Evidence-Based Medicine

Joseph Bernstein, MD, MS

Abstract

Evidence-based medicine refers to an explicit process of using and evaluating information to make medical decisions. Evidence-based medicine, perhaps contrary to popular perception, requires its users to embrace uncertainty in medical decision making because information that is simultaneously true and complete cannot be attained. Recognizing medicine's inherent uncertainty, proponents of evidence-based medicine advocate using a five-step process for sound decision making: formulate answerable questions, gather evidence, appraise the evidence, implement the valid evidence, and evaluate the process. The formulation of answerable questions requires categorizing the facts of the case in terms that allow comparison to evidence gathered from prior studies. The appraisal of the evidence uses the tools of clinical epidemiology to assess the validity and applicability of the evidence. Implementation refers to the construction of a clinical plan based on the evidence collected as well as on the physician's judgment and patient's preferences. Finally, evidence-based medicine requires continued evaluation and refinement. The methods of evidence-based medicine are especially germane to contemporary medicine as physicians practice under increasing demands to deliver optimal outcomes yet face an ever-expanding body of medical knowledge.

J Am Acad Orthop Surg 2004;12:80-88

Evidence-based medicine is a phenomenon of burgeoning popularity (Fig. 1), but what is it? To assess the value of evidence-based medicine[1] and estimate its impact on practice, we must agree on a definition. My primary purpose, therefore, is to define evidence-based medicine and to review the methods of critical appraisal and clinical epidemiology on which it is founded. A second question, more subtle but perhaps more interesting, is why there has been this recent emphasis on evidence-based medicine. To portray evidence-based medicine as a new mode of practice is to imply that traditional methods are not evidence based. This criticism is offered obliquely; nobody in the mainstream has yet equated Western medicine with necromancy or astrology. Still, the insinuation endures that medical practice lacks a basis in science.

The strongest allegation that medicine lacks a scientific basis comes from studies reporting practice variations that far exceed the variation in disease prevalence. For instance, the rate of back surgery among Medicare patients in Santa Barbara is five times the rate in the Bronx.[2] Because it is highly unlikely that the prevalence of disease differs by a factor of five, there probably exists a type of patient for whom the doctors in California, ostensibly basing their opinions on science, would recommend surgery, and for whom doctors in the Bronx, claiming the same mantle of scientific validity, would not. It would seem that the California doctors could be scientifically correct or the Bronx doctors could be scientifically correct, but not both.

The notion that wide variation in practice proves that medicine is not based on science demands rebuttal. That rebuttal takes the form of presenting how medical evidence is used in medical decision making and how even the best evidence can only minimize, but not eliminate, uncertainty and variation.

Evidence-Based Medicine Defined

The shibboleth "evidence-based medicine" refers to an explicit process of using and evaluating information to make medical decisions. The commonly quoted definition is attributed to Sackett et al:[3] "Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical

Dr. Bernstein is an Attending Surgeon, Veterans Hospital, Philadelphia, PA, and Senior Fellow, the Leonard Davis Institute of Health Economics, University of Pennsylvania School of Medicine, Philadelphia.
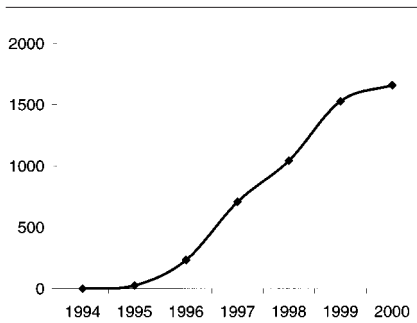
Figure 1 A plot of the number of MEDLINE citations of the phrase "evidence-based medicine" by year, 1994-2000.

expertise with the best available external clinical evidence from systematic research." This definition does not mention the randomized controlled trial. It does not refer to cost-effectiveness studies, algorithms, or managed care. What the definition does say, however tacitly, is that the process of making medical decisions is a difficult one. Why else would practitioners need to use "explicit and judicious" efforts?

Medical decision making is difficult because all clinicians come to face the following dilemma: to base medical decisions on either imperfect information or inadequate information. Call it the Uncertainty Principle of Clinical Decision Making: simultaneously true and complete information cannot be attained. To understand the horns of this dilemma, we must examine the nature of medical knowledge.

Medical Knowledge

The Dean of Cornell University Medical College once greeted the class of incoming students with an apology: "Half of what we will teach you is not true. The problem is, we don't know which half." Although obviously meant as a joke, like many jokes the apology derives its punch from the truth it expresses. If history is any

guide, much of what we currently know likely will be replaced or revised.

Given that at least some of our information is wrong, what actions should be taken? Limiting one's practice to only proven and established facts is no guarantee of safety: shunning everything unproven forces us likewise to reject the many statements that are true, just not proven. A Venn diagram representing the universe of medical knowledge illustrates this phenomenon (Fig. 2). One subset of this universe is those statements that are indeed true, shown in circle A. There is also the subset of statements that have been tested, circle B. The intersection of A and B, area C, represents the tested and true—the validated information all clinicians desire. But area C obviously does not contain all truths; area C is but a subset of area A. Therein lies the quandary. By limiting himself to statements from area C only, a physician will deny himself access to those true statements that have not been proved, area D. But if the physician reaches among the untested in hopes of grabbing from D, he also risks using false statements outside area A. This, then, is our choice: to use only validated

data, which may be inadequate, or to venture beyond the proven and employ possibly incorrect ideas.

One may attribute the existence of unproven truths to a deficiency of resources: there is not enough time and money to prove everything. But the more fundamental issue is that even with limitless resources, there still will be some true statements whose truth we cannot establish. This was demonstrated in 1930 by Kurt Gödel, in his groundbreaking Incompleteness Theorem.[4] A vernacular translation of the theorem is that "not all true statements are provably true."

We are therefore left with a set of perhaps unappealing options: imperfection to one side, inadequacy to the other. But still one must choose. In the fable, the hungry donkey was placed exactly between two bales of hay; pulled equally in both directions, it starved to death. In clinical practice, we cannot in every case rely on finding perfect information. Decisions must be made with the information at hand, and this information at times may be shaky. The methods of evidence-based medicine can help us make smart choices in the face of this uncertainty.
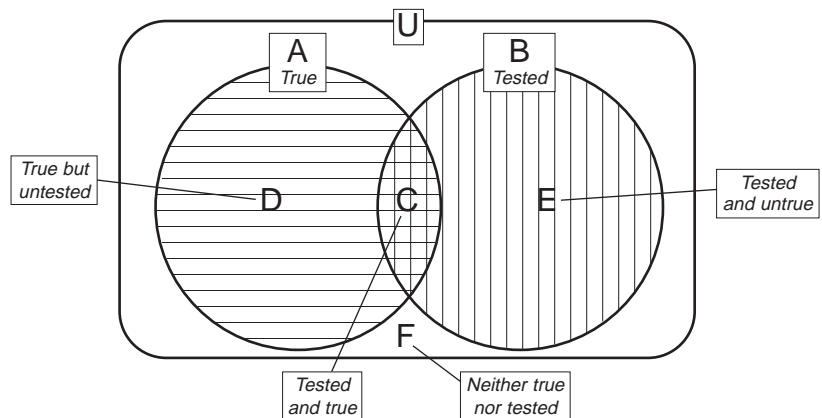

Figure 2 A Venn diagram representing the universe (U) of medical knowledge. Area A represents all true statements and area B, all tested statements. Area C represents all validated (tested and true) statements, but it does not contain all truths: it is missing Area D, which represents truths that have not been proved (true but untested). Area E represents invalidated knowledge, that is, all tested and untrue statements. Area F is those statements that are neither true nor tested. A clinician using untested knowledge hopes to use statements that fall within area D but may in fact sample from Area F.

Using Evidence-Based Medicine

Consider the case of Mr. Smith, a patient with an open tibial fracture, who was placed in a splint at the local hospital and sent to a trauma specialist who wishes to practice evidence-based medicine. The doctor recognizes that this patient requires an operation. He is uncertain, however, whether internal or external fixation is the better choice. If one's rule is to use a treatment only if it has been proved in large-scale prospective randomized double-blind controlled studies, then neither treatment is a viable option because such studies never will be found. (At the least, it would be impossible to blind observers to the presence or absence of a bulky external fixator.) But if the specialist stands by his rule and chooses neither treatment, he is in effect choosing yet a third option: leaving the patient in the splint. So a choice must be made. To choose wisely and not randomly, Sackett and Haynes[5] recommend employing these five steps of evidence-based medicine: (1) formulate answerable questions, (2) gather evidence, (3) appraise the evidence, (4) implement the evidence, and (5) evaluate the process.

Step 1: Formulating Answerable Questions

One cannot consult the literature asking, "What shall I do for Mr. Smith?" The question must be translated into an answerable form, which involves asking three questions: Who is Mr. Smith? What are the options? And what may happen if I choose one of those options? Formulating answerable questions thus requires defining the patient, the possible treatments, and the outcomes of interest.

Defining the patient begins with making a diagnosis and then identifying the attributes of the patient that one believes is relevant to management. These descriptors of the patient include age, gender, medical history, and, for example, location of the fracture and its grade as well as concurrent injuries. Recall that there are no studies of Mr. Smith himself, only of patients who share some of his attributes. Thus, the clinician must decide which of these attributes are determinant using "individual clinical expertise." For instance, most clinicians would agree that a study of tibial fractures in adult women is relevant to the case of Mr. Smith, whereas a study of tibial fractures in preadolescent boys is not. It may seem obvious that age is a more determinant attribute than gender, but this discernment is really the expression of individual clinical expertise: it is doubtful that the authors of the study on pediatric fractures will include an explicit warning not to apply the findings to adult patients.

The next step is to identify the possible treatments. This, too, is a personalized decision because some of the options presented in the literature may not be available or practical: scarcity of resources, bounds on physician skill, or the presence of other injuries may constrain the choices. (The concept that "the literature suggests technique X but in my hands, Y is better" is valid. Of course, one must consider that if one's hands are so tied, it may be best to send the patient to another physician.)

The final step is to define the outcomes of interest. Not all outcomes that researchers report are necessarily applicable to the decision of treating Mr. Smith. A study reporting outcomes after management of tibial fractures might discuss the length of time to bony healing, the percentage of patients who achieved bony union, the length of the hospital stay, the incidence of complication, or other outcomes. Deciding which of these are important, and in which proportion, must be done uniquely for each case.

Step 2: Gathering Evidence

Collecting a large number of research papers on a topic is relatively easy. Gathering "the best available external clinical evidence from systematic research,"[3] on the other hand, can be more difficult. Done ineptly, a literature search may be overwhelmed with irrelevant citations. (This echoes the lament of Coleridge's Ancient Mariner: "Water, water everywhere, nor any drop to drink.") Faced with a profusion of studies, identifying the "best" becomes a herculean task.

This article does not aim to teach how to search the literature efficiently; perhaps no single article can do that well. One acquires this skill by receiving hands-on instruction and by practicing. In some ways, learning how to search the literature is similar to the process of learning arthroscopy: one begins by taking a course, but constant refinement is needed.

Step 3: Appraising the Evidence

Proponents of evidence-based medicine say that physicians must evaluate medical evidence, just as courts process legal evidence. This comparison reminds us that evaluating medical evidence is a human endeavor open to error (Appendix, available at http://www.jaaos.org/cgi/content/full/12/2/e1). The legal comparison also highlights the folly of critics who say that medicine is not based on evidence simply because some medical statements have not been proved by perfect evidence. After all, we do not say that most murderers are sentenced without evidence just because few have been photographed in the act. Rather, we accept that there are various types of legal evidence, some better than others. (Indeed, eyewitness identification and even confession have been shown to be imperfect.) Also, in medicine as in law, the importance of the matter to be re-

solved by the evidence rightly affects the degree of certainty we demand.[6] Accordingly, the question is not, "Is the evidence perfect?" but rather, "How good is the evidence?"

In clinical research, not all sources of evidence are created equal. Among studies reporting on treatment outcomes, most epidemiologists would agree with the following pyramid of evidence:

Randomized controlled trial
Prospective cohort study
Retrospective cohort study
Case-control study
Case series
Case report
Expert opinion
Personal observation

The unique characteristics of each type of study are given in Table 1.

This pyramid of studies is a hierarchy in which each item on the list represents a method deemed more probative than the item below it. For example, a cohort study demonstrating X outweighs a case series that suggests the converse. Yet all forms of evidence, even those with low ranking, are legitimate. Therefore, absent other evidence, a case report can be the legitimate basis for action. Weak evidence is not the same as no evidence.

Considering the hierarchy of evidence as a pyramid—a structure that narrows at the apex—also hints at something important: the best forms of evidence are typically the rarest. For example, in a review of a sample of the orthopaedic literature composed of nearly 500 clinical research papers, only 33 were randomized controlled trials, and many of these, in turn, were marred by inadequate statistical power[7] (Fig. 3).

The Randomized Controlled Trial Versus the Case Series

The findings from a randomized controlled trial usually are the strongest form of medical evidence. To appreciate the advantages of a randomized controlled trial, it should be considered in contrast to weaker

Table 1

A Definition of the Study Types Used in Clinical Research

Study Type	Definition
Case report	A report of single patient
Case series	A report of multiple patients with the same treatment, but no control group or comparison group
Case-control	A study in which patient groups are separated by the current presence or absence of disease and examined for the prior exposure of interest
Retrospective cohort	A study in which patient groups are separated non-randomly by exposure or treatment, with exposure occurring before the initiation of the study
Prospective cohort	A study in which patient groups are separated non-randomly by exposure or treatment, with exposure occurring after the initiation of the study
Randomized controlled trial	A study in which patients are randomly assigned to the treatment or control group and are followed prospectively

methods, such as the more common case series. Imagine a hypothetical case series reporting the results of 50 patients with knee arthritis, all of whom were reported to be treated successfully by arthroscopic débridement. Does this series prove the efficacy of the operation? Perhaps not. One must consider that the study could be marred by various biases, such as the following:

Patient selection: If only those patients who are apt to respond to surgery are offered it, then the reported efficacy will be overstated.

Lack of context: The operation may be termed successful, but compared with what? To say that a treatment is effective, at minimum the contrast of a second group treated differently is needed. Even improvement shown by objective metrics is not sufficient
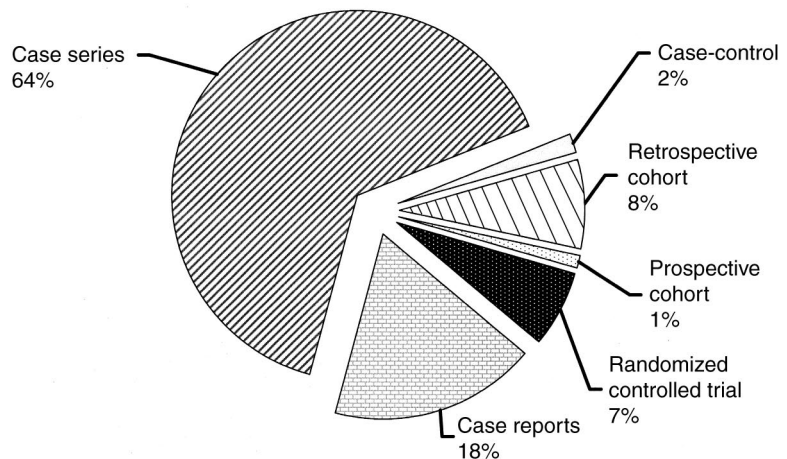

Figure 3 Distribution of the various study types among the clinical manuscripts in one year's volumes of the *Journal of Bone and Joint Surgery* (both American and British volumes) and *Clinical Orthopedics and Related Research*.[7]

to indicate that the treatment is actually effective; perhaps the natural history of arthritis includes periods of random improvement. The addition of a group that received no treatment (not even a placebo) would be needed to ascertain the actual value of the treatment.

Lost data: A retrospective study may not include all patients who were treated. Charts can be lost. Also, patients asked today about their symptoms years ago may not recall all things with equivalent reliability.

Subtle exclusions: Retrospective studies of an operation characteristically do not include data from patients for whom the treatment was intended but who did not receive it. This, too, is a bias. Patients too sick for surgery and therefore treated nonsurgically are excluded from the operative group, but selectively pruning the sickest patients in this way may affect the results.

Cognitive dissonance: A patient who chose to have an operation and subjected herself to all of the burdens associated with surgery may convince herself that the operation was helpful—even if it was not—to validate the decision to undertake it. By the same token, a surgeon who performs an operation may be less than impartial in evaluating his handiwork.

A double-blind prospective randomized controlled trial can offset these biases. In such a trial, all patients who meet the inclusion criteria are randomly assigned to treatment groups. This arbitrary allotment prevents skewing the results by offering the treatment to only those with the best chance of responding. A control group is created to offer contrast. Also, the prospective nature of the trial allows the researcher to define the variables of interest in advance and to make certain that comprehensive data are collected from all patients. Exclusion biases can be corrected by an "intention-to-treat analysis," in which patients are considered by the group to which they were assigned, not by the treatment they received. In addition, denying the patient and the evaluators knowledge about which treatment was given (so-called double blinding) limits the bias of psychological effects.

The value of the randomized-controlled trial lies in its resistance to biases, but these trials are not infallible. At the extreme are randomized controlled trials that seem to demonstrate the benefits of homeopathy,[8] a system that Park[9] has characterized as not only at odds with the tenets of medicine but also in conflict with the laws of biology, chemistry, and physics. It is far more likely that such trials were wrong or at least represent statistical noise. (By setting $P = 0.05$ as the criterion for statistical significance, we accept, a priori, a 5% chance that the study may see distinctions where none exist).

Randomized-controlled trials offer good evidence but are not perfect, of course. If you doubt that, consider this anecdote, related by Skrabanek and McCormick.[10] A patient was enrolled in a double-blind prospective randomized control trial studying the effect of a certain drug. One day, he abruptly asked his doctor why his medicine had been changed. The doctor, blinded to which group the patient was in, replied that he did not know what the patient was talking about, but still he asked, "What makes you think your medicine was changed?" To this the patient answered, "Up until now, when I threw the pills in the toilet, they sank; this week, they floated."

Case-Control Studies

Beyond its limitations, even a flawless randomized controlled trial may not be the best form of evidence. Some questions are more effectively addressed by other methods. To assess the long-term risk of harm, for example, a case-control study may be the most appropriate. To determine whether long-distance running causes osteoarthritis of the hip, for example, it would be absurd to create a randomized trial of subjects, compel some of them to run marathons, and then wait a few decades to measure the rate of arthritis in both groups. A better approach would be a case-control study, in which a researcher collects data from patients who currently have osteoarthritis and compares the amount of prior long-distance running in this group with the amount in a control group of subjects without arthritis.

The case-control method is no doubt imperfect because subjects may be more apt to recall prior exposures if they develop symptoms. Also, a cause-and-effect relationship is at best suggested, not proved. In general, however, if there is a long interval between exposure and disease presentation, or if the disease under study is rare, case-control studies may be best.

Expert Opinion

Expert opinion rests near the bottom of the pyramid of evidence, a position that has metaphoric significance: the teaching of experts is the foundation upon which all other knowledge rests. Good students turn to teachers and textbooks (not journal articles) to begin their study of a given area. But as a form of evidence, expert opinion is subordinate to systematic research. The reason is that history is full of examples in which experts were egregiously wrong.[11] For instance, William Harvey was criticized harshly by the "experts" for his radical notion that blood circulates. Scott[12] points out that the field of orthopaedics has its own examples.

In less rigorous fields, an appeal to authority is the highest form of evidence. For instance, many winning arguments in psychoanalysis not so long ago began with the phrase: "As Freud himself has noted…." Orthopaedics is based on a more objective

foundation than psychoanalysis, but we share with that field a method of professional training in which the novice is placed in the role of apprentice to the master. Because we are appropriately conditioned to accept the teachings of the experts when it comes to the basics, we may find it hard to reject their pronouncements when they veer into speculation. Yet we must. We are obliged to remember the hierarchy of evidence: expert opinion certainly trumps nonexpert opinion, but it is weaker than good clinical research.

Personal Observation

Personal observation is the weakest form of evidence. We can be easily fooled by our own personal observations, even when we realize that fact and try to compensate. To begin with, personal observations are studies based on small samples. Consider the clinician who withholds anticoagulants after knee replacement. If he observes a series of 10 patients who receive no medication and yet do not suffer a pulmonary embolism, he may wrongly conclude that anticoagulants are unnecessary. But in fact, if the risk of a pulmonary embolism without prophylaxis were, for example, 3%, the odds are still approximately three to one that the physician would not see an embolism in a group of 10 such untreated patients. Because the complication is rare, the fact that anticoagulants are necessary may be borne out only in a larger series.

Furthermore, personal observations are studies without controls. As Redelmeier and Tversky[13] noted, people want the world to make sense and will invent explanations for what they see, whether or not these explanations are warranted. If a patient is treated and gets well, it is human nature to attribute the cure to the treatment simply because the cure followed the treatment. This invites the post hoc ergo propter hoc fallacy, confusing associations in time with a cause-and-

effect relationship. If a control group were seen to get well, too, we would be less likely to make this mistake.

Finally, personal observations are subject to recall bias. If we fail to detect a synovial cell sarcoma of the posterior tibial tendon sheath in a patient who twisted her ankle, we are apt to overestimate the chance of this rare diagnosis in subsequent patients with ankle sprains.

Step 4: Implementing the Evidence

Correctly framing the clinical question indicates the treatment options for a patient. Appropriately consulting and evaluating the evidence suggests the possible outcomes from these treatments. Medical decision making should be no more complex than choosing the best option from the resulting list. Sometimes the process is very straightforward; at other times, it can be fraught with uncertainty. The source of the uncertainty is twofold. First, there can be gaps in the evidence, forcing the clinician to interpolate. Second, the value assigned to some of the outcomes may be variable. Thus, even with a robust scientific understanding of the possible treatments, uncertainty and variation abound.

Consider the following hypothetical case. A 21-year-old athletic patient presents with pain, swelling, and joint line tenderness 2 days after a twisting injury. After collecting and reviewing the evidence from the history and physical examination, and the imaging studies and joint fluid aspirate, the physician concludes that the patient has sustained a tear of the medial meniscus at the meniscosynovial junction (the so-called red zone) and no other injury.

There are three treatment options: meniscal repair, meniscectomy, and nonsurgical management. The formal means of choosing the best option among these three is to use a decision

tree[14] (Fig. 4). In the tree, all options are listed. Branches emanating from these options lead to terminal nodes representing the possible outcomes for a given treatment option. The branches are assigned weights proportional to the probability that a given outcome will be reached, and the terminal nodes are assigned values corresponding to the worth attributed to the outcome.

One can calculate the expected utility of a treatment option by computing the weighted average of the outcomes that result from choosing that option. The expected utility is the sum of the utility values of each terminal node (outcome state) multiplied by the likelihood that this outcome will occur. Choosing the best treatment option is therefore choosing the option with the highest expected value.

To fill in the decision tree for this example, the following (hypothetical) information about outcomes can be used. Red zone tears <1 cm have a 90% chance of healing without surgery. Meniscectomy in this population has a near-certain (≥99%) chance of symptom relief within 6 weeks, but it is associated with a 40% chance of developing arthritis within 10 years. Red zone meniscal repair has a 75% chance of success. It may seem that the selection of the best treatment now can be found with these data, but actually the most crucial information is missing: how much the patient values the possible outcomes. It is probable that the values ascribed to the outcomes will vary by patient.

In the case of a red zone meniscal tear, a repair may seem to be the proper answer. But imagine the attitude of a college senior on a basketball team looking forward to the national tournament in 2 months. The near certainty of being able to return to play in 6 weeks after meniscectomy may induce her to discount the risk of developing arthritis and to overvalue the short-term benefits of meniscectomy. Converse-
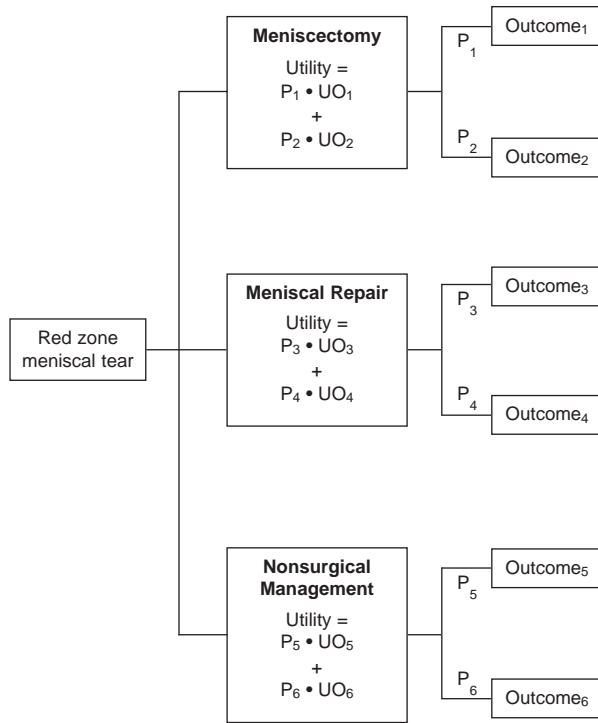
Figure 4   A simplified decision tree of the options and outcomes associated with a red zone meniscal tear. The boxes to the right represent the outcomes for treatment. The lines leading to them are weighted according to the probability that the given outcome state will be reached. P_1 is the probability that $Outcome_1$ will be reached. For clarity, only two outcomes per treatment are shown, although more are possible. The value of a treatment is $P_1 \cdot UO_1$ summed across all outcomes possible from the treatment. P_x = the probability of achieving an outcome. UO_x = the utility of $Outcome_x$.

ly, a patient whose father died during surgery may overvalue nonsurgical treatment, even if it puts his knee at risk.

Because many operations in orthopaedics are done for symptom relief, the crucial factor when deciding to recommend an operation is often the utilities assigned to the potential outcomes of treatment. Finding these utilities can be difficult. One must ensure that all outcomes are presented without bias and that utility values are correctly recorded. One must invent a scale on which all forms of happiness can be expressed. For instance, "perioperative death" and "able to walk 18 holes on the golf course" are some of the outcomes of interest for total hip arthroplasty. Although it appears that these outcomes are not at all comparable, patients in fact will risk the former to achieve the latter. One also must be able to account for risk aversion and discount future utility to its net present value. Thus, the "proper" treatment of a peripheral meniscal tear will hinge on patient preferences—even with good outcomes evidence in hand. For this reason, there may be more back surgery in Santa Barbara than in the Bronx simply because residents in each locale have different utility values and personal preferences. (There are, of course, additional plausible explanations.)

In addition, this list of evidence on meniscal outcomes has critical gaps. There is no information on the size of this patient's tear or on the accuracy of an estimate based on imaging studies. The healing rate for tears <1 cm is specified, but is the rate zero for larger tears, or is that rate dependent on the size? What happens if the tissue does not heal? Finally, what are the complications of surgery, and how likely are they to occur?

To make the right decision in this case, additional data are needed, so that a more intricate decision tree results than that shown in Figure 4. The outcome for failed nonsurgical therapy must be expanded, for example, because a meniscectomy or repair may be attempted after nonsurgical management fails. Also, the surgical option is more accurately considered a diagnostic arthroscopy because it may be that a small tear is found and attempted repair would be declined; or an irreparable tear may be encountered, necessitating a meniscectomy. Finally, dichotomous variables such as "healed" or "repair failed" are actually points on a continuum.

This example demonstrates that even a fairly straightforward case with fairly extensive outcomes data in hand produces complexities that, in Eddy's apt phrase, "far exceed the capacity of the unaided human mind."[15] Until better tools are developed for discerning what patients want and until guidelines are invented for applying these preferences, the collected medical evidence cannot be used to greatest advantage.

Step 5: Evaluating the Process

Evaluating performance is essential to practicing evidence-based medicine, although problems can arise. The main difficulty is that evaluation traditionally has been based on personal observation, and such evidence is subject to bias. Also, one can be fooled by statistical quirks when dealing with small numbers. Because of the limitations placed on personal observation, evaluation is best done as

part of a larger monitoring effort. For example, a registry of joint replacement procedures may be helpful. Even the data collection of managed care organizations can provide useful positive feedback. Consider the primary care physician who orders not even one bone density study in a year. Review of this doctor's utilization data could be used to suggest that he may not be giving sufficient consideration to the prevention of osteoporosis. Of course, the process of evaluation must elicit the cooperation of physicians to succeed.

Overcoming Resistance to Evidence-Based Medicine

Resistance to the practice of evidence-based medicine is likely to be found among physicians who generally abide the theory behind evidence-based medicine but nonetheless do not embrace it. These physicians may be concerned that the practice of evidence-based medicine demands high-level mathematics. In point of fact, however, evidence-based medicine requires no special facility with numbers. The practitioner of evidence-based medicine need not become a statistician; he or she is charged only with knowing how to use what statisticians produce. In Greenhalgh's useful analogy, doctors are not asked to build cars, only to know how to drive them.[16]

More vigorous resistance may be shown by doctors who think that evidence-based medicine is a pretext to undermine physicians' professional autonomy, that under the constraints of evidence-based protocols, doctors will be reduced to automata. Echoing the actions of 19th century Luddites in Nottinghamshire who smashed the cotton looms to protect their jobs, these practitioners may attempt to undermine evidence-based medicine in the name of halting progress in the "wrong" direction.

Of greatest concern is the resistance offered by well-meaning, well-intentioned physicians who deem evidence-based medicine simply too unwieldy for practice. This concern is legitimate. More than ever before, doctors are bereft of discretionary time, and reading and evaluating the literature is time consuming. For example, there were 479 MEDLINE citations for the journal *Spine* in the year 2000 alone. As such, even a brief contemplation of this one journal may consume more time than some doctors are able to devote to reading. The creators of evidence-based medicine recognize this and therefore encourage groups of doctors to perform systematic reviews, then from those reviews to develop guidelines for practice. A systematic review is not one in which an expert offers a point of view and then supports it with literature citations. Rather, it is one in which a bibliography is collected according to a predetermined scheme and evaluated against an articulated standard. As such, a systematic review does not have to be written by an expert, and perhaps should not be.

Systematic reviews logically yield practice guidelines. A guideline based on a systematic review amalgamates the first three steps of evidence-based medicine: the formulation of an answerable question, the collection of evidence, and the evaluation of that evidence. It offers a clinician a thorough review without forcing him or her to consult the primary literature personally. The key is that the guideline must be robust enough to account for variation in doctors' skills and patients' preferences. Doctors, in turn, must not apply their guidelines with procrustean rigidity. As noted by the 14th century French surgeon, Henri de Mondeville: "Anyone who believes that the same thing can be suited to everyone is a great fool, since medicine is practiced not on mankind in general, but on every individual in particular."[17]

Summary

Evidence-based medicine may be a new phrase, but it rests on an old concept: empirical outcomes matter. This idea dates back 100 years to the orthopaedic surgeon Codman, who stressed the measurement of end results. In fact, one can find the methods of evidence-based medicine used in 19th century Paris, where bloodletting was discredited by Pierre Louis. Louis performed what the *New York Times* described as one of the first clinical trials: "He treated people with pneumonia either with early, aggressive bloodletting or less aggressive measures; at the end of the experiment, Dr. Louis counted the bodies. They were stacked higher over by the bloodletting sink."[18] This result—not some theoretical musing—is what lead to the abandonment of bloodletting from medical practice.

Despite its venerable pedigree, evidence-based medicine is poised to respond to contemporary demands. The proliferation of information, primarily through the expansion of the Internet,[19] but also within the growth of traditional journals, requires that clinicians take a new tack in gathering and evaluating information. The problem is not the lack of data but the excess. We need to find meaning within a welter of information, and evidence-based medicine allows us to do that.

Finally, we must bear in mind the constraints on what science can inform us. The dramatist Berthold Brecht (who, incidentally, also attended medical school) wrote, "The chief aim of science is not to open a door to infinite wisdom; it is to set a limit to infinite error" (*The Life of Galileo*). Thus, when we say that evidence-based medicine is scientifically based, we mean, not that evidence-based medicine is a strategy for being right all the time—science cannot produce that—but rather that it is a tool to decrease the chance of being wrong. The goal of minimizing error is probably

within our grasp; and striving for it is certainly within our responsibility.

References

1. Lalwani SI, Olive DL: Problems with evidence-based medicine. *J Am Assoc Gynecol Laparosc* 1999;6:237-240.
2. Dartmouth Medical School: *The Dartmouth Atlas of Musculoskeletal Health Care.* Chicago, IL: AHA Press, 2000.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS: Editorial: Evidence based medicine: What it is and what it isn't. *BMJ* 1996;312:71-72.
4. Hofstadter DR: (ed): *Gödel, Escher, Bach: An Eternal Golden Braid.* New York, NY: Basic Books, 1999.
5. Sackett DL, Haynes B: On the need for evidence based medicine. *Evidence-Based Med* 1995;1:4-5.
6. Pauker SG, Kassirer JP: The threshold approach to clinical decision making. *N Engl J Med* 1980;302:1109-1117.
7. Freedman KB, Back S, Bernstein J: Sample size and statistical power of randomised controlled trials in orthopaedics. *J Bone Joint Surg Br* 2001;83:397-402.
8. Linde K, Melchart D: Randomized controlled trials of individualized homeopathy: A state-of-the-art review. *J Altern Complement Med* 1998;4:371-388.
9. Park RL: (ed): *Voodoo Science: The Road From Foolishness to Fraud.* New York, NY: Oxford University Press, 2000.
10. Skrabanek P, McCormick J: (eds): *Follies and Fallacies in Medicine.* Buffalo, NY: Prometheus Books, 1990.
11. Hellman H: (ed): *Great Feuds in Medicine: Ten of the Liveliest Disputes Ever.* New York, NY: John Wiley and Sons, 2001.
12. Scott WN: Preface, in Insall JN, Scott WN (eds): *Surgery of the Knee,* ed 3. New York, NY: Churchill Livingstone, 2001, p xiii.
13. Redelmeier DA, Tversky A: On the belief that arthritis pain is related to the weather. *Proc Natl Acad Sci USA* 1996;93:2895-2896.
14. Bernstein J: Decision analysis. *J Bone Joint Surg Am* 1997;79:1404-1414.
15. Eddy DM: (ed): *Clinical Decision Making: From Theory to Practice.* Sudbury, MA: Jones and Bartlett, 1996.
16. Greenhalgh T: (ed): *How to Read a Paper: The Basics of Evidence Based Medicine,* ed 2. London, England: BMJ Books, 2001.
17. Skrabanek P: Fat heads. *National Review* 1995;47:43.
18. Patterson K: What Doctors Don't Know (Almost Everything). *New York Times.* May 5, 2002:6,74.
19. Beredjiklian PK, Bozentka DJ, Steinberg DR, Bernstein J: Evaluating the source and content of orthopaedic information on the Internet: The case of carpal tunnel syndrome. *J Bone Joint Surg Am* 2000;82:1540-1543.
20. Guyatt G, Rennie D: (eds): *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* Chicago, IL: American Medical Association Press, 2002.