SPECIAL ARTICLE

Problems in the ''Evidence'' of ''Evidence-based Medicine''

Alvan R. Feinstein, MD, Ralph I. Horwitz, MD, *New Haven, Connecticut*

The proposed practice of "evidence-based medicine," which calls for careful clinical judgment in evaluating the "best available evidence," should be differentiated from the special collection of data regarded as suitable evidence. Although the proposed practice does not seem new, the new collection of "best available" information has major constraints for the care of individual patients.

Derived almost exclusively from randomized trials and meta-analyses, the data do not include many types of treatments or patients seen in clinical practice; and the results show comparative efficacy of treatment for an "average" randomized patient, not for pertinent subgroups formed by such cogent clinical features as severity of symptoms, illness, co-morbidity, and other clinical nuances. The intention-to-treat analyses do not reflect important post-randomization events leading to altered treatment; and the results seldom provide suitable background data when therapy is given prophylactically rather than remedially, or when therapeutic advantages are equivocal. Randomized trial information is also seldom available for issues in etiology, diagnosis, and prognosis, and for clinical decisions that depend on pathophysiologic changes, psychosocial factors and support, personal preferences of patients, and strategies for giving comfort and reassurance.

The laudable goal of making clinical decisions based on evidence can be impaired by the restricted quality and scope of what is collected as "best available evidence." The authoritative aura given to the collection, however, may lead to major abuses that produce inappropriate guidelines or doctrinaire dogmas for clinical practice. ***Am J Med.*** **1997;103:529–535.**

From the Departments of Medicine and Epidemiology, Yale University School of Medicine, New Haven, Connecticut.

Requests for reprints should be addressed to Alvan R. Feinstein, MD, Sterling Professor of Medicine & Epidemiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510.

Manuscript submitted December 27, 1996 and accepted May 16, 1997.

Within 5 years of the first proposal,[1] "evidence-based medicine" (EBM) has received enthusiastic endorsement from editors of prominent medical journals,[2] achieved the publicational outlet of its own new journal,[3] and acquired the kind of sanctity often accorded to motherhood, home, and the flag.

Hardly anyone can disagree with the goal of getting clinicians to make "conscientious, explicit, and judicious use of current best evidence"[4] for decisions in patient care; and any expressions of doubt about the EBM activities are usually greeted with a vigorous counter attack. The critics are accused of not having "done their homework," disregarding "today's harsh realities," or ignoring "what happens in clinical medicine."[5] They are also denounced for erroneous beliefs that EBM comes only from randomized clinical trials (RCTs), that it contains "merely the mindless application of the results of megatrials," and that "other forms of evidence are heavily discounted."[6]

Because we share the general goals and have friendly admiration for proponents of EBM, but worry about its current methods, the comments here are drawn from what is actually done rather than from what is proposed. The comments refer to the novelty of EBM and particularly to the information contained as "evidence." The comments will also consider the sources of "authority" and potential abuses of the EBM movement.

NOVELTY

A major source of confusion has been the distinction between the contents of EBM itself, and the application of EBM in clinical practice. The contents are analogous to a textbook of medicine; the practice is what happens when clinicians make and carry out decisions in patient care.

With this distinction blurred, many clinicians have claimed that EBM is "nothing new." After all, most good clinicians have regularly assembled evidence when they reviewed their own experience, developed clinical judgment, read medical literature, attended medical meetings, and had discussions with one another. This activity seems entirely compatible with the statement that the practice of EBM consists of "integrating individual clinical expertise with the best available external clinical evidence from systematic research."[4]

The total information used in the practice of EBM also seems similar to what has been used by good

practitioners. For practicing EBM, the information is "not restricted to randomized trials and meta-analyses." It contains "clinically relevant research, often from the basic sciences of medicine," and it includes studies of diagnostic tests, of prognostic markers, and of "the efficacy and safety of therapeutic, rehabilitative, and preventive regimens."[4] With this description of what is done when EBM is practiced, and with the overt acknowledgment that EBM's "philosophical origins extend back to mid-19th century Paris and earlier,"[4] clinicians can easily conclude that EBM is not particularly novel, and may wonder why it has stirred so much fuss and controversy.

The novelty arises, however, not in the proposed practice of EBM, but in the contents of what is assembled for use as essentially a new kind of textbook in clinical medicine. A hint of this novelty is offered in the statement that EBM contains "the best available external clinical evidence from systematic research."[4] The restricted focus implied by this hint becomes apparent when its details are further elaborated. For "questions about therapy," according to the EBM proponents, "we should try to avoid the nonexperimental approaches, since these routinely lead to false-positive conclusions about efficacy . . . The randomized trial, and especially the systematic review of several randomized trials . . . has become the 'gold standard'. . . ."[4]

An almost exclusive concentration on the "gold standard" of randomized trials and meta-analyses (sometimes called "systematic reviews" or "overviews") is evident in the material that has been published in the new *Evidence-Based Medicine* journal, and in its predecessor and now concomitant periodical called *ACP Journal Club.* This restricted focus is also apparent in the activities of the Cochrane Collaboration,[7–10] which was developed from work that produced a famous collection of randomized trials and meta-analyses in obstetrics and gynecology.[11] The "collaboration," based at Oxford, now consists of an international consortium of workers who construct an ever-enlarging data base by contributing their own randomized trials, discoveries of unpublished trials, and meta-analyses. The collected information extends through all branches of medicine, and becomes the main source of "evidence" in EBM's new "textbook," which can appear in conventional literary formats as well as in electronic media, such as the Internet.

To allay concerns that the data base is inadequate, or that EBM does not suitably represent clinical practice, the proponents of EBM recently reported that "inpatient general medicine is evidence based."[12] In a study of 109 patients hospitalized on a university medical service, 53% of the primary treatments were found to be supported by data from RCTs, and in 29% of treatments, the clinical team had "unanimity . . . about the existence of convincing nonexperimental evidence." In an analogous subsequent study[13] at a "suburban training general practice," 30% of 101 therapeutic decisions were supported by RCTs and 50% by "convincing nonexperimental evidence."

Although supporting the contention that contemporary clinical practice is indeed "evidence based," these two studies also demonstrate that most of the therapeutic decisions, even in the special academic settings, were not derived from the data of RCTs. Some other forms of appraised evidence must have been used for the therapeutic decisions that were "unanimous" or "convincing," as well as for those that were not. Some other forms of evidence must also have been used for the many additional clinical decisions that involve etiology, pathophysiology, diagnosis, prognosis, mode of communication, and methods of reassurance, rather than a choice of specific therapeutic interventions. According to the proponents of EBM, "if no randomized trial has been carried out . . . , we must [find] the next best evidence and work from there."[4]

The rest of this essay is concerned with both the "evidence" and "the next best evidence" used for all those decisions, and with the way that the decisions may be improved or impaired by the impact of the EBM "textbook."

CONTENTS OF "EVIDENCE"

Despite the broad range of information permitted when EBM is practiced, the evidence collected for EBM itself is confined almost exclusively to randomized trials and the meta-analyses done with those trials. Because meta-analyses can aggregate and evaluate but cannot change the basic information, the RCTs themselves become the fundamental source to be considered both for quality and scope of data, and for the scope of topics contained the in EBM collection.

Quality and Scope of Data

To obtain "trustworthy" information, randomized trials have concentrated on getting "hard" data about death, disease, and demography. The patient's baseline condition, before therapy, is regularly characterized with the "reliable" information of age, gender, race, imaging, endoscopy, biopsy, cytology, and laboratory tests. The therapeutic outcome is cited, whenever pertinent, as death, or in global ratings for certain symptoms (pain, insomnia, etc.) that need not be specified more precisely because the "double-blind" observations will presumably avoid bias.

The information seldom includes "soft data," distinctive to individual patients within the spectrum of a particular diagnosed disease, that identify cogent patterns of illness and pathophysiology. These patterns demarcate the pertinent prognostic and therapeutic subgroups when clinicians make individual decisions in patient care. Important but usually omitted "soft" data are the types of symptoms, severity of symptoms, auxometry (rate of growth) of illness, and severity of the co-morbidity produced by concomitant associated diseases.[14]

A good clinician constantly uses this "soft" information for diverse clinical decisions. In diagnosis, for example, the patient's history can distinguish such phenomena as transient cerebral ischemia from orthostatic dizziness, and the physical findings can often help separate pneumonia from congestive heart failure. After a diagnosis has been made, the clinician considers pertinent clinical subgroups—not just an undifferentiated collection of people with the same disease—when evaluating a patient's "condition" to estimate prognosis and choose treatment.

Beyond omitting the important symptoms and other clinical variables that identify these subgroups, randomized-trial information also often omits clinical details that may be crucial for many other therapeutic decisions. Among those details are responses to previous therapeutic agents, short-term (24-hour) response to remedial therapy, ease of regulation when the dose of therapy must be "titrated," difficulty in compliance with therapy and reasons for noncompliance, psychic or nonclinical reasons for impaired functional status, the "social support" system available at home and elsewhere, the patient's expectations and desires for therapeutic accomplishment, and the patient's psychologic state and preferences. This list of important omitted variables is incomplete; and knowledgeable clinicians could readily add many others that have sometimes been critical for clinical decisions in the care of patients.

A separate problem arises from the information that *is* contained in RCTs. It is often entered directly into meta-analyses, without further evaluation of quality. For example, the morphologic data routinely provided by pathologists and radiologists is readily accepted, despite the major inconsistencies and biases revealed whenever the providers are exposed to tests of observer variability.[15] The information reported in the trials is regularly derived from the intention-to-treat policy that classifies results according to the randomly assigned regimens, not the regimens that were actually received. This policy ignores all the events that happen after randomization, the additional therapeutic decisions and interventions evoked by those events, and the subsequent consequences. For example, a patient assigned to medical therapy but who received unauthorized surgery may be analyzed as though the surgery did not occur.

The design and conduct of the trial itself may also sometimes be neglected. Proponents of meta-analysis have described their dismay[16–18] at noting the many RCTs that were scientifically unsatisfactory despite a "gold standard" status. Although rigorous criteria might be applied to chose only those RCTs that are truly "golden,"[19–21] the criteria will also be variably applied when hundreds of Cochrane collaborators do the meta-analyses.

The last problem to be cited in scope of data arises because many RCTs enroll a restricted population that is confined to patients expected to be highly responsive to treatment. For example, the effectiveness of lipid lowering may be checked in patients having the highest 10% of cholesterol values in the general population; and the effectiveness of a multifactorial intervention to prevent falls in a geriatric population may be examined only in patients who are neither too frail nor too healthy. The trials may be conducted elegantly and may yield indisputable results, but clinicians subsequently choosing treatment have a paradoxical problem, produced by the limitations of available evidence, in deciding how to apply the results for pertinent subgroups of patients who were excluded from the trials.

Scope of Topics

The scope of randomized trials is limited by direct applicability only to "average" patients, by the absence of RCTs for prophylactic therapy of "risk factors," by "grey zones" that have not been clarified by RCTs, by pathophysiologic principles for which RCTs would be inappropriate or unethical, and by the many clinical decisions for which RCTs are not possible or pertinent.

"Average" patients. One immediate problem in using results of RCTs is that the conclusions refer to an average patient who fulfilled the criteria for admission. When transferred to clinical medicine from an origin in agricultural research, randomized trials were not intended to answer questions about the treatment of individual patients. The trials have almost always been used to offer an average value for efficacy in groups of patients receiving the compared therapies. These average results for "efficacy" have been highly successful in letting policy makers and pharmaceutical manufacturers make such decisions as whether streptomycin is beneficial enough to warrant industrial production, whether bypass surgery is worth its risks and costs in patients with coronary disease, whether tissue plasminogen activator (tPA) is generally better than streptokinase for dissolving the clots of acute coronary occlusions, and whether

chemotherapy regimen A produces more cancer remissions and longer survivals than chemotherapy regimen B.

When clinicians make decisions for individual patients, however, the information needed for pertinent clinical subgroups may not be either reported or available. For example, tPA and streptokinase can be expected to have different rates of both beneficial and detrimental effects when acute myocardial infarction is treated in a relatively old man who also has congestive heart failure, or in a relatively young man without failure. Clinicians (and patients) would want to know risk/benefit appraisals for each treatment in these subgroups, not just for an "average" acute myocardial infarction.

Prophylactic therapy for "risk factors". A different therapeutic problem occurs when the treatment is prophylactic, aimed at preventing development of a disease or complication, rather than remedial, intended to relieve pain or distress. The goal of prophylactic therapy may be to reduce a "risk factor" that is either an external exposure, such as smoking or eating a high-fat diet, or an internal abnormality, such as an elevation in blood lipids, blood pressure, or blood sugar. A different kind of "risk factor" is an unsuspected derangement or "early" disease found as an asymptomatic internal abnormality on a screening examination. These abnormalities include cervical dysplasia, a "silent" cancer, a colonic polyp, or an abnormal stress-test electrocardiogram. In all of these circumstances, the patient basically feels well (at least with respect to the therapeutic "target"); and the therapy will not improve symptoms. The prophylactic treatment has the potential for long-term benefits, but may often produce short-term psychic difficulties due to the "labeling" phenomenon, or to adverse symptomatic reactions.

Randomized trials to demonstrate the value of many of these prophylactic treatments are either not yet completed (as in many instances of asymptomatic diseases found by screening) or have yielded inconclusive results. Examples are the controversies about the multiple risk factor intervention trial (MRFIT) and mammographic screening for women at ages 40 to 49. In some prophylactic trials, the clinical side effects may have been measured inadequately. For example, when one antihypertensive agent produced an allegedly better "quality of life" than another, the investigators did an intention-to-treat analysis that did not evaluate compliance with therapy.[22] With many antihypertensive agents, however, "quality of life" can be markedly improved if patients simply stop taking the drug.

In the absence of suitable RCTs, the evaluation of many of the prophylactic treatments becomes converted to an evaluation of the merits of screening. The latter evaluation, which cannot be done with meta-analytic or other types of aggregates used for EBM, requires thoughtful clinical appraisals and the ability to distinguish evidence from advocacy. In one recent evaluation, 76 often-recommended preventive practices were classified as having inconclusive supporting data for which "decision-making must be guided by factors other than medical scientific evidence."[23]

"Grey" zones of practice. While acknowledging that the RCT is the "cornerstone method . . . abetted . . . by the widespread application of meta-analysis," Naylor[24] has pointed out the "grey zones of clinical practice where the evidence about risk-benefit ratios of competing clinical options is incomplete or contradictory." Among the procedures he cites in the "grey zone" are carotid endarterectomy, upper gastrointestinal (GI) endoscopy, hysterectomy, and percutaneous transluminal coronary angioplasty. RCTs may have been done, but the results have not produced unequivocal conclusions.

Pathophysiologic principles. Another set of decisions involves combinations of pathophysiologic and ad hoc therapeutic reasoning. Despite adequate clinical trials for the average efficacy of individual diuretics, nonsteroidal anti-inflammatory drugs, or anxiolytic agents, the patient's clinical state and previous response will usually be the impetus for adjusting individual dosage, or for discontinuing one of the "efficacious" agents and starting another. Decisions to start or stop remedial therapy with oxygen, mechanical ventilation, blood transfusion, or for patients with electrolyte alterations will almost always depend on individual pathophysiologic status, not on published evidence.

Finally, all of the previously discussed issues in personal preferences, psychosocial factors, comfort, and reassurance must be individualized, and cannot be suitably guided by published reports for an "average" patient. For these and all of the other decisions in which clinical-trial or other published data are nonexistent, too constrained, or otherwise limited, clinicians must rely on judgment, using information not available in the statistical evidence.

In the two previously cited studies,[12,13] pathophysiologic principles complemented other data to justify the treatments supported by "convincing nonexperimental evidence." On the inpatient medical service,[12] such treatments included antibiotics for infection, implanted pacemakers for symptomatic heart block, transfusion for hemorrhagic blood loss, fluids for dehydration, and catheterization for urinary obstruction. For ambulatory patients,[13] the treatments included counselling for depression or anxiety, narcotics or analgesics for pain, syringing for ear wax, incision and drainage for abscess, cryotherapy for

skin tag, and hormone replacement for menopause, as well as apparently appropriate antimicrobial agents for various infections.

The investigators in the two studies were "convinced" that all of these therapeutic decisions were "evidence-based" despite the absence of supporting data from randomized trials. In most of the cited circumstances, however, appropriate trials would be difficult to design or unethical to carry out. What would be a suitable double-blind placebo (and who would volunteer to receive it) in randomized trials comparing therapeutic efficacy for transfusions, pacemakers, bladder catherization, incising and draining an abscess, syringing ear wax, or removing a skin tag?

RCTs not pertinent or possible. In many other clinical decisions, randomized trials are either not pertinent or not possible. Intended to compare beneficial effects, the trials will usually be deemed unethical if aimed at instigating and investigating agents such as smoking or alcohol that are accused of having noxious effects in causing disease. Even if deemed ethical by an institutional review board, the trials would probably be unable to recruit enough volunteer subjects. An analogous problem would prevent the use of RCTs to evaluate open prostatectomy versus transurethral resection for benign prostatic hypertrophy. Most urologists (and patients) would refuse an open prostatectomy for conditions that seem equally well managed by the simpler transurethral procedure. Consequently, most studies of etiologic agents and many comparisons of treatment, particularly in surgery, will depend on clinical, pathophysiologic, and other data acquired in observational circumstances, without experimental designs.

Randomized trials have not been (and probably will not be) done to evaluate most diagnostic marker agents. Various scientific principles can be used to improve the methodologic quality of the observational studies, but the methods will seldom rely on randomized trials. Consequently, the literature on diagnostic marker tests cannot be easily subjected to the criteria used for meta-analysis of RCTs, and besides, in many instances the literature hardly warrants the dignity of numerical aggregation. In a recent review,[25] more than half of pertinent publications did not fulfill basic methodologic standards for scientific quality.

Finally, all of the previously discussed issues in personal preferences, psychosocial factors, comfort, and reassurance for patients are essential elements of clinical decisions in humanistic care. These decisions must be individualized, and the approaches could not be guided, even if RCTs were available, by published reports of results in an "average" patient. The investigators who did the previous cited study[13] of "evidence based general practice" concluded that the methods used in RCTs "may not be appropriate to apply to this setting." They said that evidence from RCTs, rather than being a gold-standard, may have "more the value of a coffee future—likely to be altered by tomorrow's experience."[13]

THE EBM "TEXTBOOK"

For all these reasons, the randomized trials and meta-analyses contained in the EBM "textbook" may offer a splendid collection of "the best available external clinical evidence from systematic research." The collection, however, will have inevitable scientific deficiencies in quality and scope of the raw data that have been aggregated; and RCT results will be available for only a limited scope of therapeutic topics. The EBM information can certainly enlighten the therapeutic decisions for which the information is both pertinent and trustworthy, but a clinical practitioner will have to rely on other sources for many therapeutic choices (even when RCT data are available) and for almost all the nontherapeutic decisions of patient care.

The current emphasis on the EBM "textbook" also has two other major disadvantages: it offers no guide or instruction for the pathophysiologic and other judgmental reasoning used in clinical decisions; and clinical investigators occupied with the search-and-aggregate missions of meta-analysis are diverted from doing other research that might offer better "evidence" for all the non-RCT clinical decisions. Such research would include efforts to improve the scientific quality of clinical examination and data, and thoughtful analysis of the many phenomena that must be studied with nonrandomized methods.[26]

The emphasis on constantly enlarging the EBM "textbook" also runs the risk of re-creating the intense enthusiasm followed by sad disenchantment that occurred about 25 years ago when the "Problem-Oriented Medical Record" was introduced[27] as a new method of structuring information, and promptly hailed as almost a panacea for the challenges of patient care. Like advocates of the problem-oriented record, who often spent more time taking care of the record than taking care of the patient, advocates of EBM may often be diverted from the bedside to the library or computer terminal. The problem-oriented record offered some useful advances that still endure in the structure of medical records, but the original anticipations were ultimately unsuccessful because the approach did not cope with fundamental inadequacies in the data, and did not construct a serviceable taxonomy for classifying the "problems."[28] The current approaches of

EBM seem to have analogous appeals and limitations.

Sources of Authority

A separate set of questions arises about the choice and delegation of "authority" in the authoritative EBM processes. The products of any field of activity must stand on their own merits, regardless of who commissioned or supported the work, or who did it. Nevertheless, because so many judgmental decisions are involved in the EBM processes, assurances can be sought about mechanisms of operation and accountability. All medical publications are ultimately accountable to a publisher and often to a sponsoring medical society.

How does the EBM operation work? Who decides what to accept or reject? Who determines what will be disseminated as things to believe or not believe? Who chooses the decision makers? What are their qualifications and credentials? Who determines which topics are chosen to receive meta-analytic appraisals and authoritative decisions?

As of 1995, the Cochrane Collaboration contained 9 geographic centers,[10] and about 50 collaborative review groups (devoted to topics such as stroke or incontinence), 15 field coordinators (for such work as primary health care and care of the elderly), and "more than 1000 people from over 50 countries (who) are either contributing . . . or said they want to contribute."[29] The Collaboration is also remarkably egalitarian. Iain Chalmers, often regarded as its founder, has said that "anyone with helpful ideas . . . is welcome to take part."[30] According to Brian Haynes, co-editor of the *Evidence-Based Medicine* journal, the main requirement for becoming one of the Cochrane authorities is willingness to accept "a lifetime 'sentence' to complete and periodically update a systematic review as a member of a registered review group."[31] David Sackett, probably the most prominent leader of the EBM movement, joined several colleagues[5] in offering, to someone "whose ignorance of what happens in clinical medicine remains intact," a "standing invitation to join our clinical team."

The evaluation of quality in RCTs is not an easy task. Different experts have proposed different criteria[16–18] for the evaluations; and although many of the criteria require acts of interpretive judgment, reproducibility of the judgments has seldom been tested. Consequently, interpretive decisions by old pre-EBM experts may be replaced by interpretive decisions from a new group of experts with EBM "credentials." As the EBM pronouncements appear, the new authorities who produce them may not always be prominently identified, however. Their names may be listed in the small "telephone-directory" type often used for the multiple collaborators participating in large-scale cooperative randomized trials.

The evidence used for the pronouncements will be documented, explicit, and numerate; and a collection of criteria will be stated and published for the mechanisms used in analyzing the evidence; but the use of those criteria will involve subtle judgments and decisions from the large panels of experts whose identities and credentials may be difficult to discern. Concern about the quality of this process has even been expressed by Richard Peto, a pioneer in the meta-analysis movement. He is quoted as saying[32] that the painstaking detail of a good meta-analysis "just isn't possible in the Cochrane collaboration" when the procedures are done "on an industrial scale."

Potential Abuses

Although not responsible for the ways in which EBM might be abused, the proponents offer assurance that the "bottom up approach" (integrating external evidence with clinical expertise and patients' choice) "cannot result in slavish, cookbook approaches to individual patient care," and that even if "hijacked by purchasers and managers to cut the costs of health care," EBM's identification of the "most efficacious interventions . . . may raise rather than lower the cost."[4]

Nevertheless, the products of EBM readily lend themselves to the establishment of guidelines and other "slavish cookbook approaches." Besides, almost no "financial purchaser or manager" will resist either the urge or the justification of relying on "the best available evidence" when promulgating guidelines, dicta, or other instructions for physicians employed by the state or by corporations that govern the delivery of health care. These instructions can coerce the format of clinical practice by denying payment for treatments that lack suitable accolades from the processing source of EBM authorities. Because the source will include little other than the results of RCTs, a formal approval would have been lacking for the 30% to 50% of treatments, unsupported by RCT data, that were regarded as "convincing" in the two previously cited studies.[12,13]

A separate problem is that individual meta-analyses may not always be complete or promptly updated, so that today's authoritative prouncement may be dismissed tomorrow as incorrect or misleading. Furthermore, important single studies, particularly if not done as RCTs, may be omitted from the authorized collection. For example, when insulin first achieved a rapid reduction in diabetic acidosis and when penicillin first eradicated bacterial endocarditis, the results in both instances came from observational rather than RCT research, and each set

of results was reported in a single study. Despite the extraordinary efficacy of both treatments and their dramatic impact in clinical practice, neither study, if newly reported today, would be included in the Cochrane collection of authoritative evidence.

The advocates of EBM are obviously alert to the dangers of "top down cookbooks" and say they would join clinicians "at the barricades" should such dangers arise.[4] The threat of official, corporate, or private abuse will always remain, however, whenever any collection of information has been prominently heralded as the "best available evidence." A new form of dogmatic authoritarianism may then be revived in modern medicine, but the pronouncements will come from Cochranian Oxford rather than Galenic Rome.

REFERENCES

1. Evidence-based medicine working group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA.* 1992;268:2420–2425.
2. Davidoff F, Case K, Fried FW. Evidence-based medicine: why all the fuss? *Ann Intern Med.* 1995;122:727.
3. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. *BMJ.* 1995;310:1085–1086.
4. Sackett DL, Rosenberg WMC, Muir Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312:71–72.
5. Sackett DL, Rosenberg WMC, Muir Gray JA, et al. Evidence based medicine. Authors' reply. *BMJ.* 1996;313:170–171. Letter.
6. Rosenberg W, Donald A. Evidence based medicine. Authors' reply. *BMJ.* 1995;311:259. Letter.
7. Chalmers I. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann NY Acad Sci.* 1993;703:156–163.
8. Chalmers I, Haynes B. Reporting, updating and correcting systematic reviews of the effects of health care. *BMJ.* 1994;309:862–865.
9. Sackett D, Oxman A. *The Cochrane Collaboration Handbook*. Oxford: Oxford Collaboration; 1994.
10. Bero L, Rennie D. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA.* 1995;274:1935–1938.
11. Chalmers I, Enkin M, Keirse MJNC. eds. *Effective Care in Pregnancy and Childbirth.* Oxford: Oxford University Press; 1989.
12. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet.* 1995;346:407–409.
13. Gill P, Dowell AC, Neal RD, et al. Evidence based general practice: a retrospective study of interventions in one training practice. *BMJ.* 1996;312:819–821.
14. Feinstein AR. *Clinical Judgment.* Baltimore: Williams & Wilkins Co; 1967.
15. Elmore JG, Feinstein AR. A bibliography of publications on observer variability (final installment). *J Clin Epidemiol.* 1992;45:567–580.
16. Sacks H, Berrier J, Reitman D, et al. Meta-analysis of randomized controlled trials. *NEJM.* 1987;316:450–455.
17. Detsey AS, Naylor CD, O'Rourke K, et al. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol.* 1992;45:255–265.
18. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials.* 1995;16:62–73.
19. Emerson J, Burdick E, Hoaglin D, et al. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clin Trials.* 1990;11:339–352.
20. Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet.* 1996;348:596–598.
21. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analyses. *BMJ.* 1996;313:735–738.
22. Croog SH, Levin S, Testa M, et al. The effects of antihypertensive therapy on the quality of life. *NEJM.* 1986;314:1657–1664.
23. The Canadian Task Force on the Periodic Health Examination. *The Canadian Guide to Clinical Preventive Health Care.* Ottawa: Ministry of Supply and Services; 1994.
24. Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. *Lancet.* 1995;345:840–842.
25. Reid MD, Lach MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA.* 1995;274:645–651.
26. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med.* 1990;89:630–638.
27. Weed LL. *Medical Records, Medical Education, and Patient Care. The Problem-Oriented Record as a Basic Tool.* Chicago: Year Book Medical Publishers, Inc.; 1970.
28. Feinstein AR. The problems of the "Problem-Oriented Medical Record." *Ann Intern Med.* 1973;78:751–762.
29. Dudley L. Update on the Cochrane Collaboration. In *Clinical Epidemiology Newsletter.* 1994;14:2–3. (Dept. of Clin. Epid. and Biostat., McMaster Univ.)
30. Chalmers I. What would Archie Cochrane have said? *Lancet.* 1995;346:1300. Letter.
31. Haynes B. News from the Canadian Cochrane network and centre. Clinical Epidemiology Newsletter. 1994;14:3–4.
32. Taubes G. Looking for the evidence in medicine. *Science.* 1996;272:22–24.