

WHY EVIDENCE-BASED MEDICINE ONLY SEEMS SIMPLE:

The evidence isn't always stellar // Guidelines don't account for patient differences // Experts often disagree about the facts.

# Burden of Proof

■ BY ANITA SLOMSKI // PHOTOGRAPHS BY NIGEL COX



Last November the U.S. Preventive Services Task Force unleashed a maelstrom when it recommended against routine mammography for women younger than age 50 and advised physicians not to teach women how to do breast self-exams. The response was swift and fierce, with such celebrities as Jaclyn Smith and Sheryl Crow—both breast cancer survivors—leading the vanguard in proclaiming the new recommendations to be misguided and dangerous.

Physicians lined up right behind them. “This is a giant step backward and a terrible mistake,” oncologist Marisa Weiss told the *New York Times*. “We know mammography overperforms and finds things that will never be life-threatening, and we know it underperforms,” said Weiss, who is the founder of Breastcancer.org. “But it has no chance to perform in women who don’t get it.” Added Christine Hodyl, director of breast surgery at South Nassau Communities Hospital in Oceanside, N.Y., “I can’t tell you how many times we pick up a cancer in a young woman who comes in for a baseline mammogram before age 40.”

But the USPSTF’s recommendations were based on the best research and decision-making that medicine has to offer. A panel of primary care physicians and scientists—not oncologists or radiologists, who might have a financial stake in breast cancer treatment or prevention—considered the results of numerous randomized trials and commissioned its own study before concluding that 1,904 women in their forties would

have to be screened with mammography to prevent just one death from breast cancer. That benefit did not outweigh the risks of false-positive findings, which would subject women to unnecessary biopsies, anxiety and treatment of benign breast conditions, according to the USPSTF, which urged women to discuss the benefits and tradeoffs with their doctors. The panel also found no evidence to support the utility of breast self-exams, citing data indicating that most lumps women find are benign and don’t merit the biopsies that often follow.

The new mammography recommendations joined a growing body of research that falls under the rubric of evidence-based medicine. This approach is based on the belief that medical care will be better, safer and more efficient if physicians base clinical judgments on solid empirical science. Backed by such evidence, government agencies, specialty medical societies, disease associations and large health plans have developed guidelines that set out exactly which drugs, tests and treatments doctors should use to manage a vast array of medical conditions. Insurers follow evidence-based guidelines to make decisions about which interventions they will cover. And Medicare and private health plans have utilized guidelines to extrapolate hallmarks of quality and hold physicians and hospitals accountable for meeting them.

Instead of receiving gratitude for improving care, however, the USPSTF found itself at the center of controversy. “Breast cancer is a topic of great emotional and physical meaning for women, and the guidelines were created by a group whose name suggested it was a governmental body with an interest in

**EVID**

**ENCE**

Agency \_\_\_\_\_

NO. \_\_\_\_\_

Description \_\_\_\_\_

Location of Collection New York, New York

Type of Offense \_\_\_\_\_

Victim \_\_\_\_\_

Suspect \_\_\_\_\_

**CHAIN OF CUSTODY**

Received From	<u>TS</u>	Time	<u>MS</u>
Date	<u>4-16-04</u>	Time	<u>12:01 AM</u>
Received From	<u>K.P.</u>	By	<u>K.P.</u>
Date	<u>4-03-04</u>	Time	
Received From		By	
Date		Time	

**EVIDENCE**

controlling costs,” says Steven Pearson, a physician and president of the Institute for Clinical and Economic Review. Based at Massachusetts General Hospital’s Institute for Technology Assessment and affiliated with Harvard Medical School, ICER is a policy group that weighs the effectiveness and the cost of medical treatments. In fact, says Pearson, the USPSTF is an independent organization. “The announcement explaining why the recommendations had changed was very badly managed,” he explains. “And the battles over health care reform going on at the same time just turned up the volume.”

The furor illustrates one of the biggest problems with evidence-based medicine—that it can be almost impossible to produce the kind of rock-solid evidence that will convince physicians (and the public) that a particular intervention, preventive measure or diagnostic test really is the best medicine. Though the USPSTF recommendations were never meant to apply to every patient, nuance was quickly lost in the uproar. Those who objected to the organization’s conclusions simply found medical evidence they liked better: the American Cancer Society’s guidelines calling for women to receive yearly mammograms once they turn 40. “It can be challenging to make medical decisions when august groups of clinical experts look at the same evidence and, based on their beliefs about its strength, come to very different conclusions,” Pearson says.

Yet there are compelling medical—and financial—reasons to persist in attempts to rationalize how medicine is practiced. The Institute of Medicine has estimated that only half the treatments that doctors prescribe are effective, and according to an analysis by the New England Healthcare Institute, a health policy group, \$760 billion is wasted each year in the United States because of unnecessary diagnostic tests and medical procedures, medical errors and the need to treat hospital-acquired infections. These problems might be reduced through effective practice guidelines. Many politicians and policy experts are banking on evidence-based medicine to

“Physicians do a great job of making decisions when they have compelling data to work with, but in many cases the data have been missing.”

improve the quality of U.S. medical care and rein in its cost, which now stands at an unsustainable 17.3% of gross domestic product. “The last thing you want to do is turn away from the evidence in the name of physician intuition,” says Alan Garber, director of the Center for Health Policy at Stanford University. “Physicians do a great job of making decisions when they have compelling data to work with, but in many cases the data have been missing.”

When the phrase “evidence-based approach” first appeared in the medical literature in 1990, describing the need to anchor medical decisions to scientific evidence, it reinforced what some health policy experts had been saying for decades—that physicians’ clinical judgment was based on too much art and not enough science. In 1973 John Wennberg from Dartmouth Medical School began mapping how practice patterns varied geographically, and he ultimately concluded that tradition, physician preferences and the numbers of physicians and hospital beds per capita were behind many of the differences. What’s more, subsequent studies suggested that patients often got the wrong care. One study, for example, found that for a third of patients who had carotid arteries unblocked surgically, the procedure’s risks outweighed the benefits.

Among the first evidence-based guidelines came from physician and mathematician David Eddy, who in 1980 was a professor at Stanford University’s School of Engineering. Eddy says he was motivated by the sheer complexity of medical decision-making. His initial guideline, for the American Cancer Society, held that women should be screened for cervical cancer every three years rather than annually—a recommendation that didn’t become standard practice for 20 years. Later, Eddy, as a consultant for Blue Cross/Blue Shield, urged the insurer to use evidence—or the lack of it—as a criterion for making coverage decisions. When Blue Cross, under his guidance, refused to pay for a breast cancer treatment that combined high-dose chemotherapy with a bone marrow transplant, Eddy got hate mail from physicians. A subsequent study confirmed that the therapy did not extend patients’ lives.

Soon medical societies got in on the act, not only consulting the available evidence to determine when people should

be screened for various diseases but also coming up with guidelines that sought to show the best approaches for treating everything from hemorrhoids to various cancers. Government health agencies and private insurers took things one step further, turning evidence-based guidelines into quality standards for doctors and hospitals. More than 100 “pay for performance” programs now tie reimbursement to such yardsticks as how many patients receive recommended blood pressure or diabetes monitoring. In Massachusetts some insurers assign physicians to one of three tiers according to their relative performance on cost and/or quality measures—and increase out-of-pocket payments for patients who choose to see the lowest-scoring doctors (patients are made aware that they are paying higher fees).

Supporting the push to justify how medicine is practiced, Congress appropriated \$1.1 billion in 2009 for comparative-effectiveness research, commonly known as CER, which evaluates the strengths and weaknesses of treatment options for particular medical conditions. Though evidence-based guidelines and effectiveness research aren't supposed to factor in cost, that issue is never far from view. One hope among proponents is that physicians will be swayed by the coming government-funded research to choose proven, older treatments with the same outcomes rather than automatically prescribing the latest, most expensive technology—and that insurers will integrate comparative effectiveness into coverage decisions.

**F**rom the start, however, the quest to produce evidence-based guidelines covering every medical treatment has been hampered by a persistent question: Just how good is the evidence? In the hierarchy of medical knowledge, the results of randomized, controlled clinical trials come out on top. There is no stronger scientific proof of whether a treatment, drug or medical device is effective than a trial that compares what happens to people who receive a new intervention with outcomes for those who had a different treatment or a placebo. Even with this kind of research, there may be concerns about the design of a study or how its data is interpreted, but evidence from a randomized trial has the greatest power to shape doctors' beliefs and practices—and to be turned into a quality benchmark in pay-for-performance programs. “You can



create medical standards only if there is very good evidence that this is the best way to treat a patient,” Eddy says. Yet at a cost of \$10,000 to \$50,000 per patient per year, controlled clinical trials are multimillion-dollar propositions, and the guidelines based on them are in a distinct minority.

More frequently, standards are “consensus-based”—developed by panels of disease specialists or other physicians who rely on their own expertise and studies that don't have the scientific rigor of large, controlled trials. These guidelines constitute recommended care, and physicians can follow them or not as they choose. “Some of these guidelines are self-serving and intended to ensure referrals or reimbursement for certain services, but others are really done for the right reasons”—to help physicians deliver the best care, says Richard A. Deyo, professor of evidence-based medicine at Oregon Health and Science University and author of *Hope or Hype: The Obsession With Medical Advances and the High Cost of False Promises*.

But even guidelines that seem to have been done in the right spirit may be viewed suspiciously by physicians worried about the behind-the-scenes influence of pharmaceutical companies or medical device manufacturers. Those companies often pay

specialists to extol new technologies or drugs to other physicians. “The influence of the drug and device industries has become ever stronger, so you may wonder whether you’re looking at accurate information when you read a guideline,” Deyo says.

To counter such criticism, many specialty societies have begun disclosing conflicts of interest and are trying to make sure their guideline panels include diverse points of view. But Rodney Hayward, director of the Center for Practice Management and Outcomes Research at the Ann Arbor Veterans Administration and professor of public health and internal medicine at the University of Michigan, would prefer to take this job away from specialists—partly because he considers it unrealistic to expect doctors with a stake in the outcome to make unbiased decisions.

“Guidelines are being made by people who couldn’t pass my epidemiology course—they wouldn’t even come close,” Hayward says. “That means they interpret the evidence in very simplistic ways and treat patients as though they were interchangeable.” The result is rigid guidelines that leave no room for physicians to give nuanced care to individual patients, says Hayward, who notes that clinical trials report the average benefit of an intervention for an average patient. But the reality is that a drug or a treatment does not affect every patient equally.

That’s why Hayward opposed a guideline that advocated using multiple drugs to keep a diabetic’s glycosylated hemoglobin level—as gauged by the A1C test, which measures blood glucose levels over a period of months—below 7%. There’s plenty of evidence to suggest the drug metformin lowers a diabetic’s risk of heart attack, and it may slow progression of diabetes as well, Hayward says. But he thinks that adding other drugs solely to reduce a moderately elevated A1C may lead to inappropriate care. Those additional treatments could cause significant side effects, and they may not have been tested for long-term safety when used in combination with metformin.

What’s more, a hard-and-fast A1C guideline doesn’t take into account the circumstances of a particular patient. A 45-year-old, for example, might be willing to suffer the side effects of multiple drugs if the drugs reduced his risk of becoming blind in 20 years. But the same tradeoff might look different to a 65-year-old, says Hayward. “We shouldn’t use medications in low-risk individuals for small amounts of gain,” he says. The guideline was modified in 2008 after a trial of aggressive glucose-control treatment was halted because of excessive numbers of deaths among elderly diabetics.

Hayward’s choice for creating guidelines would be a group of

internists and family physicians trained to interpret trial data so that it would more accurately reflect the benefits to patients with wide-ranging disease risk profiles. “Generalists who understand that treatment benefits and safety vary tremendously—and that optimal care includes considering which compromises patients are willing to make—would produce guidelines that physicians and patients could actually follow,” he explains.

But even flexible guidelines that accommodate a patient’s circumstances and preferences might fall short if they cover only a single disease or condition. Many patients have diverse medical problems, and treating one effectively can sometimes aggravate another, says Cynthia Boyd, assistant professor at the Johns Hopkins University School of Medicine. When she applied the best evidence-based guidelines she could find to



Who should—and shouldn’t—create evidence-based guidelines?  
Share your take at [protoeditor@mgh.harvard.edu](mailto:protoeditor@mgh.harvard.edu).

## Could Electronic Trials Provide Better Evidence? //

A mathematical model can expand on clinical results quickly and cheaply. But physicians may be skeptical.

Just as he evangelized about the need for physicians to base medical decisions on empirical evidence, physician and mathematician David Eddy is now promoting a mathematical model to fill the evidence gaps of the movement he helped start 30 years ago. The model, which he calls Archimedes, is built from hundreds of equations that represent human physiology and the effects of disease, and is overlaid with hundreds more simulating tests, treatments and physician behaviors. Archimedes expands upon clinical trial data to answer questions that trials can't, Eddy

says. For instance, it can project how changing the dose of a drug or combining it with other medications may affect patient health. It can also estimate how a broader population than the one studied might benefit from a new intervention.

What's more, Archimedes can be much faster and cheaper than a clinical trial. In two months, at a cost of several hundred thousand dollars, Eddy was able to replicate a seven-year, multimillion-dollar controlled clinical trial studying whether the drug atorvastatin could prevent heart attacks and strokes in diabetics—

and Archimedes was almost spot-on in predicting the trial results before they were announced.

The model can't replace clinical trials, however. "First you need empirical evidence from a trial that a test or treatment is effective," Eddy says. Only then, after Eddy has developed an accurate model based on those results, can Archimedes apply it to untested applications.

But if physicians find fault with guidelines created from actual trials, why would they trust a model? "All we can do is show that the model can reproduce what is seen in reality," says Eddy.

a hypothetical 79-year-old woman with five common chronic diseases—osteoporosis, osteoarthritis, type 2 diabetes, hypertension and chronic obstructive pulmonary disease—Boyd discovered that the treatment recommendations were unrealistic and potentially harmful, in part because the guidelines included limited information about potential drug interactions. One guideline called for treating osteoarthritis with nonsteroidal anti-inflammatory drugs, which have side effects that include heightened blood pressure and impaired kidney function—not a good choice for a hypertensive patient. Following the separate guidelines to the letter also meant that her fictional patient would have had to take 12 medications in 19 combinations 5 times a day. "Guidelines have an important role, but they have been developed from a single-disease perspective," Boyd says.

Boyd also worries that physicians who receive financial bonuses for adhering to guidelines may not be motivated to develop customized treatment plans for patients with complex problems. "If your income is tied to doing everything the guidelines say, you may be less inclined to talk to patients about what health outcomes they value most, think through the drug interactions and figure out the right treatment," she says. "Even when there isn't a financial incentive, most physicians want good grades on their performance report cards."

While evidence-based medicine in its purest form is blind to the cost of interventions, many of its champions believe cost should be part of evidence-based policy decisions. "Many countries apply cost-effectiveness analysis when they develop clinical guidelines in part because they find it useful for identifying patients who will receive the greatest benefit from a health intervention," says Stanford's Garber. But in the United States, where the cost of care often isn't factored in, expensive drugs and procedures

are more likely to be used and covered by insurers regardless of how little benefit they might provide. "We won't have an efficient health care system until we learn the value of individual interventions," Garber says.

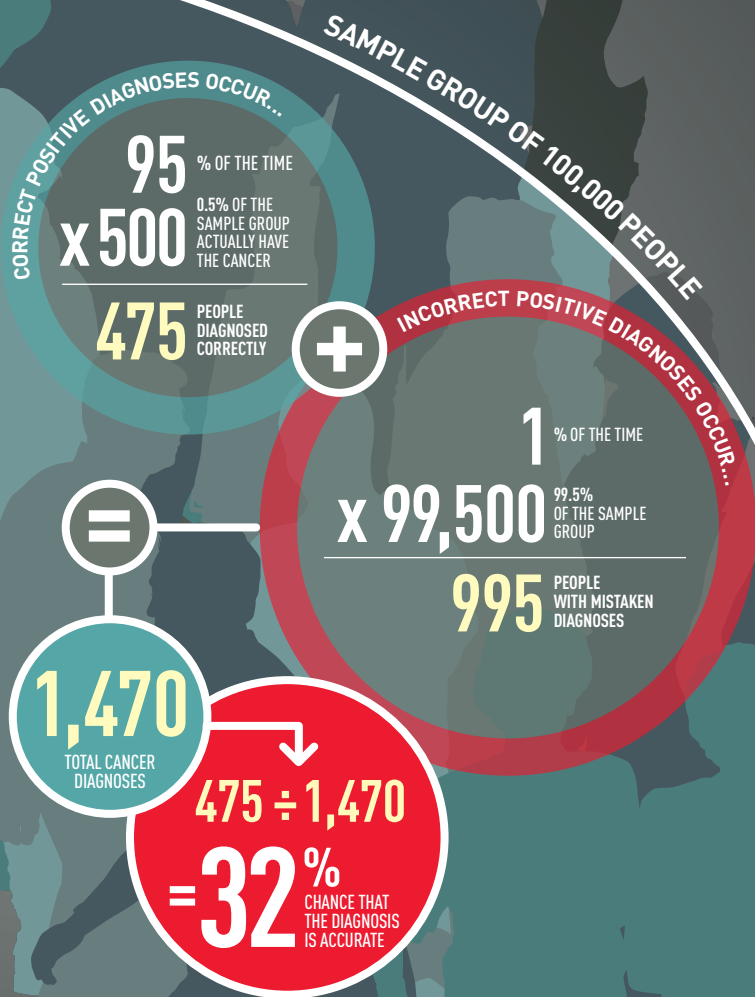
A process known as cost-effectiveness analysis uses clinical outcomes data in mathematical models to determine how many extra years of healthy life a person receives for every dollar spent on a particular treatment or procedure. That determination, in which improvements are measured in quality-adjusted life years, or QALYs, varies to a "stunning" degree across the range of medical procedures, says Milton Weinstein, professor of health policy and management at the Harvard School of Public Health, who helped pioneer cost-effectiveness analysis almost 40 years ago. According to research by his colleagues Jane Kim and Sue Goldie, annual screening for cervical cancer detects only a few more cases of cancer than tests done every other year. Doing the more frequent screening gains mere days or even hours in average life expectancy—and costs about \$800,000 more for every year of life gained.

Alternative treatment protocols for some diseases, such as using trastuzumab in addition to standard chemotherapy to treat metastatic breast cancer, may have both a lower price and a better payoff—in this case, each gain of a QALY (which already factors in chemotherapy) costs between \$125,000 and \$150,000. But it would be very hard economically to justify routinely performing a CT scan for lung cancer in 60-year-old former heavy smokers, at a cost of \$2.3 million per QALY, compared with no screening.

Some countries routinely weigh and act on such comparisons. In the United Kingdom, the National Institute for Health and Clinical Excellence evaluates cost-effectiveness when recommending which new medical technologies the country's national health service should cover. Treatments

## Screening's Strange Math //

Organizations formulating guidelines to ensure that tests are doing more good than harm must consider how often tests produce false-positive results. The U.S. Preventive Services Task Force, for example, has concluded that almost 2,000 women in their forties would have to be screened to prevent one death from breast cancer, and this could produce as many as 200 false positives, sometimes leading to unnecessary biopsies and even, in some cases, unnecessary surgery, chemotherapy or radiation. This hypothetical example of a test, put forth by mathematician John Allen Paulos, illustrates the type of arithmetic involved.



Even when the evidence shows that annual testing could confer more harm than benefit, as with women and annual mammograms, we still want to believe, says Paulos, that more frequent screenings will protect us from dying. A high rate of false-positive test results (68%, in this hypothetical case) simply doesn't carry as much psychological weight as the number of people we know who have had cancer. Nor is it easy to abandon an entrenched belief that annual mammograms are necessary.

that cost more than £30,000 (about \$46,000) per QALY are generally denied. (Cancer drugs that prolong lives by just a few months were unlikely to make the cut, until NICE recently enacted a compassionate care exception.)

Although the United States may be a long way from adopting a NICE model, some U.S. panels are starting to weigh costs. The result, in some cases, is to recommend more care, not less. A revised guideline for HIV/AIDS treatment, for example, recommends that all newly diagnosed patients receive "genotypic resistance" tests when they begin antiretroviral therapy. "Our analysis showed that the cost-effectiveness of routine genotypic testing is excellent," Weinstein says.

But such determinations aren't automatically linked to payment decisions. Historically, Medicare, which decides which treatments are covered by government health programs, has not considered the cost of medical services. Private health plans also tend not to factor in costs in deciding what to cover. Those insurers can, however, decide they won't pay for treatments or diagnostic tests that are more expensive but only slightly more effective than an alternative.

One health care payer, though, is only too happy to admit to using costs in deciding what to cover—and to invite the public to scrutinize its deliberations. In 2006 the legislature in Washington State approved a law that permits considering costs in making coverage decisions for about a million state employees, prisoners, and recipients of Medicaid or workers' compensation. A panel of physicians makes choices about surgical devices and procedures, treatments and diagnostic tests based on their efficacy, safety and cost-effectiveness. Anyone can suggest a treatment for the panel to review, all meetings are public, and the panel's decisions are posted for comment before being finalized. Now in its fourth year, the panel has made 15 rulings, and 10 more are pending.

The panel can decide to flat-out deny coverage for a procedure, as it did when it deemed an implantable drug pump for chronic noncancer pain to be unsafe. A second option is to rule that an intervention is necessary only in some cases. The panel has approved artificial lumbar and cervical disks for people younger than 60 and drug-eluting stents for those considered at high risk of having their coronary arteries narrow again after angioplasty. On rare occasions, the panel has even opted to expand coverage, as when it decided to pay for bariatric surgery for young obese people.

For help finding the best science on which to base its decisions, Washington State relies on a handful of organizations that

“We can create evidence-based guidelines, but we also have marketing campaigns that appear on the nightly news that contradict those guidelines.”

attempt to quantify the clinical and economic value of various treatments. Their research blends evidence-based medicine with cost-effectiveness analysis. One of these organizations, the MGH’s ICER, accepts funding from government agencies, health plans, and pharmaceutical and device manufacturers. Still, says ICER president Pearson, the institute maintains its impartiality by directing its own research agenda and putting its results in the public domain. “Our work is most useful when there is conflict about the strength of particular evidence,” he says. Health plans use ICER appraisals to make coverage decisions; medical societies employ them in creating guidelines.

Only government groups can commission studies from ICER, and the state of Washington asked for two. One was for virtual colonoscopy, a noninvasive imaging technology that offers an alternative to traditional colonoscopy. After evaluating studies that have looked at the new procedure, ICER concluded that virtual colonoscopy was as effective as, but likely to be more expensive than, a traditional colonoscopy. Moreover, ICER did not find that the technology would help persuade people to undergo colorectal cancer screening because they still must go through unpleasant colon cleansing. Based on ICER’s evaluation, Washington State decided not to cover the new imaging procedure.

More complicated was ICER’s appraisal of the benefits of cardiac CT angiography, a test that employs scanning technology instead of invasive angiography to diagnose coronary artery disease. ICER concluded that the new imaging test and the traditional approach worked equally well. But the more expensive CT angiography was cost-effective only in the emergency room, where it could quickly identify those who were not in imminent danger of a heart attack and could safely be sent home. The state of Washington decided to pay for CT angiography in the ER but not in outpatient settings.

So far, evidence-based coverage decisions have saved Washington State about \$27 million a year. That’s a modest beginning,

given that the state’s most recent budget allocated \$2.9 billion for health care. But it’s not about to abandon the experiment, and the national push toward using evidence-based medicine and cost-effectiveness analysis is also certain to continue. Yet, as the mammography controversy suggests, new clinical guidelines—particularly when they recommend less care and may be suspected of economic motivations—are likely to face skepticism from the public as well as from physicians.

“Today it’s no longer tenable to say that evidence-based medicine is irrelevant,” Garber says. “But it’s certainly possible to argue that current efforts promise more than they can deliver.” Those attempts have been hampered by everything from a lack of good evidence to questions about whether guidelines are corrupted by specialty biases and conflicts of interest. The issues of addressing patient differences and patients with multiple diseases further complicate progress.

There’s also a long tradition of physician and patient autonomy in medical decision-making in the United States—one of only two countries (New Zealand is the other) that allow direct medical advertising to consumers. “We can create evidence-based guidelines, but we also have marketing campaigns that appear on the nightly news that contradict those guidelines,” Oregon’s Deyo says.

Still, the stakes are too high to abandon the effort, Pearson says. “There will always be natural tensions and very strong interests in our health care system that will make interpretations of evidence challenging,” he says. “But we shouldn’t throw up our hands. Evidence-based medicine remains an important reason for optimism that we can improve our system, although there is a lot of work ahead.” ■

## → DOSSIER

1. “Where Is the Wisdom? A Conceptual History of Evidence-Based Medicine,” by Peter C. Wyer and Suzana A. Silva, *Journal of Evaluation in Clinical Practice*, September 2009. Tracing evidence-based medicine’s origins in the 1970s through its “crowning achievements” during the information age, the authors argue that EBM does not threaten individualized care, as some believe, but rather informs it.
2. “Comparative Effectiveness and Health Care Spending—Implications for Reform,” by Milton C. Weinstein and Jonathan A. Skinner, *The New England Journal of Medicine*, Jan. 6, 2010. Comparative-effectiveness research alone won’t cut health care costs, these authors say. Physicians and hospitals must make hard decisions about where limited resources will do the most good.
3. “Imaging Idolatry: The Uneasy Intersection of Patient Satisfaction, Quality of Care, and Overuse,” by Richard A. Deyo, *Archives of Internal Medicine*, May 25, 2009. Deyo explores the conundrum physicians face when evidence-based guidelines clearly indicate one approach to treating back pain but patients demand another.