RALPH I. HORWITZ, MD
Dr. Horwitz is Chairman, Department of Internal Medicine, and Professor, Medicine and Epidemiology, Yale University School of Medicine.

The dark side of evidence-based medicine

This commentary is based on a lecture Dr. Horwitz presented at the Cleveland Clinic Division of Medicine Grand Rounds.

Evidence-based medicine, or the "outcomes movement," accepts as axiomatic that a substantial portion of health care expenditure in the United States is wasted on unproven or ineffective tests and treatment. As a result, this movement figures prominently in health care reforms and in medical education. Its aspirations are not modest. In a 1992 article the Evidence-Based Medicine Working Group[1] heralded the arrival of this discipline, arguing that in patient care, evidence from clinical research should supplant intuition, clinical experience, and physiologic rationale. Many institutions, including my own, have since jumped on the outcomes movement bandwagon, developing guidelines for the care of patients.

The unspoken conviction of the outcomes movement is that impersonal knowledge of the probability of an event is the principal precondition for effective clinical medicine and superior to the more traditional physician learning-by-experience and the pathophysiologic reasoning. That assertion causes concern among physicians trying to balance the evidence-based approach with the more traditional methods.

Despite the intellectual arguments in favor of the outcomes movement, it is important to examine our intuitive reactions to the practical aspects of this approach. Imagine you were a 38-year-old woman with anemia and thrombocytopenia. Would you prefer to be treated by a naive clinician guided by a set of pathways or by an expert physician with experience treating hundreds of patients with anemia? Most of us would prefer the expertise that comes from experience guided by evidence, rather than by guidelines and evidence alone. Therefore, it is important to get beyond the evidence-based medicine hype to identify the limits of evidence and the importance of the "art" of medicine.

EVIDENCE DOES NOT VANQUISH UNCERTAINTY

The evidence for the utility of many diagnostic and therapeutic interventions of modern medicine is incredibly powerful, but there are large areas in which the evidence to guide clinical decisions is incomplete, contradictory, or inconsistently interpreted. One has only to look at studies that examined the appropriateness of clinical procedures to see large areas of medicine where evidence is insufficient to guide clinical decision.

One study of upper GI endoscopy performed on 1585 Medicare patients in three states found inappropriate use of endoscopy in about 17% of subjects.[2] But even more intriguing, 11% of cases were uncertain when examined by an expert panel of physicians using a clear set of criteria. Similarly, in a study of hysterectomy, 16% of the procedures were found to be inappropriate, but 25% were uncertain.[3]

Most physicians understand there are many clinical circumstances in which they cannot be certain about the appropriateness of indications for procedures and for the other

clinical decisions they must make. Yet these uncertainties do not paralyze physician decision-making.

REPORTING RELATIVE VS ABSOLUTE RISK REDUCTION

In addition to the continual presence of uncertainty, it is important to consider the limitations of the evidence in evidence-based medicine. Consider the way the results of studies are presented.

The utility of absolute vs relative risk reduction

In a randomized trial in 525 men with nonvalvular atrial fibrillation, for patients receiving placebo the risk of stroke at 1 year was 4.3% and for those receiving warfarin, 0.9%.[4] The investigators expressed the results as a relative proportional risk reduction of 79%, calculated as the difference in these two rates divided by the placebo rate. This is the most common way results of randomized controlled trials are summarized and reported. But this relative risk reduction of 79% is much different than the absolute risk reduction of 3.4%, which is the difference between the actual rates in the placebo and the warfarin treatment groups, ie, 4.3% vs 0.9%.

Absolute risk difference is much closer to the information clinicians need to make decisions about the care of their patients. Also, the inverse of the absolute risk difference, the "number-needed-to-treat," has enormous utility in clinical decision-making. In the warfarin study, the inverse of the absolute risk difference is approximately 30. Therefore, one would need to treat nearly 30 patients with warfarin to prevent one death or thromboembolic stroke.

Why then, do investigators present information as proportional risk reduction? They do this because the proportional risk reduction is much more impressive than absolute risk reduction. Thus, by using the more impressive number, researchers can go on the *Today Show* or issue a press release that touts a dramatic result.

However, clinicians must be cautious about this. A series of studies has shown that how results are reported affects physicians' treatment decisions.[5–7] Physicians are far more likely to use a treatment when the results are presented as proportional reductions in risk than when the results are presented as absolute risk differences.

The art of medicine includes interpreting data, not simply applying them. Physicians need to be alert to the way information is presented so that they can make judgments that enhance the outcomes of therapy for patients.

THE PERILS OF SUBGROUP ANALYSIS

Suppose a physician is treating a 58-year-old patient, not one who is 67 years old, the average age in the warfarin study. And suppose the patient is a woman, not a man (the study, which was organized through a Veterans Administration Cooperative Study Center, included only men.) On the basis of the evidence, what treatment is indicated?

A physician who asks these questions is seeking a series of subgroup analyses. Clinicians routinely ask for subgroup analyses to determine the effectiveness of treatment for specific patients. Statisticians and researchers just as routinely refuse to perform or report such analyses. When they do report them, they place serious caveats on their interpretation because subgroup analyses run the risk of having positive findings that occur by chance rather than by actual biologic effect.

An example of a study in which subgroup analyses were very important was the Second International Study of Infarct Survival (ISIS 2), which included 17 187 patients with suspected MI who were randomly assigned to receive either routine care alone, aspirin alone, streptokinase alone, or aspirin and streptokinase.[8] Among 8587 subjects who received aspirin alone, the 1-month mortality rate was 9.4%. For the 8600 subjects who received no aspirin, the mortality rate was 11.8%. The relative risk reduction was 20%, and the absolute risk reduction was 2.4%. But clinicians ask, Does treatment with aspirin benefit specific classes of patients? That was a reasonable question, but the ISIS 2 statistical coordinators were very reluctant to provide those subgroup analyses. To illustrate why, they published an analysis that is celebrated as a clear example of why subgroup analyses should not be done in large-scale randomized trials.

The absurdity of subgroup analyses

The ISIS 2 researchers analyzed the data according to the patients' zodiac sign. Subgroup analysis showed that for patients born under Libra or Gemini, the number of deaths at 1 month were approximately equal in the aspirin and placebo groups, but all

other astrological signs had a substantial difference between the two groups.[8] This led the investigators to facetiously wonder whether the aspirin should be withheld from those born under the astrological signs Libra and Gemini and given only to patients born under other birth signs. This clearly made the point that some differences in subgroup analyses are found by chance alone.

The necessity of subgroup analyses

Conversely, one of the most powerful observations in clinical medicine came from a subgroup analysis of a trial that found negative results. The VA Cooperative Study of Surgery for Coronary Artery Occlusive Disease found no overall difference in the mortality rate between patients with stable angina pectoris treated with medicine or surgery.[9] Yet one of the most important observations of that trial, in fact of any surgery trial, was in a subgroup of 113 patients out of the total of 1015 patients.[10] In that analysis, patients who had significant lesions of the left main coronary artery and were treated with surgery had much lower mortality rates than patients treated with medical therapy alone. This difference, although substantial, was not statistically significant. Still, this subgroup analysis formed the basis for a treatment strategy of surgery over medicine for patients with left main disease that has been borne out by further trials.

We need to be cautious about subgroup analyses, but clinicians also need to challenge their research colleagues to ask clinical questions of the data. It is not enough to know the average benefit of treatment. We need to know for which patients, under what conditions, treated for how long, and with what other therapies a given strategy is best. If we do not get the answers to those questions, we have abrogated our responsibilities to our patients.

When subgroup analyses are reasonable

Subgroup analyses should be considered, I believe, whenever a subgroup is clinically relevant, or when a treatment difference is anticipated on the basis of previous studies or biologic mechanisms. These analyses should be believed when the difference by subgroups is large, when it cannot be explained by bias and chance, and when it is replicated in other studies. Only then will these subgroup analyses inform and improve the care of our patients, not weaken it.

TREATING REAL, NOT AVERAGE, PATIENTS

Another issue surrounding the use of evidence-based medicine is the treatment paradox. Investigators restrict study populations to patients they anticipate will respond to treatment and exclude patients they suspect will not respond.

For example, most coronary primary prevention trials of cholesterol-lowering therapy have enrolled patients whose cholesterol values were 300 mg/dL or greater, the highest 5% to 10% of the population. But 70% of coronary heart disease cases in men and 50% in women occur with cholesterol levels less than 240 mg/dL. What do we do? Do we restrict treatment to the portion of the population actually tested in the trials, or do we extrapolate the findings to the patients who have never been included in trials, but who account for most cases?

This treatment paradox is a fundamental issue that physicians face every day, because the results of randomized trials rarely include the patients they most commonly encounter in clinical practice. Clinical experience requires that we interpret data from these studies, not that we apply a fixed answer. Attempts to resolve the treatment paradox have produced two groups of physicians: evangelists, who favor a broad application of available evidence to those not studied, and snails, who propose a strict application limited to those included in the research. Thus, controversy abounds.

ARE PHYSICIANS TREATING PATIENTS OR POPULATIONS?

Increasingly, physicians are being asked to take a population approach to the care of patients. Given exactly the same information, do physicians make the same or different decisions for groups than for individuals?

One study asked physicians to make decisions in different scenarios, involving individuals and groups.[11] One scenario involved a college student with fatigue, insomnia, and difficulty concentrating. Physicians were asked whether an additional blood test should be ordered for an uncommon, but treatable condition. They were more likely to order the test when asked to decide for an individual patient than for a population of similar patients included in the health plan.

The source of tension between health plans and physicians

For health plans, achieving the best outcomes for the population of patients is the principal objective, whereas physicians' principal obligation is to achieve the best clinical outcome for individual patients. But health plans pay for the care of individual patients. Conversely, the decisions physicians make for individual patients add up to an aggregate population for which physicians would like to achieve an optimal outcome.

This tension surrounding the choice between treating individuals and populations is often thought to occur solely between health plans and physicians. It is not so simple. The tension exists within the health plan itself, which must make decisions for the care of individual patients. Physicians feel similar tensions as they struggle to care for individual patients yet assure that they are providing the best care for the group of patients who constitute their practice.

THE ART OF MEDICINE AFFIRMED

I believe clinical science affirms the art of medicine. Clinical judgment — with its use of personal experience, analogy and extrapolation — is still necessary in a system of evidence-based medicine. But this interplay between the science and art of medicine increases the responsibility of physicians to learn their patients' risk preferences. A patient who knows that 100 patients need to be treated to prevent three embolic strokes at the cost of one hemorrhage may make a different treatment decision than another patient presented the same data.

The art of medicine will flourish where data are incomplete and blurred, which is much of medicine. It will flourish whenever caring doctors strive to meet the needs of their individual patients. Care of patients is an act, not of application of guidelines, but of interpretation of information.

Physicians will be returned to grace because of this act of interpretation. Scientific evidence must be blended with a physician's experience, reasoning, and knowledge of his or her patients and their preferences. ■

REFERENCES

1. Evidence-based medicine working group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992; 268:2420–2425.
2. Kahn KL, Park RE, Vennes J, Brook RH. Assigning appropriateness ratings for diagnostic upper gastrointestinal endoscopy using two different approaches. Med Care 1992; 30:1016–1026.
3. Bernstein SJ, McGlynn EA, Siu AL, et al. The appropriateness of hysterectomy: a comparison of care in seven health plans. JAMA 1993; 269(18):2398–2402.
4. Ezekowitz MD, Bridgers SL, James KE, et al. Warfarin in the prevention of stroke associated with nonrheumatic atrial fibrillation. N Engl J Med 1992; 327:1406–1412.
5. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting results alter perceptions of therapeutic effectiveness. Ann Intern Med 1992; 117:916–921.
6. McCormack JP, Levine M. Meaningful interpretation of risk reduction from clinical drug trials. Ann Pharmacother 1993; 27(10):1272–1277.
7. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet 343(8907): 1209–1211.
8. ISIS 2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS 2. Lancet 1988; 2(8607):349–360.
9. Detre K, Hultgren H, Takaro T, et al. Veterans Administration cooperative study of surgery for coronary arterial occlusive disease. III. Methods and baseline characteristics, including experience with medical treatment. Am J Cardiol 1977; 40:212–225.
10. Takaro T, Hultgren H, Lipton M, et al. The VA cooperative randomized study of surgery for coronary arterial occlusive disease. II. Subgroup with significant left main lesions. Circulation 1976; 54: Suppl III: III-107–III-117.
11. Redelmeier DA, Tversky A. Discrepancy between decisions for individual patients and for groups. N Engl J Med 1990; 322(16): 1162–1164.

ADDRESS REPRINT REQUESTS to Ralph I. Horwitz, MD, Chairman, Department of Internal Medicine, Yale University School of Medicine, 333 Cedar St. 1072 LMP, New Haven, CT 06520.