Commentary

The Use of Evidence and Cost Effectiveness by the Courts: How Can It Help Improve Health Care?

David M. Eddy
Kaiser Permanente Southern California

Although the relationship between health care and the courts has always been strained, recent developments in health care are creating new pressures that threaten to make it worse. Two particularly important ones concern the use of evidence and the application of cost-effectiveness analysis. This commentary will address these issues from the viewpoint of someone in health care. I will describe an ideal relationship with the courts; the current ways the courts address evidence and cost-effectiveness; the strengths and weaknesses of those methods; and what both sides can do to help the health care system function better. Both the legal and health care systems are enormously complex. To avoid obscuring the main points, I will focus on the big picture and plead leniency for the inevitable oversimplifications.

What Can We Expect from the Courts?

From the viewpoint of someone in health care, the purpose of the legal system is to help the health care system function smoothly and correctly, consistent with the principles laid out in the constitution, its amendments, and our laws. We expect the courts to do this primarily by ensuring that contracts are followed, by resolving disputes, by judging whether reasonable expectations of performance are met, and by meting out awards and penalties. The presumption is that we in health care know what we are trying to do, and the legal system is there to help us do it. The legal

Journal of Health Politics, Policy and Law, Vol. 26, No. 2, April 2001.

system should be a facilitator and referee, not an initiator. Only if we fail to articulate our goals, or if our goals violate the principles contained in our constitution and laws, should the legal system be expected to define our goals for us.

What Is the Health Care System Trying to Do?

The first sign of trouble ahead is that this is a surprisingly difficult question to answer. Our health care system is phenomenally amorphous, heterogeneous, and complex. There is no single structure or coherent leadership; we have highly variable coverage of different populations; and there are dozens of different ways of organizing, financing, and delivering care, all of which are undergoing constant change. Depending on where they sit in the system, different participants can have goals that are not only different but that actually compete with one another. Even single participants can have multiple goals that are in internal conflict. For example, a pharmaceutical company's goal of improving health can argue for putting a low price on a drug, at the same time that its goal of maximizing shareholder profits can argue for a higher price. A physician paid a fee for a service has an incentive to do a test of little value, while the same physician working in a capitated setting has the opposite incentive. Indeed, it is pushing things to talk at all about a "system," with the implications that word has for coherence and consistency.

But despite the complexities and inconsistencies of the current system, it is still helpful to imagine that there is some overriding purpose to the collection of activities we call health care. To articulate that overall purpose, we have to look beyond the immediate commercial and personal objectives of the participants and think about the ultimate purpose—the lofty goal to which all participants would lay claim. For the ultimate purpose I would propose something like "to deliver the highest possible quality of care to people, within whatever cost they want to pay."

The generality of these words intentionally skips over a variety of difficult issues. For example, it finesses all the issues around how health care is organized, financed, or delivered; whether we have a single-payer system or a heterogeneous one; who pays the immediate bills (people, employers, governments); whether access to health care is a "right" or a "commodity"; and the tension between maximizing care for an individual versus optimizing care for a population.

Furthermore, this objective does not make any ethical judgments about whether or not there *should* be any limit on the amount of money we spend on health care; it merely acknowledges that there *may* be a limit on what at least *some* people can or want to pay to receive health care, and that *if* there is such a limit, then the system that delivers health care should respect it.

In fact, let us be quite explicit about the latter point and identify two possible settings vis-à-vis costs and budgets. Imagine that in one setting people are willing to pay any cost for any care the physician recommends. In this setting, which we might say has an open-ended or unlimited budget, providers do not need to consider costs when making decisions about appropriate treatments, and aphorisms such as "costs should play no role in medical decisions" are applicable. Imagine that in the other setting there is a limit to what people can or want to pay. They either directly or indirectly (through market pressures) instruct plans to limit their costs. This in turn limits the resources available to plans for providing care. (I will use the word "plan" in a general sense to include any organization that is contracted to pay for or deliver a service, whether it is through traditional insurance or prepaid "managed" care.) In this setting, which we might say has a closed or limited budget, providers *have* to consider the cost of a treatment, if for no other reason than to ensure that the total stays within the budget each year. Needless to say, settings with limited budgets can vary widely in the tightness of the budget and therefore in the extent to which costs need to be considered. But notice again that at this point we are only acknowledging that these two settings and their variants can exist; we are not making any judgments about which is better, more desirable, or more ethical.

Now let us return to the statement of ultimate purpose. As general as this one is, it has the virtue that, no matter how any of the specific issues around the organization and financing of care are resolved, or even if they are never resolved, there will always be a need to do certain things. Those things include determining the effectiveness of different treatments and making judgments about whether a treatment's benefits outweigh its harms. (I will use the word "treatment" to include any type of health intervention, including prevention, testing and support care, as well as the more usual treatments.) In settings where budgets are limited, additional needs are to determine the costs of different treatments and to make judgments about their relative "values"—whether a treatment's net benefits are worth its costs or it is a good use of resources. These deter-

minations and judgments are "fundamental" in the sense that they are intrinsic to the provision of health care and must be done no matter how we choose to organize and finance care.

Interaction with the Courts

These determinations and judgments are fundamental in another way; they are the sources of the conflicts that bring health care before the courts. They do so under two main headings: coverage and malpractice.

Coverage. The question is whether a plan is obligated to pay for a treatment. In theory a plan's obligation to cover a treatment will be specified in the contract written between the plan and those who purchased the services (the "purchasers"). The question then is: Does the contract that describes the covered services include or exclude the treatment?

In theory this should be an easy question to answer, because most plans and purchasers agree in principle on what they want the contract to cover. They all generally agree that coverage should be limited to activities that address health conditions such as diseases and injuries, that involve health interventions, that are effective in doing what they are supposed to do, whose expected benefits outweigh any potential harms, and that are "reasonable" is some sense that we assume should be obvious. Good language for these criteria should do the trick.

Unfortunately, the actual language that specifies coverage suffers from several problems. The most fundamental is that there are very few precise ways to define the boundaries of coverage. For example, there is no easy way to separate a medical condition or disease from, say, a normal variation or aging; or a treatment from, say, a lifestyle. Nor are there unambiguous ways to define "effective." The magnitude of a treatment's effectiveness, the probability of its effectiveness, and our certainty about both of those, all vary continuously. A third problem is that even if there were clear definitions, they would be difficult to see because of incomplete data. Poor evidence especially plagues attempts to determine effectiveness, harms, and cost-effectiveness.

These problems have several consequences. One is that contract language varies enormously from plan to plan. Another is that definitions tend to be extremely vague (e.g., "reasonable and necessary" or "appropriate"), subjective (e.g., something is "investigational" if it is "not a generally accepted medical practice"), and unpredictable (e.g., something is covered if the treating physician recommends it). References to cost-

effectiveness are especially elusive. The concept may be included as an explicit criterion (rarely), as an implicit criterion (e.g., buried in the notion of "appropriateness"), so narrowly that its applicability is trivial (e.g., "the least expensive way to achieve an identical benefit"), or not included at all. In general, existing contract language is so flabby that it is almost worthless for creating accurate expectations, making decisions, avoiding disagreements, or settling disputes.

This situation would not be so bad if we didn't expect many disagreements. Unfortunately, two additional problems create a set up for disagreements. One is that most of the homework and negotiations around contracts are done by third parties. That is, those who are actually paying the premiums (e.g., employers) are usually different from those who are actually receiving the treatments (e.g., employees). Individuals are often unaware of what is in a contract until some treatment they want is denied. The second is that, even when people are making their own purchasing decisions, their incentives and desires change dramatically between the time the purchasing decision is made and the time the care is needed. As with any transaction that involves the prediction of future events and chance, choices are different when an event is a possibility versus a reality; people want to change their bets after the wheel has been spun. When we add the fact that for coverage decisions the stakes are usually very high—often involving life and death, and thousands if not hundreds of thousands of dollars—the recipe is for trouble.

Malpractice. The issues here are different but equally slippery. Here there is no dispute about whether the treatment is covered. The disputes are about whether a patient got a treatment that was indicated and/or whether the treatment was performed properly. To determine if a patient got an indicated treatment, the issues are whether, for that particular patient with his or her particular indications and contraindications, the treatment in question is effective, has expected benefits that outweigh its expected harms, and is the most appropriate compared to other available treatments. That is, does the treatment meet the standard of care? For performance, the issue is whether the actual performance corresponds to reasonable expectations (e.g., the correct drug was given at the correct dose, the correct leg was operated on, no instruments were left in the abdomen).

At least three problems make issues of malpractice difficult. First, because individual patients have a bewildering variety of individual characteristics, histories, signs, symptoms, and behaviors, the range of uncer-

tainty around the effectiveness of any particular treatment in any particular patient can be enormous and easily capable of harboring wide differences of opinion. Second, the outcomes of virtually all treatments are highly probabilistic: a treatment does not cause a particular outcome to occur, it only changes the probability it will occur. Furthermore, the effect of the treatment on the chance of the outcome is often small compared to other factors such as the patient's characteristics (risk factors) and the severity of the patient's disease. This means that practitioners cannot control the outcome nearly as much as a patient might expect. The third problem was also stressed in this special issue in the article by John M. Eisenberg: that medical decisions have to be made prospectively, whereas determinations of malpractice are made retrospectively. Prospectively, the practitioner may not know the patient's condition with certainty and never knows the outcome the treatment will produce. The choice of a treatment has to be based on the odds. Retrospectively, everything is known, it is obvious whether the choice of the treatment was right or wrong, and the patient is either happy or upset. It can be very difficult for juries and even other physicians to appreciate the appropriateness of a decision made prospectively, after the ultimate answer has become known. Given all these problems, there is considerable room for misunderstandings, disappointments, and disagreements.

The Method of the Courts

When cases of coverage or malpractice come before a court, the ultimate problem for the court is to determine whether what was done was appropriate in some sense. To deal with the inherent ambiguities in a word like "appropriate," it is extremely helpful for the court to have some reference or comparison. The underlying idea is that there is a correct "standard of care" out there, either a correct conclusion about the coverage of a treatment or a correct way to manage a patient. If several treatments provide roughly equal benefits to a patient, there may be more than one way to achieve the standard of care, but there is still a threshold of quality that we call the standard of care. The task of the court is to determine the standard of care and make a judgment about whether what was done meets that standard.

The traditional method used by the courts to determine the standard of care is to apply what might be called an "internal test." Rather than try to determine the standard of care for itself, the court hears the testimony of experts from within health care who are expected to know the stan-

dard of care and compares their answers to what was done. What the court actually hears depends on what the lawyers want to present, but it generally involves the answers to any of three questions: What do the majority of practitioners do? (What is the "community standard"?) What do groups of experts say (e.g., the recommendations of specialty societies or national panels)? Or what does an individual expert believe? While the community standard could conceivably be determined empirically, perhaps by polling community physicians or analyzing databases, it is usually learned by asking experts.

So in the end, it comes down to the testimony of experts. If all the experts agree, the court's job is easy. If the experts disagree, which they always do in an adversarial setting, then the court has to evaluate the credibility of the experts and decide which ones are correct. To do this, the court has two main sources of information. One is an assessment of any supporting evidence presented by the expert. The other is an assessment of the overall credibility of the expert based on such things as the expert's training, academic affiliations, experience, articulateness, and ability to build rapport with a jury.

At this point we should pause to note that the heavy reliance on experts by the court raises a question about differences in the "standards of evidence" used by the courts and health care. In health care, evidence means empirical observations. They may be systematic observations using rigorous experimental designs or nonsystematic observations (e.g., experience), but they are all empirical observations of real events. In the legal system, "evidence" means whatever is put before a court. On the surface this may appear to be fundamentally different standards of evidence. But in fact, at least for the use of experts in health care cases, the theoretical intent is the same. To see this we first need to appreciate that in the end, the only way for anyone to learn a fact is through empirical observations. (The only other methods are things like revelation and divination, which do not meet either the court's or health care's definition of evidence.) When the court asks an expert to express a belief about a fact, the court is presuming that that expert will be basing his or her belief on empirical observations, not revelations, dreams, or ancient texts. Thus the role of the expert is to serve as an interpreter of empirical observations. The assumption is that the beliefs developed by the expert will be accurate interpretations of all the pertinent empirical evidence. Thus the expert is not there to introduce a different kind of evidence that is nonempirical, but to introduce the conclusions learned from the empirical evidence. The evidence ultimately desired by the courts is the evidence also used in health

care, it is just obtained secondarily, through the expert, rather than primarily, through direct study of empirical observations.

Having said all this, it is also important to note that what the court actually hears can vary from pure primary evidence to pure secondary evidence. The actual information imparted to the court by an expert may be a detailed description of the empirical evidence, without any judgments about what the evidence says about the case, or a pure opinion about the case, without any reference to the evidence behind the opinion, or anything in between these two poles.

The Assumptions behind the Court's Methods

On its face, the court's approach seems quite reasonable. Certainly it is far easier for a court to listen to and evaluate experts than it is for a court to learn medicine, learn experimental methods, read scores of reports of clinical trials, weigh benefits against harms, and so forth. It is also very respectful of the professionalism in medicine. However, the validity of its approach rests on several assumptions. The assumptions differ slightly, depending on whether the method of learning the standard of care is by reference to a community standard, through the testimony of individual experts, or through the conclusions of a group of experts.

When the reference is to a community standard, the main assumptions are that

- There is a standard of care. That is, everyone agrees generally on the threshold of quality that should be achieved and that there is at least one way to achieve it.
- There is a community standard and it matches the standard of care. That is, the majority, or at least a substantial proportion of practitioners know what the standard of care is and do it.
- There is no other practice that a majority or substantial proportion of practitioners do that does not meet the standard of care. (Otherwise such a non-standard-of-care practice might be interpreted as the standard of care.)
- To be meaningful as an indicator of quality, the standard of care that is revealed through the community practice should be consistent across communities, unless there are differences in physical resources that can explain differences in practices.
- The community standard can be learned from experts. That is, at least some experts know what practitioners are actually doing, presumably through observation or analysis of data.

- These experts will present their beliefs without bias.
- If experts disagree about the community standard, the courts can differentiate the "true" experts from the "false" experts. This in turn requires assumptions that (1) to the extent an expert has presented the empirical evidence on which his or her beliefs are based, the courts can evaluate that empirical evidence, and (2) to the extent an expert's credibility is based on less direct measures (e.g., training, affiliations, experience), there is a good correlation between those indirect measures and the accuracy of the expert.

When individual experts are the sources of information about the standard of care, some additional assumptions are that

- There are at least some experts (true experts) who know the true standard of care.

When the reference is to a committee of experts, there is usually an additional assumption that

- A group of experts will be more accurate and less biased than individual experts. Two ways this can occur are that (1) whereas no single expert may have all the required knowledge or experience, a combination of experts will; (2) whereas a particular expert may be biased, the biases of a collection of experts will cancel out.

There is one more important assumption that is contained in the idea of the existence of a standard of care and that should be made explicit. It is that all the practitioners who make up a community standard, or all the experts who provide individual testimony or serve on committees are in basic philosophical agreement about what constitutes high-quality care. If this is not the case, then there might be a standard of care for each philosophy. This assumption harks back to the premise that the role of the legal system is to help the health care system do what it is trying to do. Unless the health care system agrees with itself about what it is trying to do, the legal system could be led down hopelessly tangled paths as practitioners and experts play out their different philosophies.

Problems with the Court's Methods

Each of the assumptions underlying the court's approach to health care cases seems reasonable, and until fairly recently there was little reason to question them. Unfortunately, we now have incontrovertible evidence that virtually none of them are reasonable.

Begin with the use of experts to learn the community practice. The assumption is that an expert can somehow "see" what all (or a random sample?) of the physicians in a community are doing either first hand or by analyzing data. In fact it is a major research task to figure out what practitioners in a community are doing. When an expert answers a question about a community standard it is extremely unlikely that he or she has any real data on actual practices. It is far more likely that what an expert believes is the practice in a community is what the expert personally believes *should be* the standard of care. In practice, questions to an expert about a community standard are really more like questions about a personal belief.

But more important, the idea that a community standard exists at all has been shattered by hundreds of studies of variations in practice patterns and inappropriate care. When rates of procedures vary across communities by factors of two, five, ten and more, and when 10 percent, 20 percent, 50 percent, even 70 percent of practices are judged by peers and experts to be inappropriate or equivocal (providing no advantage of benefits over harms), it is impossible to believe that there is a single community practice out there that the majority of practitioners are following. Indeed, given the very high rates of inappropriate care that can prevail in communities, if we actually measured what practitioners were doing and used that to define the standard of care, we would run a high risk of installing an inappropriate practice as the standard of care. The well-documented overuses of hysterectomies, antibiotics, bypasses, and C-sections are examples.

Nor can we continue to take the opinions of experts at face value. New studies continually reveal that practices that were once accepted without doubt can turn out to be worthless or even harmful. We were wrong about diethylstilbestrol, radical mastectomies, erythropoetin for anemia in end-stage renal disease, hyponatremic encephalopathy, treatment of ingested poisons, hormone replacement therapy for heart disease, and class I anti-arrhythmics for heart attacks. Experts from top universities with the most experience testified under oath that high-dose chemotherapy for late-stage breast cancer would produce 20 to 30 percent long-term cure rates. Randomized controlled trials later proved them wrong. This is not to say that all experts are always wrong. It is to say that we cannot assume they are right, and there is no easy way to tell when they are and when they are not from their credentials or enthusiasm.

Nor can we take at face value the recommendations of a group of experts, such as a specialty society committee or a national panel. In

addition to examples of committee recommendations that have later been proven wrong, we have the fact that the recommendations of many committees disagree with each other. This is especially obvious for treatments that represent organizational turfs. Examples include cancer screening and the management of back pain.

At this point it is important to stress that the fallibility of experts and the variations in community practices do not mean that experts and practitioners are inadequate or arbitrary, much less petty, dumb, or greedy. The fundamental problem here is that human biology, diseases, and medicine are phenomenally complex, and the complexity simply exceeds the capabilities of human subjective reasoning. There is no more shame in this than there is in the fact that people can't calculate their income tax in their heads. It is tempting to imagine that there was a time when the practice of medicine was simpler and experts could form accurate beliefs from subjective reasoning, but treatments like leaching and strychnine suggest that the mismatch in complexity has always existed. But whatever the strengths of subjective reasoning in the past, it is clearly inadequate now. Recognizing this fact does not imply disrespect, it implies confidence that we all want to provide the best care possible, and that there is no way to correct a problem without facing it.

There is one more threat to the court's method that needs to be addressed. It strikes at the fundamental assumption that there is *a* standard of care. As stated above the concept of a single standard of care implies that everyone who is a part of determining that standard—physicians, patients, purchasers, and others—agrees on what constitutes high-quality care. Until recently that was generally true. However, two recent developments have caused serious splits in the principles people use to define the standard of care. Not surprisingly, the principles involve evidence and costs. Depending on which principle one applies, one can arrive at different standards of care.

With respect to evidence, the split is between a traditional approach that gives heavy weight to it, and a more recent approach that is much more skeptical of subjective reasoning and much more strict about requiring good evidence. In this context it is important to distinguish "subjective reasoning" from the related ideas of "clinical judgment" and the "art of medicine." Subjective reasoning is the attempt to pull all the facts that are pertinent to a question about a treatment into one's head and synthesize them into a conclusion, without using formal, explicit methods such as a statistical analysis of data. Clinical judgment and the art of medicine both can include subjective reasoning. But they have additional

elements that are not included in subjective reasoning, such as empathizing with a patient, discussing a patient's preferences for benefits and harms, addressing a patient's hopes and fears, and helping a patient cope with a problem. Everyone agrees that these are important and there are no formal methods for accomplishing them that can replace clinical judgment and the art of medicine. But there *are* formal methods that can enhance subjective reasoning when it comes to estimating the expected effect of a treatment on the probability of some health outcome.

With this distinction in mind, the different weights placed on subjective reasoning have profound implications for the standard of care. The skepticism about subjective reasoning and requirement for empirical evidence can lead to a policy that a treatment should not be affirmatively recommended through a guideline or made the standard of care (or if it is a new treatment, covered by the plan) unless there is good evidence it is effective and beneficial. On the other hand, those who prefer the former approach would acknowledge that good experimental evidence is highly desirable, but they would argue that it is unrealistic and too strict to require good evidence, and that a practice might be recommended or even made the standard of care if there is suggestive evidence, or if it is supported by biological theory, or even if it just provides a hope of effectiveness.

With respect to cost, the split is between a position that took root in the era of open-ended budgets fed by indemnity insurance and a more recent position that is trying to respond to marketplace demands that costs be controlled. The former holds that costs should play no role in medical decisions. The latter says just the opposite, that costs have to be considered in medical decisions. The differences can create two very different standards of care. In settings where there are no limits on costs, a standard of care should include treatments that provide even small benefits, regardless of their costs. Examples are tPA as a blood thinner (instead of streptokinase) for acute heart attack, nonionic (instead of ionic) contrast agents (dyes) for certain radiographic procedures, and annual Pap smears (instead of three-year Pap smears) for cervical cancer. However, in settings where budgets are limited, a very different standard of care should apply. There, the standard of care should include streptokinase, ionic contrast agents, and three-year Pap smears, because the marginal gains in benefit of the more expensive alternatives are tiny compared to their costs, and the money could be put to better use elsewhere.

These disagreements about the roles of evidence and costs in defining the standard of care can wreak havoc on court cases. The risk is that

decisions made in one setting, where one standard of evidence or cost is being applied, will be judged by experts from another setting that applies different standards. A good example is the case of high-dose chemotherapy and autologous bone marrow (HDC/ABMT) or stem-cell transplant for late-stage breast cancer. Refusals to cover this treatment, on the grounds that it was experimental, were based on a lack of evidence from well-controlled trials of an effect on health outcomes. But the experts who testified for coverage developed their beliefs from biological models ("more is better"), uncontrolled clinical series, and the effects on intermediate biological outcomes (response rates). If the philosophical differences had been made clearer to the courts, and if contract language had been more precise about which standard of evidence would be applied, the courts might have been able to sort it all out. But as the cases actually played out, the courts came to very different decisions in different cases, reflecting the conflicts in philosophies behind the testimonies presented to them. Far from helping the courts deal with our internal philosophical disputes, we are using the courts as a battleground for fighting them.

What Can Health Care Do to Address These Problems?

These issues raise grave problems for health care, and it is going to take every bit of intelligence, honor, and courage to address them. Both conceptual and practical steps need to be taken.

Conceptual Steps

The first and most important conceptual step is to resolve the issue of conflicting philosophies and standards of care. This does not necessarily require that one position be agreed on and the others discarded. To be sure, agreement on a single vision for health care would be desirable for the simplicity, consistency, and public confidence it would create. But if that is not possible, which appears to be the case for at least several years, this conceptual step only requires that each position be recognized and respected as ethical and legitimate. The key is for all the actors in health care to acknowledge that there can be different settings in which health care is delivered, and that the standards of care can be different in those different settings.

To understand what this might look like let us use the two possible set-

tings introduced earlier as examples. In the setting of limited budgets, costs would be considered as critical variables in determining the standards of care, and cost-effectiveness would be used when data and methods permit. Because ineffective treatments waste money and expose patients to unnecessary risks, in this setting let us also imagine that there would be a requirement for good evidence before a new treatment is covered or an old treatment is affirmatively recommended. Old treatments that are not supported by good evidence would still be covered but would not be affirmatively recommended through guidelines or as a standard of care. In the other setting, which has an open-ended budget, costs would not enter medical decisions or determinations of standards of care. In this setting we can also imagine that the standards of evidence might be looser. A new treatment might be covered if it is *possibly* effective and/or it is the patient's last hope. An old treatment that is not supported by good evidence might be made the standard of care if it is "time honored" and there is a general consensus about its appropriate role.

Each of us might have a personal belief about which of these two settings is preferable, but our immediate need is not to force agreement on a single position, but to acknowledge that multiple settings can exist, and the standards of care can *appropriately* be different in the different settings. In fact, the existence of the two settings and the balance between them should ultimately be determined by people and patients—the ones who actually receive the treatments, live with their outcome, and pay their costs. People who are willing to pay any costs and want to receive investigational treatments (and are willing to pay the higher premiums to have them covered), will choose plans that operate with an open-ended budget, with a relatively loose standard of evidence. People who balk at paying higher costs and only want to pay for treatments that are known to provide benefit, will choose plans that operate under a limited budget, with a tight standard of evidence. Both settings and their respective standards of care are correct and ethical, provided they accurately reflect what people want and are willing to pay for.

This step of acknowledging the existence of more than one setting and different standards of care should greatly improve the ability of the courts to resolve our disputes. If experts respect the existence of different settings and different standards of care, and if they are careful to provide testimony that is appropriate to the settings in which the disputes arose, they can concentrate on the matters at hand and spare the courts confusing and conflicting testimonies that reflect differences in philosophies more than differences in the facts or evidence.

The second step is to acknowledge that there is no such thing as a community standard. We can still talk of a "standard of care" or "standards of care," with all that that implies about coverage and malpractice. What we need to drop are the assumptions that the majority of physicians are currently following a single standard of care, and that one can learn the standard of care by observing the practices of a community of physicians. This step by itself will be enormously useful in everyone's attempts to reduce inappropriate care and waste. Currently, many physicians claim to feel that a particular practice is inappropriate, and that they would personally prefer not to do it but are compelled to do it out of fear that they will be compared to a community standard. The result is that they all do it, which in turn makes it the community standard, which further entrenches the practice. This conceptual step will break that chain.

The third conceptual step is to reaffirm that ultimately truth is learned from empirical evidence. This may seem like a trivial step to take, given that the scientific revolution has been under way for more that a half a millennium, but there are still vestiges of the thought that truth is what great authorities say it is. A practical implication of this step is that we should all be humble, even skeptical, about our ability to form accurate beliefs subjectively. We should challenge ourselves and consciously hold back on locking in a belief until we have seen a systematic review of the evidence.

A fourth conceptual step is to acknowledge that the subjective beliefs of experts or a consensus of subjective beliefs cannot necessarily be taken at face value. This does not imply that all experts are wrong. Indeed, it is probable that most are right most of the time. This step only states that an expert or consensus of experts is not necessarily right, and there is no easy way to determine when they are and when they are not (without looking at the actual empirical evidence). It is also important to understand that this step does not mean we do not need experts. They will always be needed to survey and interpret the evidence. It is only to say that when experts offer conclusions about the standard of care, they need to describe the underlying evidence.

Practical Steps

These conceptual steps lead to several practical steps. The first has already been introduced: we need to be much more careful about anchoring beliefs and claims to empirical evidence. The ideas here were first described in the context of "evidence-based guidelines" (Eddy 1990) and

have subsequently blossomed into "evidence-based medicine"(Guyatt 1991; Sackett and Guyatt 1992; Sackett et al. 1997). For the development of guidelines and standards of care, "the evidence based approach explicitly describes the available evidence that pertains to a guideline and ties the guideline to evidence. . . . It consciously anchors a guideline, not to current practices or the beliefs of experts, but to experimental evidence. The usual question is whether the practice under consideration has been shown to be effective in improving the most important outcomes. Merely providing evidence as background material or peppering a guideline with occasional references to support particular positions do not count" (Eddy 1990: 1272). The most commonly cited definition of evidence-based medicine is less directive about the balance between subjective reasoning and empirical evidence but still emphasizes the role of evidence: "The conscientious, explicit and judicious use of current best evidence in making clinical decisions about the care of individual patients" (Sackett et al. 1996: 71).

The second practical step is for those who fund and conduct clinical research to improve the collection and interpretation of empirical evidence. The problem is not just the paucity of good evidence but the mixed quality of the evidence that does exist. Empirical evidence can be very difficult to collect and interpret. It is easy to be misled, especially if a misinterpretation coincides with one's prior beliefs and self interests. While a full description of experimental methods would obviously be inappropriate here, two things are especially important. One is the need for controls to minimize the effects of patient selection biases. The other is the need to track a treatment's effect all the way to health outcomes to avoid being misled by changes in biological outcomes. A high proportion of the evidence collected and reported by researchers, and cited by proponents of treatments, is at best worthless, and at worst misleading. This needs to be corrected.

None of the previous steps will be fully effective unless plans, purchasers, and patients communicate much more precisely about what a plan is expected to cover. The main vehicle for accomplishing this is the benefits language in contracts. Thus the third practical step is to improve the contracts that describe what plans will cover. This should involve the following:

- Include explicit descriptions of (1) the standard of evidence that will be used to determine coverage and to design guidelines, and (2) the role that costs and cost-effectiveness analysis will play in coverage determinations and guidelines.

- Describe any methods and processes that will be used to make specific determinations.
- Include examples to illustrate the principles, methods, and processes.
- List specific treatments that are likely to be controversial.
- Make the language as clear as possible. This includes using big print, paying attention to different language needs, discussing the coverage with people face to face, providing counselors, and avoiding any advertising or other public statements that may mislead people about the extent of coverage. The goal is to achieve truly informed consent that will not only create accurate expectations but that will hold up in court.
- Create an appeals process that is fair to patients but is consistent with the terms of the contract. The process will involve a review by external impartial experts, but the instructions to the experts will be different depending on the setting and the terms of the contract. For example, in a setting that is committed to practicing evidence-based medicine, the question to reviewers should be: What is the evidence and does it meet the standard described in the contract? In a setting that wants a looser approach to evidence, the question can be: In your opinion, is this treatment appropriate?

It is difficult to overemphasize the importance of this step. This is where plans can make clear the type of setting in which they are practicing, the role of costs, the approach to evidence, and the implications of all these for coverage and the standards of care. And this is where people will make their choices about the type of coverage and care they want to receive and the premiums they want to pay. If people prefer a different setting than is being offered by a particular plan, they can and should seek another plan that offers the setting, care, and costs they prefer. But once the plan has made its commitment in the contract, and once a person accepts that commitment, both parties should be held to it by the courts.

What Can the Courts Do to Help?

If we imagine for a minute that "the courts" is an entity that can respond to social needs and make changes, there are several things it can do to help support the conceptual and practical steps health care must take.

The first is to make a similar commitment to empirical evidence. Experts will continue to play the critical role of introducing evidence to the court, but it is the expert's role as interpreter of the evidence rather

than holder of a belief that is desired. When an expert expresses an opinion, the court can and should ask to see the evidence behind that opinion.

The court should not assume that the validity of an expert's testimony is correlated with the expert's training, affiliations, or other indirect measures, even experience. In fact, these characteristics may well correlate better with personal and professional biases than with validity. For example, the principle investigator of the world's largest research project to demonstrating the value of some new treatment arguably has the world's biggest incentive to have that treatment covered. Furthermore, credentials and experience in some medical specialty do not necessarily mean that an expert is trained in the formal methods of interpreting evidence, even in that specialty. For a reminder, experts in high-dose chemotherapy, including the principle investigators of some of the most prominent trials, testified vehemently that the treatment was effective, only to be proved wrong. To the greatest extent possible, the credibility of an expert should be based on the credibility of the evidence he or she presents.

The third thing courts can do is honor the contract. Specifically, to the extent that a plan's contract describes the standards of evidence it will apply, and the role of costs in coverage decisions and guidelines, the court should respect what the contract says and ensure that the case is evaluated within the standards and methods agreed to in the contract. If the contract specifies a strict requirement for evidence of effectiveness before covering a new treatment, the court should look for evidence that meets that standard. If the contract says that costs will be considered and cost-effectiveness will be used, the court should accept and apply the results of cost-effectiveness analyses. If the contract implies that a treatment will be considered noninvestigational and covered if the patient's personal physician believes it is in the patient's best interests, then that should be the test applied by the court. Similarly, patients should be expected to make reasonable attempts to read and abide by the contracts they sign. In *deMeurer v. HealthNet*, Mr. deMeurer admitted he "threw it in a pile with all the other papers" without reading it (Larson 1996). That should not invalidate the contract.

A fourth step is to strive for consistency and predictability. There is no way that health care can reduce variations in practice patterns and high rates of inappropriate care—two of the biggest contributors to bad quality and waste—without making decisions that will be controversial. (Notice that every instance of inappropriate care is initiated by someone who will argue it is appropriate.) In order for individual physicians, groups of physicians, or plans to make controversial decisions, they have

to be confident that they will be treated fairly if they end up in court. A big part of the burden of achieving consistency rests on the plan's contract; court decisions cannot be consistent until contracts are clear. But the court itself, even setting issues of ERISA aside, also have a responsibility for increasing consistency. Judgments that bewilder outsiders can destroy confidence that the courts can address difficult cases fairly.

The importance of this problem calls for some examples, all drawn from HDC/ABMT for breast cancer. In one case the court ruled in favor of the plaintiff on a finding that the following statement in the coverage policy was ambiguous: "Autologous bone marrow transplants of other forms of stem cell rescue (in which the patient is the donor) with high dose chemotherapy and irradiation are not covered" (Bailey v. Blue Cross & Blue Shield of Virginia, No. 94-2531 [4th Cir. Oct. 11, 1995]). In stark contrast, a different court found the following criterion for excluding coverage to be sufficiently precise to find for the defendant: A treatment is "experimental or of a research nature" if it is not "generally accepted medical practice" (Peruzzi v. Summa Medical Plan, No. 98-0069 [6th Cir. Feb. 27, 1998]). In an especially puzzling case, to avoid any possibility of bias, HealthNet asked a group of cancer experts from UCLA, UCSF, and Scripps to define for themselves the indications for high-dose chemotherapy for breast cancer. A case of breast cancer occurred that the experts agreed did not meet the indications they themselves had defined. HealthNet held the line, only to lose the case, have to pay $1.3 million in damages, be castigated by the court for "extreme and outrageous behavior [that] exceeded all bounds usually tolerated in a civilized society," and be featured on the front cover of *Time* (Larson 1996). Needless to say, we need more order than this if we are to successfully address the issues of evidence and cost-effectiveness.

A final recommendation is for the court to use court-appointed, neutral experts. It is difficult enough for impartial people who specialize in the interpretation of evidence to make sense out the existing research. It can be hopeless for people who have less experience with numbers. Reliance on competing experts in an adversarial mode is more likely to confuse and distort than clarify. Some of the legal issues raised by this "gatekeeper" model are discussed by Daniel W. Shuman in this issue.

Precedents and Prospects

This is a long list of difficult recommendations. Some of the precedents are hopeful; some are discouraging; and some are blatantly naive. One

of the encouraging areas is the potential of the court to press experts for the empirical evidence that supports their beliefs. Certainly there are many precedents for demanding to know the empirical evidence behind an expert's opinion. A forensic expert who specializes in footprints is not just asked: "Was the accused at the scene of the crime?" He or she is grilled on the evidence behind any answer to a question like that. In health care cases, the courts appear to be more deferential to medical experts and require less explanation or empirical evidence. But the precedent to ask much more is there.

The precedents for applying or accepting cost-effectiveness analysis (CEA) are less encouraging to me. As discussed by Peter D. Jacobson and Matthew L. Kanna elsewhere in this issue, the use of CEA in other types of cases is mixed. In some cases an appeal to cost-effectiveness provides a successful defense, in others it kills a defendant's case. If there is a pattern at all, it appears to be that cost-effectiveness analysis is used when it can support the little guy against the big guy. To the extent that this is true, it is not very encouraging to health plans. From the perspective of health care, the most promising case law is the reasoning of Judge Learned Hand in *United States v. Carroll Towing* (159 F.2d 169 [2d Cir. 1947]); there is negligence when the cost of preventing an accident is less than the probability of the accident multiplied by the gravity of the resulting injury. Judge Hand's formula can be adapted nicely to the types of problems seen in health care. But there appears to be considerable variability in the extent to which Judge Hand's formula is actually used, and its transferability to health care is unclear.

The most encouraging new development is the U.S. Supreme Court's recent unanimous ruling in *Pegram v. Herdrich* (120 S. Ct. 2143 [2000]). Herdrich, a member of the physician-owned Carle Care HMO, had an inflamed abdominal mass. The physician, Pegram, determined that it did not constitute an emergency and instead of ordering an immediate ultrasound at a nearby hospital, scheduled an ultrasound at a Carle Care hospital eight days later. In the meantime, Hendrich's appendix ruptured, requiring emergency surgery. Because Carle Care rewards its physician owners with a year-end bonus for controlling costs, Herdrich argued that the financial incentives compromised the medical care she received and constituted an inherent breach of an ERISA fiduciary duty. The case is a good example of the type of decision that has to be made in a budget-limited setting. A physician chose a course of treatment that has a slightly higher risk, because the magnitude of the risk was judged to be too small

to justify the cost. It has the added feature that the physician making the decision would share in the savings.

The Court found for the physician. The immediate reason was that the plaintiff had failed to create a claim under ERISA. But Justice David Souter's opinion for the Court went out of its way to include a policy discussion of the use of rationing and cost-cutting incentives by HMOs. His opinion made it clear that it is appropriate for HMOs to ration care, control costs, and even use physician financial incentives to accomplish these goals. "Imposing federal liability for efforts to reduce costs—the entire purpose of managed care—could destroy HMOs altogether." He continued: "No HMO organization could survive without some incentive connecting physician reward with treatment rationing. . . . Inducement to ration care goes to the very point of any HMO scheme, and rationing necessarily raises some risks while reducing others." The Court did not specifically address the use of CEA, but the inclusion of cost-effectiveness in benefit contracts and the use of formal methods to determine cost-effectiveness should find support in Pegram.

At this point it is worth observing that the methods of CEA are imperfect and easily abused (see Drummond Rennie's essay in this issue), and often the necessary data are unavailable or untrustworthy. But before we reject such analyses, we need to note that virtually every other aspect of medical decision making is also imperfect and suffers from incomplete data. Some examples from this discussion are the use of experts by courts, benefits language in contracts, references to community standards, clinical research, and the estimation of a treatment's outcomes. Furthermore, the options that remain if one does not use formal CEA, such as ignoring costs, or trying to guess a treatment's cost-effectiveness, are imperfect. At least this provides a formal, accountable method for addressing costs that is at least as good as any other approach. Limitations of the methods and data should not cause us to discard it altogether but to limit its use to cases the methods and available data permit.

One Final Thought

Considering all the problems, the necessary steps for health care, the help needed from the courts, and the prospects for accomplishing it all, it is easy to be discouraged. Indeed, we can expect turmoil for many years to come. This leads to a final thought. The best way for the health care system to interact with the courts is to stay out of them. We in health

care should do everything we can to prevent misunderstandings and conflicts, and to address them outside the courts through such things as counseling, mediation, and binding arbitration. In the end, the most important lessons are to be clear in our thinking, to be precise in our communications, to be fair and consistent in our applications, and to be respectful and helpful when disagreements do arise.

References

Eddy, D. M. 1990. Practice Policies: Where Do They Come From? *Journal of the American Medical Association* 263:1265–1275.

Guyatt, G. H. 1991. Evidence-Based Medicine. ACP Journal Club (*Annals of Internal Medicine* 114 [suppl. 2]): A–16.

Larson, Erik. 1996. The Soul of an HMO. *Time* Jan. 22. Available on-line at www.time.com/time/magazine/archive/1996/dom/960122/cover.html.

Sackett, D., and G. H. Guyatt. 1992. Evidence-Based Medicine: A New Approach to the Teaching of Medicine. *Journal of the American Medical Association* 268:2420–2425.

Sackett, D. L., W. S. Richardson, W. Rosenberg, and R. B. Haynes, eds. 1997. *Evidence-Based Medicine: How to Practice and Teach EBM*. New York: Churchill Livingston.

Sackett, D. L., W. Rosenberg, J. A. Muir Gray, R. B. Haynes, and W. S. Richardson. 1996. Evidence-Based Medicine: What It Is and What It Isn't. *British Medical Journal* 312:71–72.