Physician Decision Making Limited by Medical Evidence

BY JOHN BILLINGS AND DAVID EDDY

Part of the problem of variations in practice hinges on uncertainty over clinical criteria.

HEALTH POLICY REPORT

Physicians' orders differ enormously. Regardless of what it is (ordering diagnostic tests, recommending surgery or admitting patients to the hospital) or how it is measured (comparing how two physicians care for the same or similar patients or contrasting utilization rates among geographic areas), the results are almost always the same. Substantial variation in medical practice is found.

In a Medicaid program in Washington state, 18-fold differences in admission rates for common medical problems were observed. A 20-fold variation in rates for carotid endarterectomies has been documented among 14 metropolitan areas served by major teaching facilities. Physicians on the faculty at a major medical school were shown to differ as much as 2,000 percent in annual lab costs for outpatient care of their hypertension patients.

Reasons for Variations

These differences have been observed consistently since they were first measured more than 50 years ago. But definitive answers about their causes remain somewhat elusive. Five major factors are thought to explain most of the variation:

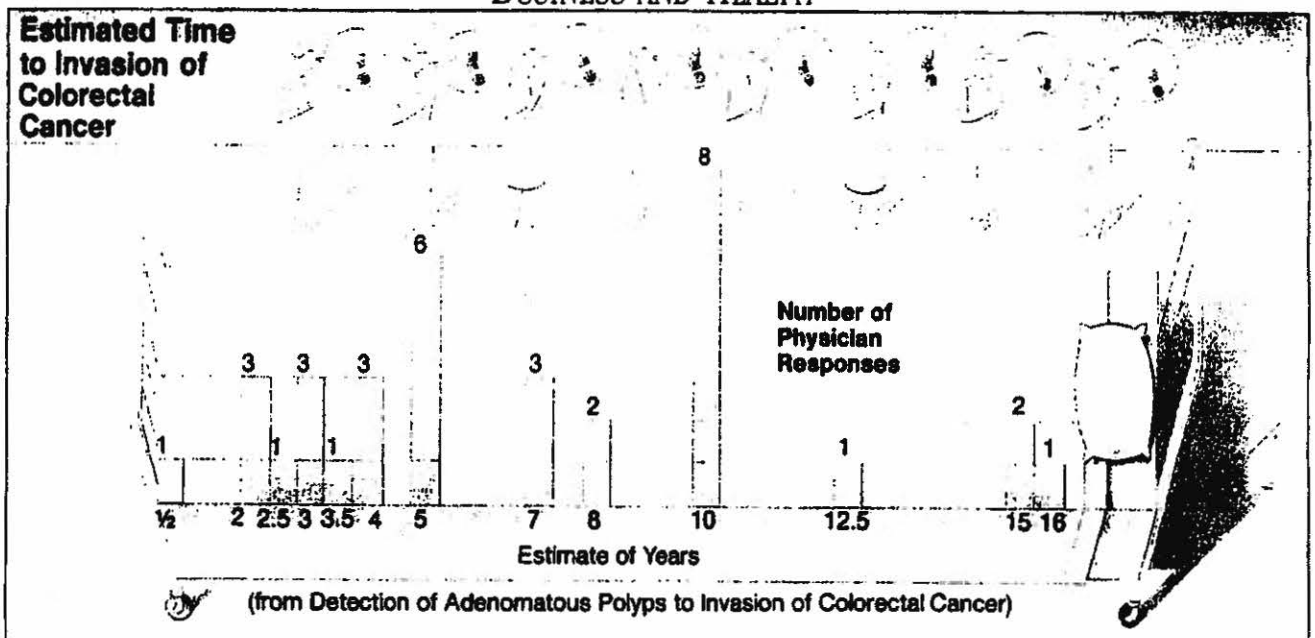
- **Patient characteristics.** Sicker patients obviously require more resources. Risk adverse patients are less inclined to undergo surgery. Poorer patients have restrictions on access, especially for chronic care needs that can reduce utilization of some services such as primary care visits, while increasing use for others such as hospitalization for acute episodes of chronic problems.

David Eddy is director of the Center for Health Policy Research and Education at Duke University in Durham, N.C. John Billings, a health care consultant in New York City, was a visiting professor at the center.

- **Physician proficiency.** Utilization levels and outcomes can be affected by how well a physician performs, what diagnostic information is gathered, how well it is interpreted, what treatment choice is selected and how effectively care is managed.
- **Physician Values.** For some medical decisions, values and attitudes of physicians on how much risk is acceptable or what levels of certainty are required for action can have a substantial influence on what tests are ordered or which treatment is recommended.
- **System Ecology.** A panoply of factors related to the supply of health system resources, how these resources are organized and managed, and how care is financed may have some effect on who has access to care and what care is provided.
- **Disagreement and Uncertainty.** For most medical decisions, physicians rely on policies or rules that have evolved from the research science base of medicine. To the extent there is disagreement about these rules, that the rules are overly broad or nonspecific, or that no rule exists, differences in medical practice are inevitable.

Although the precise effect of these factors is not well established (and is likely to differ significantly by diagnosis and type of decision involved), each undoubtedly plays a significant role in some circumstances, and all have important implications for major questions facing purchasers about whether enrollees are receiving enough care, unnecessary care or even the right care. An improved understanding of these factors will be critical to the emerging efforts by purchasers to control health costs, while at the same time assuring that patients receive quality care that maximizes improvements in health outcomes or other measures of patient satisfaction.

Disagreement and uncertainty have been suggested



as perhaps the most significant cause of variation. One important aspect of this is the effect that deficiencies in the quality of scientific evidence may have on medical decision making.

The Role of Uncertainty

The role of uncertainty and disagreement in inducing differences in physicians' orders is suggested by some of the research on utilization variation itself. John Wennberg of Dartmouth Medical School has noted an interesting pattern in the extent of variation among different diagnostic categories. For serious conditions where there is no disagreement among physicians about the appropriate course of treatment and where utilization is unlikely to be influenced by other factors such as patient characteristics and physician proficiency or values, the level of variation in hospital utilization is small. These low variation conditions include strokes, heart attacks, gastrointestinal bleeding, hip fractures and other injuries. For virtually all other conditions, where uncertainty is greater and where the impact on utilization of patient or physician characteristics may be more significant, considerably more variation has been observed.

This phenomenon is probably best illustrated by the pattern that has been observed by Wennberg for orthopedic injuries. Virtually all patients in the United States with a fracture of the hip, ankle or forearm are likely to seek and obtain medical care. After adjusting for age, differences among communities in the incidence of fractures are likely to be small. Although some fractures are no doubt missed on X-rays, physician proficiency is not a major issue. When is hospitalization required? For hip fractures, the answer is clear. All physicians would hospitalize, albeit for varying lengths stay. For fractures of the ankle and the forearm, the issue becomes progressively more complicated. Physicians may differ in some cases on the need for surgery or on the surgical technique used. Moreover, not all cases require hospitalization. In less serious cases, especially for fracture of the forearm, treatment exclusively on an outpatient basis is likely.

Not surprisingly, when the incidence of per capita hospital utilization for these injuries was examined in Maine, hip fractures showed almost no variation among areas, ankle fractures somewhat more (twofold differences among areas), and forearm fractures the most (eightfold differences). When knee and lower back injuries—conditions where disagreement among physicians about the course or setting for treatment is notorious—were analyzed, even more extensive variation was observed.

This effect of disagreement and uncertainty also was evident when researchers sought explanations for the fourfold differences among counties in hospital admission rates for diabetes that had been observed in Washington state. A more detailed examination of hospital records revealed that variation in utilization was largely attributable to higher admission rates of mild cases in the high rate counties, with almost no difference in admission rates for severe cases. Although differences among counties in disease prevalence were not examined, diabetes would have to have been four times more common in the high rate counties (and physicians in these counties substantially more proficient in managing diabetic patients and preventing severe admissions among their larger population of diabetics) to account for these differences.

A more probable explanation is that physicians have different criteria for diabetes admissions, reflecting different approaches to treatment and management of the disease. For severe cases, for example, involving very high blood sugar levels and/or a diabetic coma, there is no dispute. All physicians would admit those patients to the hospital. But for less seriously ill diabetics, there is substantial disagreement among physicians about the course and setting of treatment. Some physicians seek tighter control of glucose levels than others. In fact, the latter have concerns about the possibility of inducing glucose

levels that are too low. Some physicians use the hospital with its controlled environment to educate patients about diet, medications and self-monitoring either at time of original onset, or to reeducate patients when glucose levels become mildly elevated, as a form of negative reinforcement. Others view the hospital setting as too artificial to encourage the changes in life style required of diabetics, and conduct education programs and management of minor fluctuations in glucose levels on an outpatient basis.

Documenting Clinical Disagreement

Differences in criteria applied by physicians in making decisions about use of diagnostic tests or selection of the course or setting of treatment have been measured more directly. In a recent study at The RAND Corporation and the University of California, Los Angeles, researchers examined the medical literature and developed extensive lists of possible specific indications for use of the following six diagnostic and surgical procedures: coronary angiography, coronary artery bypass surgery, endoscopy, colonoscopy, cholecystectomy and carotid endarterectomy (see BUSINESS AND HEALTH, September 1986, p. 18). These indications then were presented to panels of nationally recognized experts in the clinical areas involved who were asked to rate the appropriateness of the indications for use of the procedure. The panel used a scale of 1 to 9, with 1 meaning the procedure was considered extremely inappropriate, 5 meaning its use was considered equivocal and 9 meaning the procedure was extremely appropriate. Initial ratings were done by mail, but final ratings were made after a discussion of each indication at a meeting of the panel of experts, except for colonoscopy, which was not discussed at the panel meeting.

The results are startling. The disagreement on indications ranged from 30 percent to 81 percent among panel members. Disagreement was defined as being when at least one of the nine panel members rated the indication between 1 and 3 on the scale and at least one rated it between 7 and 9. In many cases, at least one panel member gave an indication a rating of 1, while another gave a rating of 9 (for example, 61 percent of the time on indications for colonoscopy). The range of agreement on specific indications—that is, when all panelists gave ratings within three points of each other—varied from 3 percent to 41 percent. Individual panel members also exhibited a significant degree of uncertainty. For example, 20 percent of the panelists' ratings for coronary angiograms were between 4 and 6, and the median rating was in this mid-range for 14 percent to 29 percent of the indications for the other five procedures.

These figures actually may understate the level of disagreement or uncertainty that exists in typical community practice. Panel members in the study were experts selected by the leadership of American medicine, representing the nation's most prestigious medical schools. Moreover, panelists were provided with an extensive written summary of the medical literature including copies of seminal articles relating to the procedures under consideration, and reviewed the indications at a two-day meeting, which gave them a significant opportunity to help develop some degree of consensus. In fact, the findings described above come from the final ratings of the panel; the initial ratings made prior to the meeting showed substantially more variation. For example, the level of disagreement on coronary angiograms was cut in half in the final rating.

To some extent, the RAND-UCLA study measured something more than medical disagreement or uncertainty, such as differences about how the human body works, about the natural courses of disease, or about the effectiveness of diagnostic and treatment technologies. Some of the disagreement also may reflect differences of opinion about the relative weights assigned to the various risks, benefits and costs that are considered in determining whether a procedure is appropriate or inappropriate. This was seen, for example, in the ratings for cholecystectomy, where four levels of patient comorbidity (none, low, medium and high) were included for each indication. In many cases where there was unanimous agreement that surgery was extremely appropriate for patients with no comorbidity, disagreement was exhibited for the indication when high levels of comorbidity were present, possibly indicating different attitudes towards the risks involved.

"Differences in criteria used by physicians in making decisions about the use of diagnostic tests or selection of the course or setting of treatment have been measured. . . . The results are startling. Panel members disagreed on their final ratings for 30 percent to 81 percent of the indications."

The Case of Colorectal Cancer

It is possible to get beyond any effect of these value considerations by asking physicians directly about their assumptions concerning human physiology, disease pathology or diagnostic and treatment effectiveness that necessarily underlie such indications or decision rules. A group of physicians led by the Center for Health Policy Research and Education at Duke University recently developed a mathematical model to help estimate the effectiveness of various strategies to screen high risk people for colorectal cancer. The modeling technique was believed necessary because there were no data from randomized controlled trials for the majority of the possible strategies, and even where results were available, the studies were difficult to interpret or did not involve high risk patients.

To construct the model, it was necessary to incorporate information about incidence rates for colorectal can-

cer, the natural history of the disease such as the percent of cancers coming from polyps and how rapidly cancers develop, the effectiveness and risks of the various screening techniques, and the effectiveness of treatment approaches at alternative stages of the disease.

Because in many cases adequate data were not available from the published literature, questionnaires were sent in 1983 to more than 70 experts to help estimate parameters for each of these factors. Participants in the survey were selected for their knowledge of the incidence, pathology, detection and treatment of colorectal cancer, and were encouraged to consult with colleagues and the literature. Fifty-four questionnaires were returned.

Again, the extent of disagreement often was substantial. For example, although most experts believed that the overwhelming majority of cancers arise from adenomatous polyps (the median estimate of those surveyed was that 90 percent come from polyps), 10 percent of the respondents estimated the proportion to be 20 percent or less. Similarly, although most experts believed that relatively few polyps ultimately become cancerous (the median estimate was 5 percent), some believed the proportion to be as high as 80 percent to 90 percent. There also was considerable disagreement about how fast cancers develop, with some estimating the average time from detectability of adenomatous polyps to development of an invasive cancer to be as rapid as six months, while others believed 16 years was more probable (see graph).

Screening Methods Debate

Answers to these questions are critical in deciding which screening tests to use and how frequently. But there was even more disagreement about the effectiveness of some of the screening technologies. For example, the fecal occult blood test (FOBT) detects blood in the stool, a possible indication that an adenomatous polyp may be present in the colon. But when asked what porportion of polyps actually bleed prior to becoming cancerous invasions, and therefore are detectable by the FOBT as part of a possible screening strategy, the answers of the experts were scattered broadly between 2 percent and 100 percent, with no apparent consensus among those responding. Not surprisingly, there also was considerable difference of opinion about the accuracy of the FOBT, with estimates for the false negative rate (the probability of missing adenomas that were present) ranging from 1 percent to 50 percent, and estimates of the false positive rate (indicating a possible adenoma when none is present) varying from 1 percent to 80 percent. Similar variation was displayed for screening technologies such as rigid and flexible scopes.

Another survey of experts illustrates the potential effect that this disagreement and uncertainty about the science base can have on clinical decisions and medical policies. At a 1981 meeting of international leaders in colorectal cancer detection, the attendees were asked the following question: "What is the overall reduction in colorectal cancer incidence and mortality that could be expected if men and women over the age of 50 were tested with fecal occult blood tests and a 60-cm. flexible sigmoidoscope every year?" The answer to the question is central to individual patient decisions and the development of any broader policies about cancer prevention. Again, the responses of the experts varied enormously from estimates close to zero reduction in incidence and mortality to estimates approaching 100 percent. It is tempting to suggest that physicians with less expertise might evidence even more disagreement, but a wider degree of variation is hardly possible.

How can this be? How can experts in a field of science disagree so extensively and display so much uncertainty?

Unfortunately, full and complete answers are not possible. Although the medical profession has long recognized that controversies exist about some medical practices, the extent and degree of disagreement and uncertainty only recently has become apparent as purchasers and policy makers seek explanations for variation in utilization rates and as economic pressures stimulate efforts to define and promote optimal medical practices in terms of health outcomes and costs.

Three Principal Factors

Accordingly, the sources of disagreement and uncertainty seldom have been studied directly. What is known often must be drawn from other unrelated research, supplemented by conclusions based on an understanding of the nature of medical care and human behavior. But it is becoming more clear that much of the disagreement and uncertainty may be attributed to three major factors: the imperfect state of knowledge of human physiology, disease pathology and treatment effectiveness as limited by the capacity of existing technology; the quality of scientific evidence describing what is known or is knowable, and on which clinical decisions and policies must be based; and the manner in which clinical policies that guide individual patient decisions are developed and disseminated.

To the extent that in some areas of medical practice answers are beyond the reach of current knowledge, there can be few certainties, and some degree of variation is inevitable. The goal of the health system and purchasers in these circumstances can only be to assure that patients are fully informed and that the best possible effort has been expended to determine and interpret what is knowable. But for the other two contributing factors, deficiencies in medical evidence and inadequacies in design of clinical policies, the implications for costs and quality of care are more troublesome. More, in fact, appears to be knowable about the effectiveness of some medical technologies—and rules that govern many clinical decisions could be improved by incorporating this additional information on outcomes and by applying more rigorous methods in their development. The actions needed to accomplish changes are likely to be difficult. Pressure from the purchaser community will be vital to stimulating these efforts. ■

Part II: The impact on decision making of deficiencies in medical evidence.