The risk that an individual living in the United States will undergo a given surgical procedure or be hospitalized for a given medical condition is heavily influenced by where that individual lives, John Wennberg has conclusively demonstrated this fact in his studies of small area variations in the rates of both specific surgical procedures and specific episodes of medical care.¹ Studies conducted in Canada show similar variations, not all of which can be explained by economic incentives inherent in the payment system or the numbers of available physicians and surgeons.² Variations in the rates of use of medical services by different population subgroups are an attractive new target for those committed to cost containment in health care. If population A has a surgical rate for a given procedure that is four times the rate in population B then, so the argument goes, everyone would benefit and money would be saved if only A’s rate could be reduced to equal B’s. This logic can lead to requests, such as Wennberg’s own, that physicians receive better information on their local rates in the hopes that such information can be used to shame those with high rates into reducing them. Carried to its logical extreme, the same reasoning dictates some form of enforcement of low rates either through rewards, punishments, or direct control.

The idea of using population-based data to support utilization review activities is not new. Michael Goran, the director of the Professional Standards Review Organization (PSRO) Program, suggested using Wennberg’s work for review purposes in the mid-1970s. Potential reductions in high rates of use have become particularly popular recently, perhaps because the perverse incentive inherent in the system of prospective payment to hospitals is so clearly to increase rates of use in the local population, perhaps because the kinds of data bases needed to reproduce Wennberg’s studies are only now becoming generally available.

Glorification of the golden mean (or, more accurately, of the low end of the normal curve) has become so popular that a commentator at-
tempting to raise caveats is likely to be dismissed as an eccentric groucher. When the president of the American Medical Association, the policy representatives of a notably conservative administration, a foundation representative speaking for consumers, and a variety of researchers of varied political stripes all concur in a single issue of Health Affairs that “regional variations in utilization patterns constitute a challenge that will have to be met aggressively in the coming decade,” there seems little left to add.³

In fact, however, a primarily regulatory response to Wennberg’s elegant work is dangerous and irresponsible. Rate variation can clearly be attributed to wasteful variations in practice patterns in some settings and at some points in time. Rate variation is also the hallmark of improvements in the science of medical care and the direct effect of the enormous variations in the health of the underlying population.

When Variations Are Acceptable

There are three settings in which variations in the use of medical services are either acceptable or highly desirable. These are: (1) when uncertainties in the science of medicine lead to acceptable alternate practice patterns; (2) when an innovation in diagnostic or treatment modality is in a phase of active dissemination; and (3) where the variation reflects underlying differences in the population’s health status.

Uncertainty. The role of uncertainty in medical practice has been well described by David Eddy and will not be reviewed in detail here.⁴ When approaching rate variations, the real issue posed by uncertainty is: When are two treatments for the same condition of equal merit? The most simplistic and common assumption is that a surgical approach should always be avoided if a medical substitute is available. While this logic has an innate appeal to anyone trained in internal medicine, it does not stand up to critical examination in all cases. Medical treatments may be more expensive, less acceptable to the patient, even more dangerous in the long run than surgery. What is needed is a far more sophisticated ability to balance all aspects of the choice of treatments, including the patient’s preference and the long-term risks, costs, and benefits of each alternative.

Dissemination of innovations. New treatments and approaches to diagnosis are not adopted uniformly by all hospitals and all physicians at the same time. When a new procedure has been developed, tested, and reported, it typically goes through a period of dissemination. Teaching centers are apt to adopt new procedures more quickly than community hospitals; young physicians entering practice will adopt technologies learned during training more quickly than will their established colleagues. Payment methods, the availability of capital, and a variety of
other influences including how great a margin of improvement the new technology appears to offer will also affect the rate at which technology is disseminated.

As a result there is always a period for any technology when the rate at which it is used is far higher in some areas than others. Any effort to enforce average rates or to drive all areas towards low rates will work against the dissemination of all new technologies no matter how beneficial or effective these technologies may be.

Although the literature on small area variation is astonishingly silent on the subject of innovation, a close reading of the available data tends to confirm the existence of the dissemination effect. For example, in their study of surgical rate variations among the elderly in Canada, Noralou and Leslie Roos examined three types of procedures: elective surgery, complex surgery, and cataract surgery. Population-based rates were developed for each type of surgery for fifty-six small service areas. For elective surgery, the ratio of the lowest, median, and highest rate was 1 to 1.8 to 3.1; for complex surgery it was 1 to 2.2 to 3.7; and for cataract surgery it was 1 to 2.3 to 4.2. Rates of complex surgery, in other words, showed higher variation than the group of elective procedures; cataracts when studied alone showed the highest variation of all.

The study, which used data from the early 1970s, defined elective surgery to include seven procedures most commonly referred to as discretionary or elective. Complex surgery included all procedures having a hernia equivalent of 2.0 or greater; the authors cite four examples of complex procedures.

One possible interpretation of these data is that the high rate variations observed in the complex group were, at least in part, dissemination effects. In complex surgery, two typical procedures included in the study above are in no sense elective: bilateral nephrectomy is done only in anticipation of kidney dialysis or transplantation; bilateral orchiopexy only in the aggressive treatment of certain cancers. Neither procedure has much likelihood of being performed unnecessarily. There are two plausible explanations for differential rates in procedures such as these: one is a differential rate of the availability to Canadians in 1971 of lifesaving treatments for renal failure on the one hand and of effective treatment of prostate cancer on the other; the other a differential rate in the underlying population of renal failure-and prostate cancer.

Cataract surgery presents a more complex picture. The procedure has improved markedly in ease, safety, and outcome in the last twenty years. As a result, the criteria for surgery have steadily changed as well. The old common wisdom that cataracts should "ripen" until the patient is near blind no longer holds. Unfortunately, no studies documenting outcome and identifying appropriate criteria for surgery have been conducted so that no clear-cut identification of an "acceptable" rate can be made.
While high rates cannot be uncritically accepted as suitable, the assumption that the lowest rates are ideal is difficult to support. In fact, rates of cataract surgery in the United States have continued to increase steadily since the 1970s, a fact which suggests that the large differential rates observed by the Roos may well have represented a pattern of change over time.

Data such as these are only a tantalizing hint. There is, however, a substantial literature on the dissemination of new technologies in medicine which deserves careful linkage to studies of rate variation. Such a linkage requires a critical look at the outcomes of care; new technology cannot be uncritically accepted as good. At the same time it should certainly not be uncritically rejected as bad. Enforcement of low rates of procedures essentially implies the latter point of view.

Population differences, How much health care does a population need? The question is a thorny one that has occupied—and supported—several generations of scholars. Multiple factors affect use: age, sex, socio-economic status, out-of-pocket costs, and the numbers of available doctors have all been shown to be influential. These factors interact in unpredictable ways; one of the interesting and less-discussed findings in the Rand study of national health insurance is that families with high out-of-pocket costs for care will reduce the use of services by adults far more than they reduce use by children. This suggests that some episodes of care are viewed as necessary across a wide range of costs while others are not. A careful look behind the aggregate data would probably reveal that, even for adults, different types of care show different levels of price sensitivity.

Even though a simple needs model cannot explain all use of services, need as determined by underlying health status must be considered a major factor in the equation. This influence is hard to describe and characterize; it is best seen in the simple fact that one of the best predictors of an individual’s use of services in any given year is the same individual’s use in the previous year. A more sophisticated description of this phenomenon would require the ability to synthesize the epidemiology of a myriad of diseases, each of which shows up in relatively low numbers in any population studied.

Such a synthesis is by no means impossible. One problem to date is that major studies of of health sevices, such as the Rand experiment, have tended to focus on the “normal” population: the able-bodied and employed. The effects of health status would be more easily observed and might well be better understood if more effort were made to study in detail the high users such as the disabled or the victims of specific diseases.

Epidemiologists have known for a long time that illness is not evenly distributed among the population. Some of the maldistribution can be attributed to life-style and behavior- hence the old saw about the differ-
ential life expectancy in the neighboring states of Utah and Nevada. Others, such as chronic obstructive pulmonary disease and lung cancer among miners are heavily influenced by known environmental causes. Still others remain mysteries: multiple sclerosis, for example, is not only more prevalent among whites than blacks and among women than men, it is also almost twice as common in individuals living above the 37th parallel than in those living below.

These examples are presented not as a series of interesting anecdotes, but as evidence for our need to integrate epidemiology into health services research. Chronic obstructive pulmonary disease and multiple sclerosis are severely disabling diseases; their victims have a long life expectancy after the onset of illness and a high need for costly services. Maldistribution of these two conditions within the general population is unlikely, in and of itself, to explain large variations in cost and use, but it does explain some. Careful analysis of the distribution of a wide range of similar diseases among the population will explain still more. Only by integrating sophisticated epidemiology into the examination of use patterns can we accurately determine the influence which need exerts on use.

Conclusions

Studies of small area variations in medical practice have made a major contribution to our understanding of the use of health services. The accumulated data clearly show that certain specific procedures and treatments are used more frequently in some areas than others. The power of this finding is enhanced by the fact that to date the work has been done on relatively homogeneous, usually rural, populations in which little variation in health status has been reported.

The available information on rate variations can be used to focus utilization review efforts towards specific procedures or to guide employers towards contracting with low-use and low-cost health centers. The information cannot, however, be extrapolated to mean that overall variations in the use and cost of services are attributable solely to practice variations or that overall reductions towards low existing rates are a desirable goal.

Future work in small area variations should not focus exclusively on practice patterns but should be expanded to explore the interactions between practice patterns and other factors. Integration of existing research on the dissemination of medical innovations and on the epidemiology of chronic disease is needed before we can say with real authority that the nationwide variations in use are in fact a challenge that requires aggressive correction.
NOTES


3. The entire Summer 1984 issue of Health Affairs was devoted to the subject of small area variations. This quotation is taken from the article by James H. Sammons, representing the American Medical Association.


6. Lens extractions, prostatectomy, hysterectomy, cholecystectomy, hemorrhoidectomy, varicose vein stripping, and inguinal hernia repair are included in the definition of elective surgery.
