Commentary

Improving The Medical Decision-Making Process
by John E. Wennberg

I want to argue the case for the evaluative clinical sciences and the need to support a national program to assess the outcomes of medical care. For too long, and at peril to the welfare of patients and the public as well, we have neglected that essential part of medical science whose job it is to establish the validity of clinical theory and help physicians and patients reach the right decision in their choice of medical care. The result is an intellectual crisis in the scientific basis of clinical practice, which is of growing significance for medicine.

The outcomes of many treatments, including major surgery, invasive and risky diagnostic procedures, and such costly clinical choices as treating a host of acute and chronic medical conditions in the hospital, are not well understood simply because the necessary evaluations have not been done adequately. In many clinical situations, the physician’s treatment and outcome preferences unduly influence the clinical decision process, in part because of confusion about the probabilities for outcomes but also because so little attention is now given to the crucial task of helping patients discover and choose the care they really want. The scientific basis for quality assurance in medicine remains retarded in its infancy, unable to cope with the growing demand for sophisticated technologies to evaluate differences in outcomes among health care providers.

Consider the following statistics. A resident of New Haven, Connecticut is about twice as likely to undergo a coronary bypass operation as is a resident of Boston; for carotid endarterectomy, the risks are the other way around. The numbers of knee and hip replacements per capita are much more common among Bostonians, while New Havenites experience substantially higher risks for hysterectomy and back surgery. The risk of hospitalization for Boston is substantially higher for a host of acute and chronic medical conditions, including back pain, gastroenter-

John E. Wennberg, M.D., M.P.H professor of epidemiology in the department of community and family medicine at Dartmouth Medical School, Hanover, N.H. He has been tracking variations in health care, both in this country and abroad, for nearly a decade and a half.
itis, pneumonia, chronic bronchitis, and diabetes, even though the residents of the two communities are very similar in demographic characteristics related to the need for care.²

These statistics illustrate the intellectual confusion in the heartland of scientific medicine. The residents of Boston and New Haven receive most of their care in hospitals and from physicians who are affiliated with some of the nation’s finest medical schools. The practice styles in these communities have very different implications for costs, but the alternative theories about appropriate practice they represent have gone unchallenged or examined by academic medicine. We remain ignorant of the consequences for patients of spending vastly different proportions of the gross national product (GNP) on health care. The scientific basis of medicine as now constituted does not distinguish the outcome value of an investment equivalent to upward of 16 percent of GNP (as invested in the health of Bostonians) from 9 percent (as invested in New Havenites).³

The crisis has broad implications for the status of the medical profession. The doctor-patient relationship is based on the notion that it is rational for patients to delegate decision making to physicians. The doctor knows best: physicians, because of their formal training, continuing education, and extensive experience, are assumed to know the scientifically correct way to treat disease. Moreover, they are assumed to understand vicariously the needs and wants of patients and thus are qualified to make utility or value judgments for patients.

A number of policy analysts, Kenneth Arrow among them, have argued that it is rational for society as well as patients to delegate decision making to physicians. A key assumption of the rational agency theory is that demand for medical care is controlled by a central consensus among physicians on what constitutes “correct” medical practice.

It is no longer reasonable or feasible to base health policy on rational agency theory. The evidence from small area analysis, from the critical appraisal of strengths and weaknesses of the scientific basis of medicine, and the failure of expert panels to reach consensus on appropriate practice build a consistent and strong case against the rational agency hypothesis and associated assumptions about the nature of demand in medical markets. The intellectual crisis has important economic implications. The unsettled nature of professional opinion on correct practice, combined with the high prevalence of chronic diseases in an aging population and the steady increase in the numbers of physicians trained in invasive technology, ensures a continuing increase in per capita costs and risks of medical care.

Over the next decade or so, the issue of what is appropriate practice will dominate the health policy debate. The debate threatens to be increasingly acrimonious and divisive, pitting physician against physician, specialty group against specialty group, and the profession itself
against the payer and the government, with the patient lost somewhere in the rhetoric. We must move to prevent this caricature of the informed policy debate that is needed to sort out the complicated problems the profession, patients, and the public face, now that it is clear that the scientific basis of clinical practice is much less well developed than previously assumed.

The Promise Of The Evaluative Clinical Sciences

Thomas Kuhn has argued that science does not exist as a progressive, unified body of theory, knowledge, and technique. Rather, there are multiple sciences, each making its appearance at times of intellectual crisis as a “small” revolution, as a shift in “paradigm” that establishes new disciplines in response to anomalies in theory or experimental evidence. A shift in paradigm involves new ways of posing problems and, often, new methods and techniques to address these problems; it requires exemplary research that provides models for how the disciplines should take their place as part of regular science.

Medicine may now be experiencing such a Kuhnian evolution. Long-held assumptions about the efficacy, the ethical sufficiency, and the legal basis of the physician’s role in making vicarious utility assessments for patients as well as the validity of many specific theories physicians hold on appropriate practice are now recognized as problematic. Professional uncertainty rather than consensus about the scientific basis of clinical practice is emerging as the dominating reality. At the same time, new methods and technologies and exemplary applications of these techniques to problems of professional uncertainty also are emerging. It is now possible to speak about a new set of disciplines that together constitute the evaluative clinical sciences. They offer the promise of a scientific program that can greatly decrease uncertainty about the probabilities and the value to patients of the outcomes of care, and can improve the information base for clinical as well as policy decisions.

A number of important disciplinary advances contribute to the evaluative sciences. First, advances in statistical theory and methods make it possible to manage new classes of problems relevant to predicting outcomes and testing theory about alternative treatments. These advances, which permit statistical adjustments for differences in severity of illness in the study of outcomes, are useful for measuring the quality of care and for assessing outcomes in nonexperimental study designs.

Second, advances in medical care epidemiology make it possible to use health insurance claims data and other large data systems to monitor the use and outcome of care in specific locations and to base postgraduate education and quality assurance programs on the feedback of measures of performance. Claims data technology also provides new methods for
estimating outcome probabilities with improved accuracy and at lowered costs. The claims data systems also represent a registry of medical care events and can be used to locate patients efficiently for follow-up studies, thus offering additional ways of lowering the cost of outcome studies.

Third, advances in psychometrics have provided objective and valid measures of patient symptoms and functional status ("quality of life"), opening up new domains for assessing outcomes. These developments are particularly significant for evaluating surgical theory when the reasons for surgery are to improve the quality of life.

Fourth, decision analysis, adapted to medical decision making and applied to a series of prototypical clinical decision problems, allows objective testing of clinical theories through simulated experiments. It also provides a means for assessing the importance of uncertainties about the true probabilities for specific outcomes and for evaluating the importance of patient utilities in a specific clinical choice.

Advances in information technology have enormous significance for the evaluative sciences. It is now possible, using the personal computer, to make calculations and conduct analyses that ten years ago were only possible on large, expensive mainframe computers. Physicians in their offices and on the wards of hospitals can now have at their fingertips the computing power necessary to obtain "real time," precise and specific information relevant to the medical care decisions of their individual patients. Progress in the related area of interactive, computer-driven video disc technology and conceptual breakthroughs in accessing and presenting information ("hypermedia") provide revolutionary new ways for synthesizing, conveying, and individualizing information that can support a luxurious and active cross-communication between the patient and the physician. Through video and other graphic means, patients can see vignettes of the possible futures they face, according to the treatments they may choose. The technology thus holds the promise of greatly activating the patient as a partner in the decision process. This will be particularly useful when the choice involves complex tradeoffs that require evaluation of patient preferences or utilities, such as the decision to live with a symptom or to undergo a risky treatment in the expectation of reducing symptoms.

These methodologies and techniques have been applied in assessments of outcomes. The best-established and understood paradigm is the evaluation of new drugs in which a progression of evaluative techniques are systematically applied. Nonexperimental studies (phase I and II) provide a screen to establish which treatment theories fail, which show promise (and therefore need further evaluation with prospective clinical trials—phase III), and which, as rarely happens, are so clearly effective that they warrant immediate application in clinical medicine.
Because this strategy has been applied systematically to all new treatment candidates, the scientific basis of medical practice with regard to drugs has improved substantially over the past twenty years.

By contrast, theories concerning the value of diagnostic procedures, major and minor surgery, and the value of the use of hospitals are not routinely evaluated. Indeed, many common treatments have not received the careful nonexperimental studies that are considered a routine first screen in the case of new drugs. It is, however, quite feasible to undertake such studies.

One line of research, exemplified by work at The RAND Corporation, uses panels of experts to review in great detail the various clinical indications for which a specific treatment is used. Based on an extensive review of the literature and the experts’ own experiences, the research uses consensus methods to evaluate different theories to classify them as “inappropriate” or “appropriate” care from the perspective of patient outcomes. The panel has achieved considerable success in reaching consensus about inappropriate care, and the method as constituted is quite useful for defining practice patterns that fall outside the pale of reasonable theory, given the current state of uncertainty. For these examples of care, the panel is able to provide outcome-based standards of care of considerable practical use in improving the practice of medicine by reducing the costs and risk of unnecessary care.

The RAND researchers have also shown that, for at least some procedures, it is possible to abstract patient records retrospectively to identify inappropriate practice patterns. The method thus holds considerable promise as a screening approach similar in philosophy to the phase II screens used for drugs and also as a means of feedback to physicians to help them improve the quality of care.

The RAND model is less useful for evaluating ambiguities and uncertainties concerning theories that seem plausible to experts but whose validity has not been tested. This, unfortunately, may well represent the majority situation. Another research model, exemplified by work by David Ransohoff and colleagues and by our own research efforts, shows how the various disciplines discussed above can be used to evaluate reasonably held theories and to come to conclusions about their value. These studies, which compare medical management to surgical management of silent gallstones and prostatism, involve primary data collection to obtain new evidence on key outcome probabilities and a formal decision analysis to test the alternative theories. They show that empirical studies analogous to phase II studies can be undertaken to assess the theoretical basis for medical practice to reach important conclusions on correct theory. The studies point out an important fallacy in a commonly held theory. The evidence refutes the appropriateness of cholecystectomy as a means for extending the life expectancy of patients with silent
gallstones and of prostatectomy for extending the life expectancy of most patients with symptoms of prostatism. For such patients, the operation can be justified only on the basis of expected improvements in the quality of life.

A National Program To Assess The Outcomes Of Medical Care

The evaluative sciences stand poised on the threshold of achievements that can address substantially the underlying causes of the intellectual crisis in medicine. They have achieved virtually all of the requirements to become part of regular science. The methodologies are there. A cadre of scientists is available to do the research, and channels of communication have been established through journals, international societies, and the founding of centers of collaboration, such as the Copenhagen Coordinating Center for the Study of Regional Variations in Health Care. What remains to be secured is public policy to focus priorities and bring the scientists together into an assessment project that meets priorities through a program of peer-reviewed grants in aid similar to that now commonplace for the biomedical sciences.

In the United States, this critical step can be taken easily. Congress has passed legislation to establish a national program for the assessment of patient outcomes to be administered by the National Center for Health Services Research and Health Care Technology Assessment (NCHSR). This legislation targets money from the Medicare trust fund to apply the various methods and strategies discussed above to a set of well-defined assessment priorities. It provides for the establishment of assessment teams to evaluate the major clinical controversies and uncertainties responsible for the large differences in risks and costs of care uncovered by geographic variation studies. Under the priorities of this program as specified in the legislation, assessments would be completed to establish the outcome significance of the major differences in use of surgical technology versus alternative practices, and the outcomes of the costly decision to use hospital rather than ambulatory care as illustrated by the Boston-New Haven comparisons. The program also would actively promote research to fill in the critical gaps in the scientific evidence about effectiveness uncovered by the assessments, promote further development in methodology, and train new scientists. The legislation also requires the essential peer review mechanism that insures that the research proceeds as part of regular science.

A reasonably funded program in this area will have a number of benefits. For the first time, rigorous and scientifically valid assessments that distinguish appropriate from inappropriate practice useful to patients, physicians, and policymakers will become available on an ongoing basis. Scientific rigor is brought to the methods and procedures used in
the increasingly important field of quality assurance. The program also provides the nidus of support to ensure the growth of the evaluative sciences in the nation’s medical schools with several positive long-range effects. It increases the teaching of the evaluative sciences to medical students and physicians and gives these disciplines a stronger voice in the future direction of medical education. It also can serve as a focus for the growth and elaboration of the computer sciences and information technologies, which promise an increasing number of “small” revolutions in the years to come.

An adequately funded program has great benefit for the policy debate on appropriate practice and the related issues of cost containment and medical malpractice. But the most important benefits are for the physicians and patients. The evaluative sciences offer the intellectual tools to strengthen the relationship between the patient and the physician, based on an improved understanding of treatments on health and better decisions concerning the care that patients truly want.

NOTES

1. The weaknesses in the evidentiary basis for clinical decision making are seen most directly in the critical evaluation of medical literature. See: Archie Cochrane, Effectiveness and Efficiency (Nuffield Provincial Hospital Trust, 1972); and Bunker, Barnes, and Mosreller, Costs, Risks, and Benefits of Surgery (Oxford Press, 1977). The literature also documents the divergence between physician utilities and those of their patients and shows how the preferences of physicians for outcomes can dominate the decision process. See McNeil et al., “Fallacy of the Five-Year Survival in Lung Cancer,” The New England Journal of Medicine 299 (1978): 1397–1401. The scientific basis for quality assurance is particularly opaque: the little research that exists is often proprietary, unavailable for critical review.


3. I based the estimate for differences in GNP on a simple ratio of the Medicare per capita reimbursements in Suffolk County (Boston) and New Haven County to the national average reimbursement. In 1982, per capita reimbursements under the Medicare program were $2,647 and $1,561 for Bostonians and New Havenites, respectively. In that year, the average for the U.S. was $1,691, and the percentage of GNP invested in health was 10 percent. Most of the differences in costs between New Haven and Boston are attributable to greater use of hospitals for medical conditions. In 1982, 739 more beds were used to treat Bostonians than would have been used if the admission rates and lengths-of-stay for New Haven had applied. Most of the beds were used to treat patients with a variety of acute and chronic medical admissions, most frequently low back pain, gastroenteritis, pneumonia, heart failure, and diabetes.

4. The focused application of the evaluative sciences to test theories about efficacy of new drugs is in sharp contrast to lack of focus on the evaluation of diagnostic tests, surgery, and the use of hospitals. Licensure requirements prior to market entry make it easy to establish incentives to test new drugs. In contrast, treatment theory in the other fields of medicine often develops as part of the ongoing practice of medicine where the supply of resources and professional practice styles are in a dynamic relationship. Regulation similar to new drugs is not easily feasible, and the cottage industry nature of most provider organizations means...
that capital to support innovation and its assessment is not readily available. Within this context, much of the medical theory and associated practice styles (which evolve as part of the ongoing effort of practicing physicians to solve emergent clinical problems) go unassessed. One policy justification for public investment in a national program to assess the outcomes of care is the need to correct this market defect.


6. While the list of problem areas that need assessment is quite long, the number that achieve priority on the basis of their cost to society and risk to patients is surprisingly small. Some twenty-three operations cover approximately 60 percent of major surgical admissions, and about forty acute and chronic illnesses make up 70 percent of medical admissions.

7. The appropriations authorized under Section 9316 are not sufficient to meet the priorities of the program. The funding level needed is about that of a moderately prosperous institute in the National Institutes of Health. In addition to the public need for unbiased, valid information on outcomes, a substantial investment in this area is justified by the lack of private-sector incentive and capital. Even with the substantial investment I suggest here, the resources allocated to the evaluations of surgery, diagnostic technology, and the use of hospitals will remain minuscule compared to that now invested in the evaluation of new drugs.

8. Past experience emphasizes that a systematic evaluation of medical theory may well lead to lower costs. Poorly evaluated procedures are more likely to be judged efficacious than those that are subjected to careful study. In a review of randomized clinical trials of new surgical and anesthetic technologies, fewer than half represented improvements over existing technologies, and only one of seven represented a marked improvement. See Gilbert, McPeek, and Mosteller, “Statistics and Ethics in Surgery and Anesthesia,” Science 198 (1977). Strengthening the evaluative sciences should also lower the risk of malpractice: many successful suits now are brought against physicians on the assumption that a diagnostic test or treatment should have been done, when in fact the available evidence for effectiveness is inconclusive and sometimes is in the opposite direction. A good example is suits (based on community standards of care, not outcome-based standards) against physicians who fail to use fetal monitoring in low-risk patients. These suits have had the effect of accelerating the use of an unproven intervention.