speaking, there are 2 types of cohort studies: prospective and retrospective. In prospective studies, the investigators assemble the cohort, then collect baseline data in the present and outcome data in the future. In retrospective studies, investigators assemble the cohort and amass baseline data from the past, then collect outcome data from the past or present.

One example of a clinical cohort study is the Framingham Heart Study,¹⁸ which evaluated baseline serum potassium levels and the subsequent risk of cardiovascular disease in 3151 participants free of cardiovascular disease and not taking medications that would affect serum potassium levels. Potassium levels were measured between 1979 and 1983, within which time 313 cardiovascular disease events occurred, including 46 deaths. After adjusting for age and sex, no associations were found between baseline serum potassium levels and risk of cardiovascular disease. This type of study design has power in high numbers of participants with substantial amounts of data.

One example of a medical education prospective cohort study assessed the relationships among medical students' clinical experiences in their respective clerkships, their performance on final examinations, and their learning styles.¹⁹ Two cohorts of medical students (1478 students in 1 school and 2399 in another) entered the study at different time periods (1980 and 1986, respectively) and were assessed through 1987-1988 and 1991-1992, respectively. Results indicated that students' clinical experiences during clerkships were not related to success in their respective clerkships' final examinations. Rather, knowledge gained from clinical experiences was mediated by strategic and deep learning styles in both early and late phases of medical education. As measures of knowledge

gained, assessments of learning style may be more valuable than results of final examinations.

Case-Control Studies. Case-control studies are used for studying rare events or diseases. They involve participants with a condition (cases) and those without the condition but as similar to the cases as possible (controls). Examples in clinical medicine are studies that have assessed the relationship between use of diethylstilbesterol (DES) for treatment of corpus luteum insufficiency in early pregnancy^{20,21} and the influence of DES on development of vaginal cell adenocarcinoma (VCA). In 1 study, VCA was noted in only 0.1% of women who used DES.²⁰ However, at least 1 case-control study that matched the daughters of mothers who used DES during pregnancy (cases) to daughters of mothers who did not use DES during pregnancy (controls) found that the daughters of cases were more likely to experience several reproductive abnormalities, including VCA, compared with the daughters of controls.²¹ This work led to subsequent biological research²⁰ on mutation screening for polymorphisms in human progesterone receptor genes, which may prevent or identify VCA among offspring of cases in early stages, when treatment is most effective. This work provides an example of how epidemiologic research can translate back to the laboratory and then to the clinic.

A medical education case-control study might involve rare events, such as students' receiving failing scores on Objective Structured Clinical Examinations (OSCEs) or on licensing examinations. This would involve selecting a control group (students who passed the examinations) and a case group (those who did not). The cases might then be matched to controls based on a ratio of 1 case to 2 controls. Example hypotheses might include that examination scores would be lower among the cases because of a previously unidentified learning disability or that study habits were better developed in controls compared with cases. Understanding specific characteristics

of the 2 groups may assist in identifying students at high risk of receiving failing scores on critical examinations so they might be referred for early counseling or remediation. Such possible findings serve as a foundation for future research and illustrate how educational research findings might be used to improve medical education over time.

Experimental Designs

Trials involving random assignment of participants have long been considered the criterion standard for evaluating interventions and outcomes. Minimizing bias is a hallmark of random assignment to treatment and control groups, in part because observational study designs have overestimated treatment effects.²² Despite the strengths of trials involving random assignment, there are some caveats, which include difficulties implementing blinded study designs and possible cross-intervention contamination.

Randomized Controlled Trials. Randomized controlled trials involve enrolling a defined group of study participants and assigning them at random to an intervention group, a control group with no intervention, or a comparison group that might receive usual care. A particularly interesting historical example of an RCT in clinical medicine is the study by James Lind, *A Treatise of the Scurvy in Three Parts*,²³ which determined that citrus fruits cure scurvy. In this study, 12 sailors with the same symptoms were provided with the same diet, kept in the same location on their ship, and randomly assigned to 1 of 6 groups receiving either 1 quart of cider per day; 25 "guts" of elixir vitriol 3 times per day; 1 half-pint of sea water per day; 2 oranges and 1 lemon per day; a concoction of nutmeg, garlic, mustard seed, balsam of Peru, gum myrrh, barley water, and crème of tartar; or no specific treatment. By the end of 6 days, the 2 sailors who ate citrus fruits were the only sailors able to return to active duty.

We chose this study not only for its historical relevance as being among the

first controlled trials in clinical medicine but also because the first 2 parts of the treatise outline several observational studies that provided the foundation for the subsequent intervention trial. This example indicates how a series of studies in one specific area can result in an effective "treatment." This work was performed 270 years ago, yet we have not consistently applied these rigorous techniques to education.

One educational RCT example assessed the One-Minute Preceptor program and residents' teaching skills.²⁴ This study involved 57 second- and third-year internal medicine residents randomly assigned to receive either no intervention or a 1-hour session that incorporated lecture, group discussion, and role playing as an educational intervention. Students working with the residents rated those who received the intervention higher on teaching skills than residents not involved in the program.

Randomized Crossover Designs. Randomized crossover designs have great potential in educational studies. This design involves randomly assigning participants to either an exposure/intervention group or a control group for some period of time and then crossing the groups over so that the intervention group becomes the control group and vice versa. The first assignment in the design is most rigorous, since obvious contamination occurs in the "control group" after crossover. In clinical medicine, these designs are often applied in drug trials; for example, 1 study evaluated body surface area-based dosing vs fixed dosing of paclitaxel.²⁵ Paclitaxel disposition was significantly related to body surface area in this study, providing its rationale for body surface area-based dosing. More than a decade of work assessing this study design has found that it applies best if the exposure or intervention is intermittent and its effect on outcomes is immediate.²⁶

Although we could find no published report of an educational study using a crossover design, this type of design has great potential for use in edu-

educational settings. For example, students might be randomly assigned to receive course content either presented using an interactive Web-based educational system or presented didactically, with the presentation techniques switched later in the course. There is a sense of balance and fairness in a crossover design that is especially attractive when studying new learning strategies. Outcome measures might include scores on final examinations, while process measures might include time spent learning the material using each teaching approach. The implications of such a trial would expand across all stages of medical education.

Challenges of Experimental Designs

Despite successful implementation of the RCT design by Furney et al,²⁴ this type of study is especially difficult to apply in educational settings. For example, RCTs involve actively obtaining informed consent from participants, whereas other types of epidemiologic studies may allow for an institutional review board (IRB) exemption because they involve low risk or use completely anonymous data. Students and residents who choose not to be involved in RCTs may introduce a self-selection bias into the study and reduce sample size, affecting power and generalizability. Therefore, RCTs must be carefully undertaken and have a strong evidence base for the intervention to be tested. Examples of such studies are rare in medical education,⁴ though they do exist.²⁷

Several recent studies on problem-based learning provide additional examples of the challenges involved in RCTs.^{28,29} One controlled evaluation study of problem-based learning found an average effect size in medical education of less than 0.5,²⁹ not dissimilar to the modest effect sizes found in health care research. However, to reliably detect an effect size of 0.5 with 80% power using an α level of .05 would require 126 medical students (63 in each study group). However, this study was inconclusive, since a recent review of RCTs on problem-based learning in

medical education²⁹ found that no trial reached this minimal sample size. To address this problem, institutions would need to collaboratively conduct a rigorous assessment with adequate power. Other articles have debated the usefulness and utility of conducting RCTs in educational settings.^{28,30,31} An RCT should not be undertaken without the support of educational epidemiologic studies, such as those conducted by Lind prior to his randomized trial,²³ to justify the interventions to be tested.

Rigorous study designs are needed to determine educational effectiveness. Interestingly, the notion that RCTs are at the top of the research hierarchy is coming into question. For example, a dual meta-analysis, with one meta-analysis performed on case-control and cohort studies and another performed on RCTs,³² involved studies assessing the same intervention: the effectiveness of BCG vaccine in preventing active tuberculosis. This meta-analysis of 13 RCTs yielded a relative risk of 0.49 (95% confidence interval [CI], 0.34-0.70) among those vaccinated compared with an odds ratio of 0.50 (95% CI, 0.34-0.65) in 10 case-control/cohort studies. Thus, if well conducted, RCTs and observational studies can both be quite powerful and not overestimate the magnitude of treatment effects.³²

BUILDING AN EDUCATIONAL EPIDEMIOLOGY RESEARCH NETWORK

To conduct research as previously outlined would require 3 underlying elements: existing infrastructure, institutional motivation, and a national commitment.

Existing Infrastructure

It is important to consider the availability of epidemiologists and their potential role as educational researchers, the availability of traditional IRBs and changes needed to improve oversight of educational research, and how the educational culture would need to evolve. Virtually every medical school teaches classical epidemiology. Therefore, faculty who understand study de-

sign, database design and management, statistical analysis, and research costs should generally be available. These faculty can contribute as consultants or investigators in educational research.

Institutional review boards exist in most academic medical centers. They may have special challenges when evaluating educational research. This includes understanding the potential burden of numerous educational studies involving medical students, residents, and faculty. Also, educational exemptions have commonly been applied by traditional IRBs, even though research designs were often used in educational settings. Increased surveillance regarding educational research and IRB issues is highlighted by recent events at the Association of American Medical Colleges, which was scrutinized for its administration of the annual Graduation Questionnaire to fourth-year medical students.³³

Learner privacy is another important matter, especially for research requiring longitudinal collection of data, in which names and addresses are required for initial and follow-up mailings. Traditional IRBs have important expertise with this issue and should be consulted whenever this is a concern. However, because IRBs cannot determine if proposed educational research may be competing with required academic activities, we recommend developing a policy for low-risk evaluation studies informing students, residents, and faculty that routine evaluation activities may lead to published articles that include anonymous data reflecting their work. For more complex or higher-risk research designs, we recommend convening a committee of medical school and IRB faculty who can review potential research for merit, quality, and overlap. Using this approach can prevent students and faculty from being overwhelmed by research projects and ensure that approved recruitment, privacy, and consent procedures are in place.

Finally, the educational culture may need to change. Teaching faculty must

be open to self-examination, as occurs in the clinical research arena, allowing for a critical review of existing weaknesses and mechanisms for adoption of effective educational strategies. The level of collaboration among medical schools and residency programs would also need to increase to address concerns regarding sample size and generalizability issues.

Institutional Motivation

Schools of medicine need to prioritize and actively support the teaching mission, the development of epidemiologic approaches to study educational effectiveness, and the faculty leading these efforts. Increased pressure on clinical faculty to generate patient care revenue competes with and therefore demands efficient approaches to teaching in the clinical setting. Rigorous evaluation could reform the medical curriculum by introducing empirical evidence to determine whether a change would be beneficial.

The growing body of scientific information in all biomedical fields requires continuous curricular evolution if medical school is not to expand to a 5-year program. Rigorous application of epidemiologic analysis across educational institutions should inform the choices about what could be deleted from the formal medical curriculum without compromising graduating students' clinical competence. Faculty participating in educational research should enhance the clinician-teacher academic track by making important contributions to the educational literature. Faculty should be developing their own academic portfolios as well as raising the bar of academic achievement to parallel that of clinical research. In addition, feedback to faculty about their teaching beyond the level of global satisfaction is currently scarce. Faculty could improve their teaching with more objective data, which should enhance learner satisfaction as well.

National Commitment

A centralized organization not affiliated with any accreditation process or

governing body is needed to facilitate ongoing multi-institutional research nationally. Such an organization could serve as a central statistical coordinating center to which medical schools could electronically send encrypted institutional data that could then map into a common data structure. This approach could allow teaching faculty across the country to collaborate on projects leading to research publications and grant proposals, thereby promoting the rapid development of high-quality research, and to minimize costs for those institutions that lack aspects of existing infrastructure and that would want to contribute to educational research. This work could then inform medical educators and accrediting bodies on best practices and serve as an effective means of dissemination and application. Unfortunately, our current research foundation is not large enough or rigorous enough to meet the evidence-based standard that would deem current educational processes best practice.

Finally, more funding should be devoted to educational research in the health professions. The current pool of funding for educational research is very modest and its ongoing existence is threatened annually. Perhaps this has occurred in part because the national outcomes from such programs are either not assessed or not visible or are otherwise unknown. In addition, determining educational effectiveness must become core to the educational mission for any university and college. Too many elements are now in place for medical educators not to take the lead in this. Medical schools must commit the funds either through departmental subventions or through alterations in rates of indirect costs assessed for educational research, just as they do for biomedical research.

The yield of a national program of educational epidemiologic studies would be an ever-evolving research base in medical education that could provide an appropriate foundation for more-discriminating studies. With well-defined and validated outcome mea-

asures readily available, the timing of such a plan could not be better. Although the social, political, and economic factors that affect any science may become increasingly significant as a national research agenda arises in medical education, objectivity is a core element of scientific investigation that should be applied in medical education just as it is in biomedical research. The power of this objectivity and its application to how actual medical practice is and should be taught would promote both achievement and the recognition it is due. The most effective medical education would benefit everyone, including educators, learners, and especially patients, by minimizing costs, reducing medical errors and medical malpractice, and maximizing quality of care.

In conclusion, it is clear that over the past 2 decades, the teaching mission in many academic medical centers has been subsumed by the clinical enterprise due to the economic imperatives of health care. If this trend does not change, who will teach the physicians of the future? Many key components discussed in this article are in place, but national commitment to the importance of evidence-based medical education must be present for these changes to occur in a timely fashion. Educational epidemiology could generate a powerful research base to allow for studies designed to determine educational effectiveness.

Acknowledgment: We gratefully acknowledge M. Scottie Eliassen, MS, for her important editorial contributions to the final draft of this article. This work was supported by the mission of the Offices of Medical Education and Community-Based Education and Research, Dartmouth University.

REFERENCES

1. Boissel JP, Cucherat M, Amsallem E, et al. Getting evidence to prescribers and patients or how to make EBM a reality. *Stud Health Technol Inform.* 2003;95:554-559.
2. Writing Group for the Women's Health Initiative (WHI) Study. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA.* 2002;288:321-333.
3. Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA.* 2004;291:1701-1712.

4. Lurie SJ. Raising the passing grade for studies of medical education. *JAMA*. 2003;290:1210-1212.
5. Elnicki DM, Cykert S, Linger B, Ogden P, Hannis M. Effectiveness of a curriculum in telephone medicine. *Teach Learn Med*. 1998;10:223-227.
6. Jekel JF, Elmore JG, Katz DL. *Epidemiology, Biostatistics and Preventive Medicine*. Philadelphia, Pa: WB Saunders Co; 1996.
7. Liaison Committee on Medical Education Web site. Available at: <http://www.lcme.org/>. Accessed April 29, 2004.
8. United States Medical Licensing Examination Web Site. Clinical Skills Examination. Available at: <http://www.usmle.org/news/cse>. Accessed November 11, 2003.
9. Association of American Medical Colleges. Medical Schools Objectives Project. Available at: <http://www.aamc.org/meded/msop>. Accessed December 10, 2003.
10. Accreditation Council for Graduate Medical Education Outcome Project. General Competencies: minimum program requirements language. Available at: <http://www.acgme.org/outcome/comp/compMin.asp>. Accessed December 10, 2003.
11. Goodwin R. The prevalence of panic attacks in the United States: 1980-1995. *J Clin Epidemiol*. 2003;56:914-916.
12. Eaton WW, Holzer CE, Von Korff M, et al. The design of the Epidemiologic Catchment Area surveys: the control and measurement of error. *Arch Gen Psychiatry*. 1984;41:942-948.
13. Wanf PS, Berglund P, Kessler RC. Recent care of common mental disorders in the United States: prevalence and conformance with evidence-based recommendations. *J Gen Intern Med*. 2000;15:284-292.
14. Carney PA, Poor DA, Schifferdecker KE, Gephart DS, Brooks WB, Nierenberg DW. Computer connectivity and use among community-based primary care physician preceptors: does age matter? *Acad Med*. 2004;79:580-590.
15. Quinn MJ. Cancer trends in the United States—a view from Europe. *J Natl Cancer Inst*. 2003;95:1258-1261.
16. Houn F, Elliot M, McCrohan JL. The Mammography Quality Standards Act of 1992: history and philosophy. *Radiol Clin North Am*. 1995;33:1059-1065.
17. Reed VA, Jernstedt GC, Boudreau D, et al. Meeting the challenge of MSOP: comprehensive measurement of profiles of student characteristics across medical schools. *Acad Med*. 1999;74(suppl 10):S96-S98.
18. Walsh CR, Larson MG, Leip EP, Vasan RS, Levy D. Serum potassium and risk of cardiovascular disease: the Framingham Heart Study. *Arch Intern Med*. 2002;162:1007-1012.
19. McManus IC, Richards P, Winder BC, Sproston KA. Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ*. 1998;316:345-350.
20. Engehausen DG, Schrott KM. PROGINS polymorphism of progesterone receptor is increased in female offspring with maternal exposure to diethylstilbestrol. *Anticancer Res*. 2000;20:5145-5149.
21. Swan SH. Intrauterine exposure to diethylstilbestrol: long-term effects in humans. *APMIS*. 2000;108:793-804.
22. Preventive Services Task Force. *Guide to Clinical Preventive Services: Report of the US Preventive Services Task Force*. 2nd ed. Baltimore, Md: Williams & Wilkins; 1996.
23. Lind J. A treatise of the scurvy. In: Stewart CP, Guthrie D, eds. *Lind's Treatise on Scurvy: A Bicentenary Volume*. Edinburgh, Scotland: The University Press; 1953.
24. Furney SL, Orsini AN, Orsetti KE, Stern DT, Gruppen LD, Irby DM. Teaching the One-Minute Preceptor: a randomized controlled trial. *J Gen Intern Med*. 2001;16:620-624.
25. Smorenburg CH, Sparrenboom A, Bontenbal M, Stoter G, Nooter K, Verweij J. Randomized crossover evaluation of body-surface area-based dosing versus flat-fixed dosing of paclitaxel. *J Clin Oncol*. 2003;21:2804-2805.
26. Maclure M, Mittleman MA. Should we use a case-crossover design? *Annu Rev Public Health*. 2000;21:193-221.
27. Linzer M, Brown JT, Fraizer LM, DeLong ER, Siegel WC. Impact of a medical journal club on house-staff reading habits, knowledge, and critical appraisal skills: a randomized control trial. *JAMA*. 1988;260:2537-2541.
28. Smits PBA, Verbeek JHAM, deBuissonje CD. Problem-based learning in continuing medical education: a review of controlled evaluation studies. *BMJ*. 2002;324:153-156.
29. Albanese M. Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. *Med Educ*. 2000;34:729-738.
30. Torgerson CJ. Educational research and randomized trials. *Med Educ*. 2002;36:1002-1003.
31. Torgerson CJ. Researching outcomes of educational interventions: initial formation of unbiased groups is most important issue in randomised controlled trials. *BMJ*. 2002;324:1156.
32. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-1892.
33. Cohen JJ. Compliance of the Graduation Questionnaire With Federal Human Subjects Requirements. Washington, DC: Association of American Medical Colleges; February 9, 2004. Memorandum No. 04-07.

What we have to do is to be forever curiously testing
new opinions and courting new impressions.
—Walter Pater (1839-1894)