

International differences in medical care practices

by Klim McPherson

An overview of several aspects of international comparisons of medical care utilization is presented with a discussion of the usefulness of such comparisons in identifying geographic variations in utilization and in elucidating the nature of clinical decisionmaking regarding various procedures. The

discussion includes the purposes of conducting international studies as well as the methodological and policy issues involved. Brief descriptions of some of the studies that have been conducted are also provided.

Introduction

Health care is consuming ever-increasing proportions of developed nations' budgets. As populations age and the ability to provide effective intervention increases, medical care inflation continues to outstrip retail price indices. The aggregate utility of these expenditures, as well as each new increment that results from new diseases such as acquired immunodeficiency syndrome, new techniques such as organ transplants, technological advances in diagnostic equipment, and more sophisticated drug therapies are being questioned by governments faced with the provision of adequate health care that requires more real funding in each year than it did in the previous one.

We are entering an age, therefore, where questioning will be axiomatic in health care provision. New techniques will no longer be universally implemented without evaluating value versus cost. Even common procedures will come under more intense scrutiny as the need for justification increases. The nature of this progression increasingly becomes a rationing process. However, to ration in medicine is to do something which is quite alien to health care provision as it has evolved. One has to be absolutely certain that real benefit is not being withheld; incontrovertible evidence of efficacy or lack of it is needed as a prerequisite for rationing.

The resolution of the health care dilemma is hindered by two factors. The first is that this era of questioning is somewhat threatening to the medical profession, which has taken, and been given, decisionmaking responsibility and power (Friedson, 1972). The second, and in the end the real hindrance, is the difficulty with which many of the important questions can actually be answered.

If limited resources are to be focused on the provision of appropriate care, one must know what appropriate care is. In health care, there is a diversity of accepted opinion on the need for and value of alternative treatments. In many situations, equally qualified physicians might disagree on which treatment is optimal. There is often no scientifically correct way to practice much of medicine. Many

accepted theories concerning the treatment of illness have not been adequately assessed, and consensus based on knowledge of treatment outcomes is the exception rather than the rule.

Overall efficiency

The aggregate cost to a population of hospital health care, measured in terms of annual costs per capita, is the product of two independent components. The first is the average cost per admission. This is intensively studied and relatively easily measured, and attempts to monitor contributing factors, such as diagnostic tests, length of stay, or manpower costs, can greatly affect its magnitude. The second component, less intensively studied but often more important, is average annual admission rates per capita. This component is often assumed to reflect medical need and, as such, is not subject to questioning; to question admission itself assumes a broader concept of efficiency than is usual. However, many causes of admission are associated with large variations in their per capita rates and, therefore, can be strong determinants of per capita health expenditure.

Overall efficiency requires that the aggregate activity of the hospital service maximizes the benefit-to-cost ratio of all alternative admission and process options. This means that those patients for whom the greatest benefit is realistically expected are admitted in preference to those for whom little benefit can be expected and that, among those admitted, the therapeutic options should maximize benefit-to-cost ratios. Such criteria lead inexorably to a greater interest in measuring the outcomes associated with hospital admissions and in comparing such outcomes with those associated with alternative forms of treatment.

Measuring outcome

The crucial yardstick by which all aspects of medical care will come to be measured will inevitably be outcome and, in particular, the improvement in outcome consequent upon the particular intervention. This is the benefit. There are many problems with its measurement, not least of which is the placebo effect associated with almost all supposedly active medical intervention (Beecher and Boston, 1961).

Reprint requests: Klim McPherson, University of Oxford, Department of Community Medicine and General Practice, Oxford, OX 26 HE, United Kingdom.

Disentangling this effect from a "real" therapeutic effect is one of the major problems with its measurement.

Moreover, there are many dimensions of outcome to which different people will attach different weights (Sacket and Torrance, 1978; Llewellyn-Thomas et al., 1984). For instance, for some, the length of survival with cancer is more important than the quality of life experienced (McNeil, Weichselbaum, and Pauker, 1978). While, for others, any suffering is worth avoiding even at the cost of extended life. Such individual preferences would indicate the need for discrete sets of outcome parameters (Mulley, 1989). The probabilities of achieving these sets associated with different clinical decisions, or indeed complex sets of interwoven decisions, should be known. Only then can rational choices be made about the provision of health care and whether expenditures are justified when other possibilities are considered.

Not only are the pertinent questions often complex, but the data that are available to help answer the questions are often limited. To measure outcome, one needs followup. Routine health statistics rarely provide such information because patients are lost to the system once they are discharged. Therefore, to measure even mortality 6 months after discharge is usually impossible outside the environment of special studies, and to assess quality of life is more difficult still.

Origins of clinical uncertainty

Advances in medical knowledge have come, to an extent, from undirected basic research (Comroe and Dripps, 1977). The basic knowledge that results from research and development is then formulated into a clinically usable form, sometimes through evaluation with or without clinical trials and sometimes without evaluation. The clinical practices that emerge from this inconsistent process of diffusion may begin with a strong science base, but this base is often gradually weakened as it evolves through several stages to clinical practice (Fineberg, 1985; Bunker and Fowles, 1982). Herein lies an essential paradox in the study of medical practice. Medicine is widely held to be a science, but many medical decisions do not rely on a strong scientific foundation, simply because such a foundation has yet to be fully explored, developed, and tested.

What often happens in the medical decisionmaking process is a complicated interaction of scientific evidence, patient desire, doctor preferences, and all sorts of exogenous influences, some of which may be quite irrelevant. This tends to mean that the extent to which individual clinical decisions can actually be justified by a coherent body of scientific knowledge is likely to be variable. More importantly, it is not always obvious where, within this spectrum of variability, a particular clinical judgment might lie.

The frank recognition of the existence, or the extent, of clinical uncertainty by health professionals

can be difficult, however. People, on the whole, find clinical uncertainty disconcerting both when ill and when responsible for treating illness (Ingelfinger, 1980) and there is a tendency to disguise it. Patients who are concerned with their symptoms are happier if they can believe that what is being done can be justified scientifically, and health professionals command more respect if what they do is based on professional expertise. However, to question and evaluate medical care practice fairly (so that rationing can be rational), it is necessary to recognize all important uncertainty that exists.

Detecting important uncertainty

Some insight into the variation that exists in determining medical treatment is provided by epidemiological investigation of clinical consistency. In recent years, hospital use and procedure rates have become the subject of intensive investigation in many countries with a view to describing and understanding the nature of clinical decisionmaking (Bunker, Barnes, and Mosteller, 1977; Aaron and Schwartz, 1984; Wennberg and Gittelsohn, 1982).

As early as the 1930s, differences in tonsillectomy rates were observed among school districts of Southern England (Glover, 1938). The work of Bunker (1970), Vayda (1973), and McPherson et al. (1981) have documented the extent of cross-national variation in many population-based hospital use rates and drawn attention to the generally higher rates in North America compared with the United Kingdom or other European countries for which data exist. These differences are sometimes of such magnitude that important questions are raised about the causes and consequences, such as resource cost implications, which are easily estimated (McPherson, 1988).

Small area variation studies

During the 1970s, work on variations in rates led to the study of small geographic areas. Although gross differences in morbidity, in need, or in access to health care among relatively homogeneous communities should not exist, gross differences on a per capita basis in the use of many operations or procedures were recorded. Some of this variation resulted from differences in the supply of facilities, but differences in clinical decisionmaking were also reflected. Although international differences in use rates would rarely be entirely attributable to clinical differences, such an explanation was much more difficult to avoid for small area variations. In fact, ancillary evidence from surveys of need and illness rates only served to confirm such an explanation. Moreover, the nature of the observed variation was consistent with the level of certainty involved in determining appropriate medical treatment. Those conditions, such as hip fractures, that invariably required hospitalization exhibited little variation in their rates among small areas.

Table 1

Magnitude of systematic variation (in ascending order) for selected causes of admission among 30 hospital market areas in Maine: 1980-82

Variation	Medical	Surgical
Low: 1.5 fold range	—	Inguinal hernia repair Hip repair
Moderate: 2.5 fold range	Acute myocardial infarction Gastrointestinal hemorrhage Cerebrovascular accident	Appendectomy Major bowel surgery Cholecystectomy
High: 3.5 fold range	Respiratory neoplasms Cardiac arrhythmias Angina pectoris Psychosis Depressive neurosis Medical back problems Digestive malignancy Adult diabetes	Hysterectomy Major cardiovascular operations Lens operations Major joint operations Anal operations Back and neck operations
Very high: 8.5+ fold range	Adult bronchiolitis Chest pain Transient ischemic attacks Minor skin disorders Chronic obstructive lung disease Hypertension Atherosclerosis Chemotherapy	Knee operations Transurethral operations Extraocular operations Breast biopsy Dilation and curettage Tonsillectomy Tubal interruption

SOURCE: (Wennberg, McPherson, and Caper, 1984).

Wennberg and Gittelsohn (1982) proposed that variations in procedure rates reflected supplier-induced demand. Often patterns of procedure-specific, population-based rates existed in hospital service areas that could not be explained by the characteristics of the populations served and were sustained over several years until there was a change in clinical personnel. The concept of supplier-induced demand was further supported by studies of physician feedback programs where information on rates was provided; in geographic areas where rates were found to be high, a common physician response was to reduce procedure rates (Lembcke, 1956; Dyck, 1977; Wennberg and Gittelsohn, 1973; American Medical Association, 1986). Therefore, variations in hospital use rates between communities that are similar in major determinants of health, need, or use and that are larger than could be explained by chance are likely to be a manifestation of clinical uncertainty, i.e., differences in clinical opinion (Wennberg et al., 1987). This leads to comparison of the amount of systematic variation between neighboring small areas across procedures as well as across countries or health care systems. Such comparisons require a metric for measuring variation that is robust and excludes the random component of variation (McPherson et al., 1982; Roos, Wennberg, and McPherson, 1988).

In making such comparisons, some procedures are found to exhibit much more variation than others (Table 1). If this variation is taken as a measure of clinical uncertainty among professionals, then the uncertainty attached to determining admission can be compared. Hysterectomy, for example, exhibits more variation than appendectomy. The same procedures exhibit as much relative variation in the centralized system of the United Kingdom as they do in the United States despite higher aggregate rates in

North America (McPherson et al., 1982). Therefore, the clinical uncertainty concerning the indications for these procedures and for hospitalization for medical diagnoses in general is a function of the procedure or diagnosis itself, rather than the health system through which care is provided. As shown in Table 1, there is a hierarchy of implied uncertainty. Such information is invaluable in the interpretation of international differences in rates.

For one study, 90 percent of admissions exhibited greater variation among neighboring communities than occurred for hysterectomy rates (Table 2). Such a phenomenon may indicate the level of clinical uncertainty and may encourage clinicians to question their therapeutic decisions, but it does not provide necessary information about appropriate use rates. Low rates are neither necessarily better nor worse than high rates where patient welfare is concerned.

Table 2
Percent of admissions categorized by typical diagnosis-related groups with different characteristic variation among 30 hospital market areas in Maine: 1980-82

Variation	Typical diagnosis-related group	Percent admissions	Cumulative percent
Low	Hernia	1.1	1.1
Moderate	Appendectomy	8.9	10.1
High	Hysterectomy	42.3	52.4
Very high I	Disc removal	31.7	84.1
Very high II	Tonsillectomy	15.9	100.0

NOTE: The number of admissions on which percents are based is 428,056.

SOURCE: (Wennberg, McPherson, and Caper, 1984).

Such large variations, however, can lead to acceptance of prospective clinical trials to determine the parameters of appropriate care.

Reasons for differences in rates

When comparing rates for health care practices among countries, it is important to be aware of all of the possible reasons for observed differences. Many aspects of health care differ among countries (Schieber, 1987; Poullier, 1985). The utility of any comparisons depends on the extent to which competing explanations are determined to be causative. By adopting a somewhat simplistic view of the purpose of health care, certain causes of variation can be designated as legitimate and others as artifactual.

Legitimate causes of variation

The populations being compared may have different prevailing rates of illness for which the intervention is appropriate. This alone could cause observed differences in rates. In general, comparative morbidity rates are difficult to obtain because they are measured by admission rates or consultation rates which are themselves confounded by medical practice variations.

Genuine surveys of morbidity could give insights into differences in rates, but they would be expensive to do reliably. The single exception would be cancer incidence rates (Muir, 1976) where rates of disease are uncontaminated by supply or clinical preferences. However, even these are subject to artifacts of recording, such as diagnostic ambiguities, incomplete enumeration of the population at risk, or omission of cases from private or religious hospitals. Thus, systematic differences among countries may not be entirely a manifestation of real differences in incidence (Doll and Peto, 1981).

The same difficulty exists for comparative morbidity rates that rely on cause-specific mortality rates. The designation and coding of death certificates are a function of local practices and may not reflect true epidemiological differences. However, large differences in genuine incidence of disease may exist, particularly between the developing and developed world.

Different age and sex characteristics of populations also have an impact. Diseases are usually more common with increasing age and more prevalent in one gender than the other. If a population consists of a higher proportion of elderly females (as in Sweden), then this could explain differences in rates. Therefore, comparisons among countries should be standardized for the age and sex distribution of the populations, when possible. This way, any residual differences are unlikely to be a consequence of different demography. However, among the countries of the Organization for Economic Cooperation and Development (OECD), demography should not be a major determinant of large variations in rates.

Artifactual reasons for variations

When comparing rates, it is vital to exclude all artifactual reasons for differences. For example, high rates of intervention in previous years may give rise to low rates for current time periods. For relevant comparisons to be made, the rates need to be related to the estimated population at risk (Gittelsohn and Wennberg, 1976).

Hospital use statistics may underestimate the real population because of the systematic exclusion of some hospital admissions, for example in private or religious facilities. Also, patients who are discharged on the same day as admission (day cases) will often not appear in hospital statistics (Nicholl, Beeby, and Williams, 1988). A low rate in a community for hernia repair may indicate a larger proportion of day cases and a preference for day surgery rather than reflect any other practice variation. If one must be careful that population estimates provide the proper denominators for the calculation of rates, it is equally important to account for cross-boundary flow to ensure that numerators are not inflated by people not at risk of admission in the geographic area (Wennberg and Gittelsohn, 1973).

In addition to the completeness of recording all admissions and accurate and reliable population estimates, there may be other coding artifacts that require careful scrutiny. For example, some health information systems record up to three procedures per admission while others record only one. An incidental appendectomy that would be included in the former would be excluded in the latter. Also, the nomenclature used to record operations may differ, for example, "cholecystectomy" versus "operations on the gallbladder" (McPherson et al., 1981).

Such considerations imply the need for great care in interpreting rates, particularly among countries where significant distinctions exist in the mode of care and data collection activities. In comparing rates among countries, one has to be certain that an important part of the observed variation is not attributable to any of these factors. However, known epidemiology gives insights into plausible differences in illness rates and their relationship to age and gender, and limits on likely variation can be set. The observed variations outside these limits that are not artifactual are taken to be manifestations of practice variation (and in large part, the level of professional uncertainty concerning appropriate treatment) until demonstrated to be attributable to something else.

Clinical judgment

Clinicians may differ in making diagnoses, but even assuming the same diagnosis, they may have different opinions about the relative merits of various treatment options for a given condition, in the absence of biological certainty. Their beliefs are based on their respective educations, understanding of the literature, and personal experiences in practice. Whatever the basis for these beliefs, medical evidence is open to

interpretation and may later be proven wrong. Such evidence may nonetheless appear convincing. When clouded by financial and professional considerations, such beliefs are more difficult to evaluate.

The greater concern is that strongly held beliefs can prohibit randomized comparison, which provides the most reliable information on the relative efficacy of competing treatments (Hill, 1962; Cochrane, 1971). It is the lack of this reliable information, which in turn contributes to the level of uncertainty, that ultimately impacts on the variation found in medical practice. Unfortunately, the determination of medical efficacy, in all its dimensions of outcome, is often extremely difficult. The consequence of doing one rather than another intervention for a given disease state is, in such circumstances, imprecisely understood by anyone, so clinicians must rely on their own best judgments and some medical consensus where it exists.

Prevailing custom

Some communities might eschew certain kinds of medical intervention more than others, notwithstanding availability or recommendation. This may be the result of prevailing medical opinion or of patient preferences by long-standing custom or tradition. Such things might affect the dominant case mix of admission procedures.

Supply and availability of resources

The availability of resources inevitably affects clinical decisions. Either some decisions are prohibited because the necessary resources are not available at the right time or some rationing occurs when priorities are set for the use of available resources.

Annual health budgets within countries affect availability. For those nations where per capita outlay is minimal, almost nothing is provided to the majority of the population. Yet there are some where the average expenditure might be several hundred dollars, and still others where health resources are consumed at even higher rates (Maxwell, 1981; Organization for Economic Cooperation and Development, 1989).

The method of payment for medical services also has an impact on availability. Fee-for-service systems tend to provide high levels of availability for acute services for patients with adequate medical insurance and low levels for patients without such coverage, in particular for chronic diseases. On the other hand, prepayment systems tend to under provide services because of incentives to minimize expenditures. For these reasons, the payment method is confounded by both availability and clinical judgment. As George Bernard Shaw (1971) said in 1906, "Nobody supposes that doctors are less virtuous than judges; but a judge whose salary and reputation depended on whether the verdict was for the plaintiff or the defendant, prosecutor or prisoner, would be as little trusted as a general in the pay of the enemy."

Another important consideration is the way in which patients are admitted to the hospital. In an exclusively referral system, all decisions to hospitalize are the outcome of several screening processes. For example, in the United Kingdom patients have to decide to obtain advice from their general practitioners, who in turn have to decide to refer out their patients, at which point other decisions will have to be made about the need for hospital admission. Since the decisions made at each point in this process are constrained by different exogenous influences, the outcomes could be systematically different from those that could occur when patients seek advice directly from specialists. Specialists can be expected to be enthusiasts for their specialties and have a less detached view of the need for their services. Second opinion programs in the United States have shown that hospitalization rates are reduced when an opinion is sought from an independent consultant (McCarthy, Finkel, and Ruchlin, 1982). Where such programs are a normal part of medical care, use rates for uncertain indications could be expected to be lower; where they are not, they may well be higher. In either case, the rates are not necessarily appropriate; however, the former will cost less to current budgets.

The effect of the distribution of specialists should not be underestimated. Stevens (1977) has argued that the evolution of specialist guilds in different countries gives rise to quite different influences on the referral process. Primary care is more dominant in some countries than in others. In constrained health systems, patients may be discouraged from seeking services if they suspect that a wait or a delay is involved. In this way, the availability of resources operates as a direct restraint by precluding certain actions, but it also indirectly affects both patients' readiness to seek advice and clinical decisions about priorities. By the same token, a higher level of availability may directly encourage use by patients whose perceived benefit may be marginal, but this indirectly mitigates the need for rationing.

All of these factors contribute to observed variations; the purpose in studying variations in use rates is to understand the dominant causes and to identify fruitful areas of research and evaluation. International comparisons could invoke sensible models for incorporating data from each country to measure indices of these parameters. Unfortunately, many countries still cannot provide utilization data.

Variations in admission rates

Many studies have documented large variations among countries in hospital admission rates for surgical and medical causes. The literature on variations in utilization rates among countries was recently reviewed by Sanders, Coulter, and McPherson (1989), and useful bibliographies are provided in Sanders (1988) and Ham (1988). As shown in Table 3, variations in admission rates were evident for the populations at risk for selected

Table 3
Reported admission rates for selected procedures: Selected countries for which data were reported, 1980

Country	Procedure								
	Tonsillectomy	Coronary bypass	Cholecystectomy	Inguinal hernia repair	Exploratory laparotomy	Prostatectomy	Hysterectomy	Operation on lens	Appendectomy
Number of admissions per 1,000 population									
Australia	115	32	145	202	99	183	405	101	340
Canada	89	26	219	224	105	229	479	139	143
Denmark	229	—	21	—	—	234	255	118	248
Ireland	256	4	91	100	52	124	123	64	245
Japan	61	1	2	67	—	—	90	35	244
Netherlands	421	5	131	175	—	116	381	68	149
New Zealand	102	2	99	211	110	191	431	95	169
Norway	45	13	30	78	—	238	—	71	64
Sweden	65	—	140	206	111	48	145	—	168
Switzerland	51	—	49	116	68	—	—	22	74
United Kingdom	26	6	78	154	116	144	250	98	131
United States	205	61	203	238	41	308	557	294	130

NOTES: These figures are not age standardized and assume equal proportions of men and women. Some are likely to be incomparable for artifactual reasons.

SOURCE: Organization for Economic Cooperation and Development: Health Data File, 1989.

procedures around 1980 in those OECD countries that reported such data.

The first international study was by Pearson et al. (1968), and striking differences were noted in the frequency of operations in Liverpool, England, the Uppsala hospital region in Sweden, and the New England region of the United States. The Liverpool region discharged fewer patients than the other two regions, despite having more discharges of adults than any other hospital region in the United Kingdom. Tonsillectomy and adenoidectomy were performed more than twice as often in New England as in Liverpool and four times as often as in Uppsala. Although Uppsala and Liverpool had similar surgical rates, Uppsala had significantly more gallbladder and gynecological operations than Liverpool.

Comparing operations and surgeons in the United States with those in the United Kingdom, Bunker (1970) found that there were twice as many surgeons in proportion to the population in the United States as in the United Kingdom and that the population underwent twice as many operations. Comparing specific operations, Bunker reported that tonsillectomy and adenoidectomy operations as well as hernia repair operations were performed almost twice as often in the United States, and cholecystectomy operations were performed almost three times as often.

Vayda (1973) compared surgical rates in Canada with those in the United Kingdom and standardized the rates for age and sex. Comparisons showed that surgical rates in Canada in 1968 were 1.8 times greater for men and 1.6 times greater for women than in the United Kingdom. The standardized rates for diverse elective and discretionary operations such as tonsillectomy, hemorrhoidectomy, and hernia repair operations were two or more times higher in Canada than in the United Kingdom.

Both Bunker and Vayda commented that the disparity resulted from sources other than the incidence and prevalence of disease, relating it particularly to the supply of services and number of surgeons. Subsequently, Vayda, Mindell, and Rutkow (1982) compared surgical rates in Canada, the United Kingdom, and the United States between 1966 and 1976, and again reported that overall surgical rates in the United States were twice those of the United Kingdom, while the rates for Canada were 1½ times the rates for the United Kingdom. Kohn and White (1976) examined hospital utilization rates in 12 areas of 7 countries and found that standardized hospitalization rates varied more than two-fold among areas. The availability of short-term beds was found to account for most of the variation.

McPherson et al. (1981) studied the use of common surgical procedures among the United Kingdom, Canada, and the United States, and reported that rates of surgical utilization, standardized by age and sex, varied up to seven-fold among the countries. Appendectomy was the only operation carried out at similar levels in these countries. Large variations were also found among regions of the countries. Although supply and cultural variables might account for the international differences, the variation within countries was only somewhat attributable to indices of supply, but much variation remained unexplained.

As an example of the magnitude of differences in cross-national rates for a specific procedure in the mid-1970s, McPherson (1988) reported an age-standardized rate for hysterectomy of 700 per 100,000 in the United States, approximately 600 in Canada, 450 in Australia, 250 in the United Kingdom, and 110 in Norway. Coulter, McPherson, and Vessey (1988) reported more recent rates for hysterectomy of 130 per 100,000 in Sweden, in contrast to 360 in neighboring Denmark. Several possible explanations

for these differences have been discussed at length in McPherson et al. (1981). The method of payment, supply of resources, availability of manpower, and reimbursement and referral patterns may all play a part. The definitive causes of these differences remain, in most cases, unknown, and the outcome differences associated with these variations as well as whether the benefits are commensurate with the costs also remain unresolved without explicit further study.

The role that demand for medical and surgical services might play in the observed variations in hospitalization rates has received considerable attention. Bunker and Brown (1974) demonstrated that the wives of men in different professions did have significantly different rates for hysterectomy and for several other discretionary surgical procedures. Of special interest was the observation that the wives of physicians reported operation rates as high, or higher, than those for the other professional groups. Whether this was demand-led or a manifestation of more available (and less expensive) supply is difficult to tell. Bloor (1976), in an extensive study of childhood tonsillectomy, has failed to discern a demand component in the decisionmaking process, and a study by Coulter and McPherson (1985) found little social class difference in the probability of discretionary surgery in the Oxford region of the United Kingdom, or any support for the notion of an effect of differential consumer demand.

Bridgman (1979) presented an international study on hospital utilization in nine regions of eight countries. One of the significant outcomes of the study was to show the correlation between the pattern of hospital utilization and the level of socioeconomic development of the countries. The rates of hysterectomy were much lower in Norway than in either Denmark or the State of Massachusetts in the United States (Anderson and Kamper-Jorgensen, 1984); the strikingly low rate of hysterectomy in Norway was noted by McPherson et al. (1982). Women in Italy were much more likely to have a hysterectomy than those in France (Van Keep, Wildemeersh, and Leher, 1983).

Caesarean section rates similarly show large variations among countries. National rates varied four-fold from less than 5 percent of all deliveries in the Netherlands and Fiji to nearly 20 percent in Singapore, Canada, and the United States (Chalmers, 1985). Notzon, Placek, and Taffel (1987) studied caesarean rates in 19 industrialized countries of Europe, North America, and the Pacific, and there were sharp differences in rates per 100 hospital deliveries in 1981, ranging from a low of 5 in Czechoslovakia to a high of 18 in the United States. The percentage of mothers who had a vaginal birth after a previous caesarean section was only 5 in the United States, compared with 43 in Norway. Women in the United States had a significantly higher rate of caesarean section for dystocia or abnormal labor than women in Ireland (Sheehan, 1987).

Ulcer disease accounted for 35 percent more bed-days per 100,000 population in Denmark than in

Sweden, with the main source of the difference being accounted for by admissions for duodenal ulcer (Joensson and Silverberg, 1982). The higher consumption of hospital care in Denmark was largely explained by the fact that more medical cases were treated as in-patients in Denmark than in Sweden. The appendectomy rate in the Federal Republic of Germany was two to three times higher than that of other comparable European countries (Lichtner and Pflanz, 1971).

Plant et al. (1973) compared the number of gallbladder operations in 1961 and 1971 in three similar towns in Canada, France, and the United Kingdom and concluded that the incidence of gallbladder disease was six times higher in North America than in Western Europe, because of the much higher rate of cholecystectomy in the United States. However, evidence from morbidity surveys using necropsy studies of the prevalence of gallstones indicated a lower, rather than higher, prevalence in North America (McPherson et al., 1985).

Within the United Kingdom, morbidity differences did correlate positively with cholecystectomy rates in a study of six English towns (Barker et al., 1979). It is interesting to note that the original studies of international variations showed Canada with a higher rate than anywhere else, and this high rate was most marked among middle-aged females (McPherson, 1988). As shown in the OECD data compendium section of this issue, cholecystectomy rates among females in Canada are approaching the rates in the United States. In the early 1970s, however, the Canadian rate was twice as high as the American rate. This would strongly suggest that, at that point in time, 50 percent of the women receiving cholecystectomies in Canada would not have received them in the United States. By the same token, based on the latest OECD data, two-thirds of the women receiving cholecystectomies in the United States (where the rate is 200 per 100,000 population) might not receive the operation in the United Kingdom (where the rate is 68 per 100,000 population).

In 1982, the number of cardiac operations in the United Kingdom, with 107 operations per one million population, was significantly lower than in other countries such as Australia, with 410 operations per one million, or the United States, with 750 operations per one million (English et al., 1984). Japan has a corresponding rate of around 10 operations per one million population, which must reflect the low incidence of coronary heart disease in that country (Tunstall-Pedoe, Smith, and Crombie, 1986). Unfortunately, other countries with apparently low coronary heart disease incidence or mortality, such as France, Spain, Greece, and Switzerland, do not yet produce procedure rates.

In the book, *The Painful Prescription: Rationing Hospital Care*, Aaron and Schwartz (1984) provide an illuminating analysis of the provision of 10 medical procedures in the United States and the United Kingdom. Most services were provided at

lower levels in the United Kingdom than in the United States: for example, the rate of coronary artery bypass surgery in the United Kingdom was only 10 percent of that in the United States. The overall rate of treatment for chronic renal failure in the United Kingdom was less than half that in the United States. The low rate of treatment for patients with kidney failure in the United Kingdom compared with other countries has also been commented on by Wing (1983). Three procedures—bone marrow transplants, radiotherapy for cancer, and treatment for hemophilia—were provided at essentially the same levels in both countries.

Issues for the future

Just as it is difficult to provide mass screening services without clear evidence of the benefits received, it is also difficult to provide appropriate health services without clear evidence of the relative efficacy of various treatment approaches. In the face of uncertainty, however, therapy, in contrast to screening, is difficult to withhold. Therefore, health services are provided, and implicitly rationed, but the variations shown in the OECD data compendium would indicate that all are not provided rationally. Since the information base for judging appropriateness is often inadequate, research protocols must efficiently address the important uncertainties.

Rational rationing

In comparing rates among nations, low rates are often taken implicitly to be a manifestation of under-supply. In contrast, since feedback on high rates usually results in rate reduction, high rates are taken as an indication of over-supply. These are examples of glib assumptions about the nature of the relationship between process and outcome (Donabedian, 1966), without empirical validation.

Until recently, almost all evidence of under-supply was considered detrimental to the health of the community, and indeed, in some circles (Lowry, 1988), it still is. Low rates reflected the inefficient use of resources, low productivity, and unmet need, whereas high rates reflected too much specialist enthusiasm, over-provision of resources, and unnecessary intervention. Such crude analysis pushes countries towards an average which has no rationale. It pretends that the appropriate intervention rate is known, when it is not, and indeed assumes that there is such a thing.

The policy implications of observed variations depend on knowledge about their causes and consequent outcomes, not on the magnitude of the variation alone, and certainly not on an average. If variations in rates occur because of legitimate reasons, then the policy implications are negligible. Aspects of culture and demand, when commensurate with explicit social policy and budgets, can give rise to large

international variations in use rates that are wholly unproblematic.

Procedures with low variation

When combined with knowledge about small area variation and the epidemiology of relevant diseases, observed differences among countries can be interpreted with greater precision. Large differences among countries for a procedure that is relatively invariant among small areas might well point to differences in morbidity as the first, most plausible, explanation. If the variation observed is out of proportion to feasible morbidity differences, then the influence of culture, education, or availability on clinical decisions would be the next most obvious explanation.

Examples of such procedures are cholecystectomy and appendectomy. Although these procedures vary little among neighboring small areas within countries, there appear to be four- to five-fold variations among the countries themselves. The search for systematic differences in morbidity rates has proved inadequate to justify these international differences, as has the search for artifactual explanations. The evidence suggests that there may be strong national consensus on the nature of what constitutes sufficient indications, but that this consensus appears to be quite different in different countries, and in some it may be changing.

Presumably, the within-country consensus represents a view or teaching about the correct use of an operation. As such, it is not explicitly recognized as being uncertain within a country. However, by inspecting international differences, it becomes quite clear that wholly different policies are invoked in different countries for sufficient indications to perform an operation. On the one hand, aspects of prudent prophylaxis are possibly being used where, on the other hand, only the most urgent need is being admitted. Such differences have enormous financial implications for any health sector, particularly when multiplied by other similar procedures.

The next step in evaluation is assiduous decision analysis based on the most reliable data in the literature and data from longitudinal studies (e.g., Wasson et al., 1985). For this, one requires estimates of outcome associated with various treatment options. From the literature, this would come from randomized trials and case series. Information from data bases should come from several countries where the rates are different and where the evidence for an artifactual or morbidity explanation is lacking. The application of other health information to approximate randomized comparison is possible (McPherson and Bunker, 1989). Claims data have been successfully used (Wennberg et al., 1987), as have record-linkage studies (Roos et al., 1989). It is possible to compare mortality and readmission rates at various followup times between treatment options, and such comparisons may not be seriously confounded by unmeasured indices of prognosis.

If one country has a rate twice that of another for a procedure, and if this can be attributed to practice style, it is reasonable to assume that 50 percent of the most seriously ill patients in the latter country will be receiving the procedure. In some circumstances, it might be possible to identify, from primary health care consultations, the remaining 50 percent of patients who are not admitted. It could then be argued that the combination of these two patient groups represents a cohort that is comparable to the treated cohort in the high-rate country. The advantage to using these alternative techniques is that randomized controlled comparisons could be by-passed in circumstances where they are unethical, and hence impossible, because the clinicians have few doubts. To identify all patients presenting with symptoms but not being recommended for surgery might well be to identify a cohort that could legitimately comprise a control group in a randomized study. The comparison would have to be between all those treated in the high-rate area with all those identified (both treated and not) in the low-rate area that had comparable symptoms, in essence, a comparison between a policy of intervention and a policy of conservative management.

The advantages of such studies are legion. The actual practice variations among countries can be quite enormous and, since these are often for common interventions, the sample sizes obtainable from such natural experiments can be huge. Most importantly, comparisons can be made between different treatment policies that already exist, ones that could not otherwise be duplicated in an ethical environment for the sake of experimentation. Large differences in management policy can be compared where there may be few case-mix comparability problems (Coronary Drug Project Research Group, 1980). With such potentially large data bases, it may be possible to examine the effect of policy among subgroups and even for long followup periods.

Procedures with high variation

Some discretionary procedures, such as tonsillectomy, prostatectomy, and hysterectomy, vary a great deal among small areas. Hysterectomy is variable both within and among countries. This must reflect massive uncertainty about the appropriate indication for this operation. There may well be few "correct" indications, and each decision may be an individual matter concerned with finely balanced assessments of anticipated benefits and losses (Coulter, McPherson, and Vessey, 1988). If this is the case, then such decisions ought to be made with an eye on the foregone opportunities associated with each marginal hysterectomy. It is difficult to justify marginal operations when there are any genuine unmet needs elsewhere in the health sector, which there are most likely to be. No epidemiological evidence suggests that hysterectomy rates are, to any important extent, determined by demand from

consumers. However, many opportunities exist for outcome studies and for the types of studies previously discussed to evaluate the relative benefit of interventionist versus conservative policies and to examine the detailed determinants of these varying rates.

In almost all cases for which we have data, highly variable procedures among small areas are also highly variable among countries. Moreover, countries with fee-for-service systems and/or high expenditures tend to have the higher rates (Schieber and Poullier, 1988). Health care provision which, at the margin, may provide relatively little net benefit may be provided because, in unconstrained systems where uncertainty exists, it is tempting to over-emphasize the benefit and under-emphasize the cost. It is precisely for these types of procedures that guidelines and publicly discussed policies are required so that health care budgets can be monitored in a way that is consistent with national policy. Countries could then decide which operations with marginal indications are more important, and then might end up with higher rates than in other countries as a matter of public policy.

At the moment, these policies happen in ignorance of their own determinants. They are formulated without knowing outcome, because longitudinal studies are not done, and without knowing the rates in other countries, because the data are incomplete. It is important to collect these data because many hypotheses about the determinants of health care practice, such as cultural considerations, availability of manpower and resources, and method of payment could be tested with complete data. In particular, the role of method of payment on health care practice remains impossible to disentangle from other systematic differences among countries. If these data were made available, observational information to complement the RAND Corporation's (Brook et al., 1983) randomized studies in a single country would be extremely useful.

Such policies also happen in the face of observational evidence which refuses to show a sensible association between the amount of health care provided and outcome (Cochrane, Leger, and Moore, 1978). They happen largely in ignorance of patient preferences as well, because people are unaware that such international and intranational variations exist and are, therefore, apt to view their physicians' decisions to hospitalize them as above their own preferences and inclinations. In many circumstances, patient preferences should be the ultimate determinants of these decisions (Barry et al., 1988). However, patient participation requires that information about health care practices be broadened and deepened and that the knowledge that is gained about what determines different practice styles be widely disseminated. Such knowledge can also help determine research priorities. Now that the need for answers is becoming so critical, the difficulties inherent in assessing these things can no longer be allowed too much dominance (Ellwood, 1988).

Acknowledgments

The author is grateful to colleagues and friends, particularly Angela Coulter and Diana Sanders for ideas and stimulation. A long-standing collaboration with Professors Jack Wennberg and John Bunker has been essential and also extremely fruitful. Professors Martin Vessey and, lately, David Skegg oversee a convivial working environment with lots to do and talk about. Finally, Mrs. Isobel Pairman has greatly helped with data reduction and preparation.

References

- Aaron, H. J., and Schwartz, W. B.: *The Painful Prescription: Rationing Hospital Care*. Washington, D.C. The Brookings Institute, 1984.
- American Medical Association: *Confronting Regional Variations: The Maine Approach*. Chicago, Ill. 1986.
- Anderson, T. F., and Kamper-Jorgensen, F.: Variation in surgical rates and the organisation of health care. *Institute of Social Medicine*. Denmark. University of Copenhagen, 1984.
- Barker, D. J. P., Gardner, M. J., Power, C., and Hutt, M. S. R.: Prevalence of gallstones at necropsy in nine British towns. *British Medical Journal* 2:1389-1392, 1979.
- Barry, M. J., Mulley, A. G., Fowler, F. J., et al.: Watchful waiting vs. immediate transurethral resection for symptomatic prostatism: The importance of patients' preferences. *Journal of the American Medical Association* 259(20):3010-3017, 1988.
- Beecher, H. K., and Boston, M. D.: Surgery as a placebo. *Journal of the American Medical Association* 231:1102-1107, 1961.
- Bloor, M.: Bishop Berkeley and the adenotonsillectomy enigma: An exploration of variation in the social construction of medical disposals. *Sociology* 10:44-61, 1976.
- Bridgman, R. F.: *Hospital Utilisation: An International Study*. Oxford, England. Oxford University Press, 1979.
- Brook, R. H., Ware, J. E., Rogers, W. H., et al.: Does free care improve adults' health: Results from a randomized controlled trial. *New England Journal of Medicine* 309(23):1426-1434, 1983.
- Bunker, J. P.: Surgical manpower: A comparison of operations and surgeons in the United States and in England and Wales. *New England Journal of Medicine* 282:135-144, 1970.
- Bunker, J. P., Barnes, B., and Mosteller, F.: *Costs, Risks and Benefits of Surgery*. New York. Oxford University Press, 1977.
- Bunker, J. P., and Brown, B.: The physician as an informed consumer of surgical services. *New England Journal of Medicine* 290:1051-1055, 1974.
- Bunker, J. P., and Fowles, J.: Between the laboratory and the patient. *Nature* 298: 405-406, 1982.
- Chalmers, I.: Trends and variations in the use of caesarean section. In Clinch, J., and Mathews, T., eds. *Perinatal Medicine*. Lancaster, England. MTP Press, 1985.
- Cochrane, A. L.: *Effectiveness and Efficiency: Random Reflections on Health Services*. London. Nuffield Provincial Hospital Trust, 1971.
- Cochrane, A. L., Leger, A. S., and Moore, F.: Health service 'input' and mortality 'output' in developed countries. *Journal of Epidemiologic Community Health* 32:200-205, 1978.
- Comroe, J. H., and Dripps, R. D.: *The Top Ten Clinical Advances in Cardiopulmonary Medicine and Surgery, 1945-1975*. DHEW Pub. No. 78-1521. Department of Health, Education, and Welfare. Washington. U.S. Government Printing Office, 1977.
- Coronary Drug Project Research Group: Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *New England Journal of Medicine* 303:1038-1041, 1980.
- Coulter, A., and McPherson, K.: Socioeconomic variations in the use of common surgical operations. *British Medical Journal* 291:183-187, 1985.
- Coulter, A., McPherson, K., and Vessey, M. P.: Do British women undergo too many or too few hysterectomies? *Social Science and Medicine* 27(9):987-994, 1988.
- Doll, R., and Peto, R.: *The Causes of Cancer*. Oxford, England. Oxford University Press, 1981.
- Donabedian, A.: Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly* Suppl. 44:166-178, 1966.
- Dyck, F. J., Murphy, F. A., Murphy, J. K., et al.: Effects of surveillance on the number of hysterectomies in the province of Saskatchewan. *New England Journal of Medicine* 296:1326-1329, 1977.
- Ellwood, P. M.: Shattuck lecture. Outcome management: A technology of patient experience. *New England Journal of Medicine* 318(23):1549-1556, 1988.
- English, T. A. H., Bailey, A. R., Dark, J. F., et al.: The UK cardiac surgical register, 1977-82. *British Medical Journal* 298:1205-1208, 1984.
- Fineberg, H. V.: Effects of clinical evaluation on the diffusion of medical technology. Chap. 4. *Assessing Medical Technologies*. Committee for Evaluating Medical Technologies in Clinical Use, Institute of Medicine. Washington, D.C. National Associated Press, 1985.
- Friedson, E.: *Profession of Medicine*. New York. Dodd-Mead, 1972.
- Gittelsohn, A., and Wennberg, J. E.: On the risk of organ loss. *Journal of Chronic Diseases*. 29:527-535, 1976.
- Glover, J. A.: The incidence of tonsillectomy in school children. *Proceedings of the Royal Society of Medicine* 31:1219-1236, 1938.
- Ham, C.: *Health Care Variations: Assessing the Evidence*. London. Kings Fund Institute, 1988.
- Hill, A. B.: *Statistical Methods in Clinical and Preventive Medicine*. Edinburgh, Scotland. Livingston, 1962.
- Ingelfinger, F. J.: Arrogance. *New England Journal of Medicine* 303:1507-1511, 1980.
- Joensson, B., and Silverberg, R.: Variations between and within countries in hospital care for peptic ulcer: A comparison between Denmark and Sweden. *Scandinavian Journal of Social Medicine* 10:63-70, 1982.
- Kohn, P., and White, K. L.: *An International Study*. Oxford, England. Oxford University Press, 1976.
- Lembcke, P. A.: Medical audit by scientific methods: Illustrated by major female pelvic surgery. *Journal of the American Medical Association* 162:646-651, 1956.
- Lichtner, S., and Pflanz, M.: Appendectomy in the Federal Republic of Germany: Epidemiology and medical care patterns. *Medical Care* 9:311-317, 1971.

- Llewellyn-Thomas, H., Sutherland, H. J., Tibshirani, R., et al.: Methodological issues in obtaining values for health states. *Medical Care* 24:113-118, 1984.
- Lowry, S.: Focus on performance indicators. *British Medical Journal* 296:992-994, 1988.
- Maxwell, R. A.: *Health and Wealth. An International Study of Health Care Spending*. Lexington, Mass. Lexington Books, 1981.
- McCarthy, E. G., Finkel, M. L., and Ruchlin, H. S.: Current status of surgical second opinion programs. *Surgical Clinics of North America* 62(4):705-719, 1982.
- McNeil, B. L., Weichselbaum, R., and Pauker, S. G.: Fallacy of five-year survival in lung cancer. *New England Journal of Medicine* 299:1397-1400, 1978.
- McPherson, K.: *Variations in Hospitalization Rates: Why and How to Study Them*. London. Kings Fund Institute, 1988.
- McPherson, K., and Bunker, J. P.: Application of health information to health services. In Holland, W. W., Knox, G., and Detels, R., eds. *Oxford Text Book of Public Health*. Oxford, England. Oxford University Press, 1989.
- McPherson, K., Strong, P. M., Epstein, A., et al.: Regional variation in the use of common surgical procedures within and between England and Wales, Canada and the United States of America. *Social Science Medicine* 15A:273-288, 1981.
- McPherson, K., Strong, P. M., Jones, L., et al.: Do cholecystectomy rates correlate with geographic variations in the prevalence of gallstones? *Journal of Epidemiologic and Community Health* 39:179-182, 1985.
- McPherson, K., Wennberg, J. W., Hovind, O., et al.: Small area variations in the use of common surgical procedures: An international comparison of New England, England and Norway. *New England Journal of Medicine* 307:1310-1314, 1982.
- Muir, C. S.: Classification. In Doll, R., Muir, C., and Waterhouse, J., eds. *Cancer in Five Continents*. 3rd Edition. International Agency for Research in Cancer, 1976.
- Mulley, A. G.: Assessing patients' utilities: Can the ends justify the means? *Medical Care* 27(3):269-281, 1989.
- Nicholl, J. P., Beeby, N. R., and Williams, B. T.: Role of the private sector in elective surgery in England and Wales. *British Medical Journal* 298:243-247, 1988.
- Notzon, P. A., Placek, P. J., and Taffel, S. M.: Comparisons of national cesarian section rates. *New England Journal of Medicine* 316(7):386-389, 1987.
- Organization for Economic Cooperation and Development: Health Data File. Paris. Directorate for Social Affairs, Manpower, and Education, 1989.
- Pearson, R. J. C., Smedby, B., Berfenstam, R., et al.: Hospital case loads in Liverpool, New England and Uppsala: An international comparison. *Lancet* 2:559-566, 1968.
- Plant, J. C. D., Percy, I., Bates, T., et al.: Incidence of gallbladder disease in Canada, England and France. *Lancet* 2:249-251, 1973.
- Poullier, J. P.: *Measuring Health Care, 1960-1983. Expenditures, Costs and Performance*. Social Policy Studies No. 2. Paris. Organization for Economic Cooperation and Development, 1985.
- Roos, N. P., Wennberg, J. E., Malenka, D. J., et al.: Mortality and reoperation following open and transurethral resection of the prostate for benign prostatic hyperplasia. *New England Journal of Medicine* 320(17):1120-1124, Apr. 27, 1989.
- Roos, N. P., Wennberg, J. E., and McPherson, K.: Using diagnosis-related groups for studying variations in hospital admissions. *Health Care Financing Review*. Vol. 9, No. 4. HCFA Pub. No. 03265. Office of Research and Demonstrations, Health Care Financing Administration, Washington. U.S. Government Printing Office, July 1988.
- Sacket, D. L., and Torrance, G. W.: The utility of different disease states as perceived by the general public. *Journal of Chronic Disease* 7:347-351, 1978.
- Sanders, D.: *Bibliography on Geographic Variations in Hospital Admission Rates*. Department of Community Medicine. Oxford, England. Radcliffe Infirmary, 1988.
- Sanders, D., Coulter, A., and McPherson, K.: *Variations in Hospital Admission Rates: A Review of the Literature*. London. Kings Fund Institute, 1989.
- Schieber, G. J.: *Financing and Delivering Health Care. A Comparative Analysis of OECD Countries*. Social Policy Studies No 4. Paris. Organization for Economic Cooperation and Development, 1987.
- Schieber, G. J., and Poullier, J. P.: International health spending and utilization trends. *Health Affairs* 7(4):105-122, Fall 1988.
- Shaw, G. B.: Preface on doctors. *The Doctor's Dilemma*. London. The Bodley Head, 1971.
- Sheehan, K. H.: Caesarian section for dystocia: A comparison of practices in two countries. *Lancet* 1:548-51, 1987.
- Stevens, R.: The evolution of health care systems in the United States and the United Kingdom: Similarities and differences. *Priorities for the Use of Resources in Medicine*. Fogarty International Centre Proceedings. Washington, D.C. U.S. Department of Health, Education, and Welfare, 1977.
- Tunstall-Pedoe, H., Smith, W. C. S., and Crombie, I. K.: Level and trends of coronary heart disease mortality in Scotland compared with other countries. Scottish Home and Health Department. *Health Bulletin* 44(3):153-61, 1986.
- Van Keep, P. A., Wildemeersh, D., and Leher, P.: Hysterectomy in six European countries. *Maturitas* 5:69-75, 1983.
- Vayda, E.: A comparison of surgical rates in Canada and in England and Wales. *New England Journal of Medicine* 289:1224-1229, 1973.
- Vayda, E., Mindell, W. R., and Rutkow, I. M.: A decade of surgery in Canada, England and Wales and the United States. *Archives of Surgery* 117:846-853, 1982.
- Wasson, H. H., Sox, H. C., Neff, R. K., et al.: Clinical prediction rules: Application and methodological standards. *New England Journal of Medicine* 313:793-799, 1985.
- Wennberg, J. E., and Gittelsohn, A.: Small area variations in health care delivery: A population based health information system can guide planning and regulatory decision making. *Science* 182:1102-1109, 1973.

Wennberg, J. E., and Gittelsohn, A.: Variation in medical care among small areas. *Scientific American* 246:100-112, 1982.

Wennberg, J. E., McPherson, K., and Caper, P.: Will payment based on diagnosis-related groups control hospital costs? *New England Journal of Medicine* 311(5):295-300, Aug. 1984.

Wennberg, J. E., Roos, N. P., Sola, L., et al.: Use of claims data systems to evaluate health care outcomes: Mortality and reoperation following prostatectomy. *Journal of the American Medical Association* 257:933-936, 1987.

Wing, A. J.: Why don't the British treat more patients with kidney failure? *British Medical Journal* 287:1157-1158, 1983.