

GEOGRAPHICAL VARIATIONS IN THE USE OF HEALTH SERVICES:

A Review of the Literature

Not for Quotation

Paper prepared for working conference at the  
King's Fund Centre on 12 June 1987

Chris Ham

King's Fund Institute

126 Albert Street

LONDON NW1 7NF

Tel: 01-485 9589

HPW

## Introduction

The literature on geographical variations in health care is large and growing. It is now almost fifty years since Glover (1938) reported wide variations in the rate at which tonsils were removed in different parts of England. In the intervening period, studies of health care variations have multiplied to the point where a recent bibliography listed 153 references concerned with regional variations in the provision, utilisation and outcomes of health care (Copenhagen Collaborating Centre, 1985). A comprehensive review of this literature would be a major task and is beyond the scope of this paper. Instead, the paper summarises some of the key issues which emerge from the literature and identifies the implications for policy and priorities for research.

## Background

Although research on geographical variations in health now spans half a century, it is only in the last ten years that a major effort has been made to describe and explain these variations. This is illustrated in Table 1 which shows the year of publication of the references listed in the bibliography prepared by the Copenhagen Collaborating Centre for the Study of Regional Variations in Health Care. Analysis of the publications identified by the Copenhagen Collaborating Centre indicates that research effort has been led by a relatively small number of researchers, most notably Jack Wennberg in the United States, Noralou and Les Roos in Canada and Klim McPherson in England. It is principally these researchers whose work will be drawn on in this paper, although other key studies will also be cited where appropriate.

TABLE 1

## YEAR OF PUBLICATION OF CITATIONS IN THE CCC BIBLIOGRAPHY

1938	1	1975	8
1952	1	1976	8
1954	1	1977	12
1957	1	1978	6
1961	1	1979	7
1968	4	1980	9
1969	3	1981	11
1970	1	1982	15
1971	2	1983	14
1973	5	1984	33
1974	2	1985	8

As already noted, one of the seminal papers on health care variations examined the rate at which tonsils were removed in different parts of England. Study of variations in the *use* of services as measured by the number of procedures carried out for a given population is the main theme of much of the later research that has been conducted, and it is principally this research which is examined here. However, analysis has been extended in some cases to include variations in the *provision or supply* of services (beds, doctors etc) and variations in the *efficiency* with which services are provided (length of stay, costs per case). Furthermore, researchers have used a range of population units in studying variations. These include small areas within one country, large areas

within one country, small and large areas between countries, and countries as a whole. In examining variations at different levels of aggregation, a number of researchers have turned their attention to the practice style of individual doctors as a possible source of the variations that exist. This has given rise to an important minor theme in the literature, namely variations between clinicians in their propensity to provide services.

Before reviewing the literature in more detail, it is important to ask why the issue of variations in use rates has received so much attention.

Three factors appear to be important. First, there is the longstanding concern with the *effectiveness* of medical care. Second, there is the associated interest in the *efficiency* of health care delivery. And third, there is the debate about *equity* in the provision and use of services.

The variations literature contributes to analysis and discussion in each of these fields by raising questions such as:

- Does the high rate of use of services in certain areas indicate unnecessary or inappropriate use?
- Can the higher costs associated with high rates of use be justified in terms of improved health outcomes?
- Are differences in use related to the need for care of the population concerned?

As will become apparent, it is possible to offer some answers to these questions, but much work remains to be done. Although considerable progress has been made in *describing* variations, attempts at explanation have so far proved inconclusive or at least not susceptible to clear cut conclusions. Research on causality has focussed on demand-side factors

such as population characteristics and morbidity, supply-side factors such as the provision of doctors and beds, and on the practice style of clinicians. A number of studies have speculated on the importance of these factors; others have used multivariate analysis and related forms of statistical investigation to test their significance. Despite the growth of interest in variations studies and the increasing sophistication of the methodologies used, there is a continuing debate about the relative importance of different variables. Much work points to the importance of the supply of services and practice style, but there remains a good deal of uncertainty. In the absence of convincing explanations, one of the main contributions which analysis can make in this field is to identify key issues for debate and action by policy makers, the medical profession and researchers. We discuss this more fully in the final part of the paper.

#### International Variations

A number of the early studies of health care variations were concerned with international differences in the provision and use of services. A much quoted example is the analysis by Pearson and others of hospital caseloads in three regions of England, Sweden and the United States (Pearson *et al*, 1968). One of the most important findings of the study was the existence of striking differences in the frequency of individual operations between the regions: Liverpool, Uppsala and New England. These differences are illustrated in Table 2.

TABLE 2

-DISCHARGE-RATES PER 10,000 POPULATION FOR SELECTED COMMON OPERATIONS, BY SEX							
Operation	H.I.P.E. codes	Males			Females		
		Liver- pool	Upp- sala	New Eng- land	Liver- pool	Upp- sala	New Eng- land
Tonsillectomy and/or adenoidectomy ..	261-3	29	15	68	23	19	72
Inguinal herniorrhaphy ..	402	26	35	49	3	7	5
Appendicectomy ..	441	23	31	20	26	27	14
Cholecystectomy ..	521	3	21	10	7	49	27
For peptic ulcer*	422-3	10	9	7	3	3	3
	427, 431						
	433-4						
D. & C. ..	732	..	..	..	25	76	65
Total abdominal hysterectomy ..	722	..	..	..	14	6	28
For prolapse*	724, ..	..	..	..	14	7	11
	743-4						
Mastectomy ..	381-3	..	..	..	10	8	25
Prostatectomy ..	672-7	7	14	18	..	..	..
Extraction of lens ..	173	3	5	9	5	5	10

\* Corrected to exclude operations for other conditions.

Source: Pearson et al. (1968)

To give some examples, the table shows that tonsillectomy and adenoidectomy was performed twice as often in New England as in Liverpool and four times as often in Uppsala; inguinal herniorrhaphy was performed twice as often in New England as in Liverpool with Uppsala in an intermediate position; and cholecystectomy was performed seven times as often in Uppsala as in Liverpool, with New England in an intermediate position. The authors also noted that mean hospital stays were considerably higher in Liverpool than in the other two regions. The study concluded: "inter-regional differences are real, large and important; they are found in most of the common operations. Some of the differences may be related to variations in incidence of a condition, but many are

more likely to be caused by differences in the systems of medical care" (Pearson et al, 1968, p.563).

This conclusion was supported by Bunker's comparison of surgical services in the United States and England and Wales (Bunker, 1970). Bunker found that there were twice as many surgeons in proportion to the population in the United States as in England and Wales, and they performed twice as many operations. Comparing specific operations, Bunker reported that tonsillectomy and adenoidectomy were performed almost twice as often in the United States, cholecystectomy was performed almost three times as often, and inguinal herniorrhaphy was performed almost twice as often. Bunker argued that variations on this scale could not be accounted for entirely by differences in morbidity. Rather, he maintained that the existence of different methods of organising and financing services, the more aggressive surgical philosophy of the United States, and the uncertainty surrounding appropriate indications for surgery, created a climate in which surgeons in the United States operated more frequently. Although Bunker was careful not to conclude that the United States was providing twice as much surgery as was necessary, his analysis led him to argue that until new evidence was provided "it is reasonable to assume that there is a disproportionate number of surgeons in the United States and it seems likely that some unnecessary surgery is being performed" (Bunker, 1970, p.143).

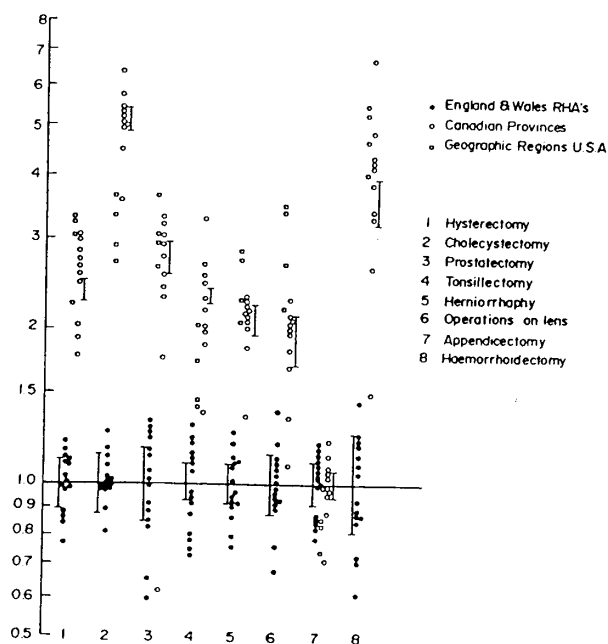
A further study along similar lines was published by Vayda in 1973. Vayda compared surgical rates in Canada and England and Wales, but unlike Bunker he standardised his data for the age of the population. Overall, Vayda found that surgical rates in Canada were 1.8 times greater for men and 1.6

times greater for women than in England and Wales. The age standardised and sex specific rates for particular operations were two or more times higher in Canada than in England and Wales. In seeking to explain these variations, Vayda argued that the key factors were the more conservative treatment styles in England and Wales, the greater availability of surgeons and beds in Canada, and the impact of financial incentives to operate in Canada. Differences in disease prevalence, as measured by mortality rates, were not found to be important. Vayda concluded that it was difficult to establish whether surgical rates were too high in Canada or too low in England and Wales. Accordingly, he called for further work through controlled trials to establish the benefits of surgical and non-surgical treatment for common diseases, and he suggested that the medical profession should initiate or expand audit programmes to establish appropriate indications for surgery (Vayda, 1973).

The study of international variations was taken a stage further by McPherson and colleagues in an analysis of variations in the use of common surgical procedures within and between England and Wales, Canada and the United States (McPherson et al, 1981). This study reported that rates of surgical utilisation standardised by age and sex varied by as much as twofold within England and Wales, as much as fivefold between Canadian provinces, and up to sevenfold internationally. The results are displayed in the accompanying table.



TABLE 3



Age and sex standardized ratio of operation rates to all England and Wales 1975 (logarithmic scale).

Source: McPherson et al. (1981)

McPherson et al noted that a major problem in interpreting these data was "the lack of any comparative morbidity rates, whose variation could, in principal, explain all the observed variation in the rates for the operations we have described here" (p.280). A further problem was the absence of an agreed set of indications for surgery which would enable a

correct rate to be established. Notwithstanding these difficulties, on the basis of a series of multiple regression analyses, McPherson et al concluded that variations in surgical rates could probably best be accounted for by supply variables, in particular the number of surgeons available. However, this applied only when comparing England and Wales with North America, and it did not hold within England and Wales. Like Vayda and Bunker, McPherson et al argued that the North American fee-for-service system provided an incentive to surgeons to operate. This incentive was lacking in England and Wales where surgeons were constrained by fixed budgets and the availability of beds.

In a further paper, McPherson et al studied variations in the use of seven common surgical procedures in seven areas in Southern Norway, 21 districts in the West Midlands and 18 areas in New England (McPherson et al, 1982). The results confirmed the findings of earlier studies. For all procedures except for appendicectomy, international comparisons showed a twofold or greater difference between at least two of the three countries for each procedure, and surgical rates in New England were consistently higher than in Norway and the West Midlands. However, the more original and important contribution of this study was the finding that variations in each country followed a characteristic pattern. Independently of the method of organising and financing health care the extent of variation was similar. This is shown in Table 4 which demonstrates that some procedures, such as tonsillectomy, had a highly variable rate of use, whereas other procedures, such as appendicectomy, exhibited much less variation. In general, the degree of variation appeared to be more characteristic of the procedure than of the country in which it was performed. The significance of this conclusion was that it suggested that differences in methods of

organising and financing health care were less important in explaining the degree of variation than controversy and uncertainty concerning the indications for a procedure. We return to this theme below.

TABLE 4

Indexes of Variation in Age- and Sex-Standardized Surgical Rates among Selected Hospital Services in New England, Norway, and the West Midlands.

COEFFICIENT OF VARIATION (%)	HERNIA REPAIR	APPEN- DECTOMY	CHOLECYS- TECTOMY	PROSTA- TECTOMY	HYSTER- ECTOMY	HEMORRHOID- ECTOMY	TONSIL- ECTOMY	ALL SEVEN PROCEDURES
New England	0.11	26	18	30	22	30	36	14
Norway	0.20	16	18	33	31	47	48	11
West Midlands	0.20	16	16	24	20	35	31	12
RANGE (HIGH/LOW)								
New England	1.7	2.3	1.9	2.2	2.2	4.8	4.2	1.69
Norway	1.3	1.6	1.5	2.2	3.0	2.9	4.7	1.34
West Midlands	2.0	2.0	1.5	2.1	2.1	4.6	3.3	1.55
SYSTEMATIC COMPONENT * ( $\times 100 \sigma^2$ )								
New England	0.6	1.7	1.7	5.0	4.8	12.7	12.2	2.08
Norway	0.2	2.4	1.9	9.3	10.4	14.7	27.5	1.28
West Midlands	4.4	2.9	2.1	6.2	3.7	12.2	18.5	1.33

\*See Appendix.

Source: McPherson et al. (1982)

The last study of international variations to be considered here is Aaron and Schwartz's comparison of the United States and Britain. Although more broadly conceived than the other studies discussed, **THE PAINFUL PRESCRIPTION** (Aaron and Schwartz, 1984) is relevant to the present review because of its analysis of how ten key medical procedures are provided in

the two countries. In particular, Aaron and Schwartz focussed on the use made of procedures which had become possible as a result of advances in medical technology. In broad terms, they found that most services were provided at lower levels in Britain. For example, the overall rate of treatment for chronic renal failure in Britain was less than half of that in the United States. Again, the rate of coronary artery bypass surgery in Britain was only ten per cent of that of the United States, and Britain had only one sixth of the CT scanning capability of the United States. On the other hand, three procedures were provided at essentially the same level in both countries: bone-marrow transplants, radiotherapy for cancer patients able to benefit from this treatment, and treatment for patients with haemophilia. Overall, Aaron and Schwartz concluded that the rationing of services was more difficult in Britain because of the lower overall investment in health services.

#### Small Area Variations

The second main strand in the literature is concerned with small area variations in health care delivery. One of the earliest studies of small area variations was the analysis by Lewis of the incidence of surgery in the state of Kansas (Lewis, 1969). The rate at which six common surgical procedures were performed in eleven areas was described and three to fourfold variations were found. Using multiple regression analysis, Lewis established some association between use rates and the provision of doctors and beds and concluded "the results presented might be interpreted as supporting a medical variation of Parkinson's Law: patient admissions for surgery expand to fill beds, operating suites and surgeons' time" (p.884).

This finding is supported by other studies. For example, Wennberg and Gittlesohn (1973), in an analysis of health care variations in thirteen areas of Vermont, found large differences between these areas in the provision of services, expenditure levels and use rates. Age-adjusted use rates for nine frequently performed surgical procedures "varied tremendously" (p.1104) over the thirteen areas, and positive and significant correlations were found between the supply of surgeons and surgery rates. In analysing these findings, Wennberg and Gittlesohn, echoed other authors in emphasising the difficulty of establishing correct rates of use. The existence of uncertainty concerning indications for treatment for many procedures and the lack of data on outcomes associated with treatment gave surgeons a large measure of discretion in deciding whom to treat and how. In this situation, "the possibility of too much medical care and the attendant likelihood of iatrogenic illness is presumably as strong as the possibility of not enough service and unattended morbidity and mortality" (p.1106).

In a specific study of tonsillectomy and adenoidectomy in Vermont, Gittlesohn and Wennberg reported that age-adjusted rates varied between areas from 4 to 41 per thousand children per year. It was estimated that for the entire state, 22 per cent of children would have their tonsils and adenoids removed by their twentieth birthday, but the risk of removal varied from 9 per cent to 60 per cent between areas. Directly adjacent communities had rates of 11 per cent, 19 per cent, 20 per cent, 27 per cent and 60 per cent, and the authors concluded, "it is unlikely that the differential tonsillectomy rates can be related to variations in the incidence of tonsillitis, recurrent sore throat, or otitis media. Rather, the major source of variation appears to be in differing attitudes by

physicians as to indications for the procedure" (Gittlesohn and Wennberg, 1977, p.95).

In parallel with Wennberg's studies in the United States, Noralou and Les Roos have studied small area variations in Canada (see Roos, N., 1984; Roos, N. and Roos, L., 1982). In an analysis of tonsillectomy and adenoidectomy in nine small areas in Manitoba, these researchers examined the relationship between surgical rates, the morbidity of the population, and the number of surgeons (Roos, N., Roos, L. and Henteleff, P., 1977). The analysis found that in 1973 the number of operations performed on children aged 14 years and younger varied from 80.8 to 163.6 per ten thousand population. No significant correlation was found between surgical rates and respiratory morbidity, nor between the supply of surgeons and surgical rates. Furthermore, a retrospective review of standards for selecting patients for operation did not reveal any significant correlation between selection standards and surgical rates. Some of the factors which did seem to be important were the age of the surgeon (younger doctors were more conservative), the place of training of doctors (British trained doctors were more conservative), and the specialty qualification (ENT specialists and general surgeons were more conservative than general practitioners). However, these factors could not explain all the variations that were found, and the authors concluded by emphasising the complexity of physician practice patterns.

In this context, a study by Bloor and Venters of small area variations in tonsillectomy and adenoidectomy in Scotland is relevant (Bloor and Venters, 1978). The surgical rate within one region varied between areas from 6.2 to 15.8 operations per thousand children per year. These

differences in part reflected variations between general practitioners in their rate of referral to specialists, but independently of this the practice style of specialists had a key influence on the number of operations performed. The authors noted the existence of two groups of specialists: a low acceptor, low operator group, and a high acceptor, high operator group. They concluded that variations between specialists in their propensity to list children for tonsillectomy and adenoidectomy was the most important factor in understanding differences in surgical rates. Of particular importance were the assessment practices used by specialists, clinic routines, and search procedures. A point of more general significance follows, namely that explanations of geographical variations in use rates should take into account variations between individual clinicians in their style of practice (see below).

The importance of practice style and professional uncertainty emerge as key themes in Wennberg's later work. A review of small area variations in six states in New England identified the importance of supply factors in accounting for variations, but concluded that the availability of beds and surgeons could not furnish a complete explanation of all the variations that existed (Wennberg and Gittlesohn, 1982). Rather, the judgements and preferences of doctors were a key factor. Furthermore, Wennberg and Gittlesohn coined the phrase "surgical signature" (p.106) to describe the phenomenon of high surgical rates for particular operations in individual areas. Lack of hard scientific evidence meant that authoritative standards were not available to guide medical practice and accordingly surgeons had considerable discretion in determining methods of treatment.

In a review article published in 1984, Wennberg reiterated this point,

noting that "the type of medical service provided is often found to be as strongly influenced by subjective factors related to the attitudes of individual physicians as by science" (Wennberg, 1984, p.7). Wennberg argued that neither demand-side variables such as population characteristics and illness rates, nor supply-side variables such as the availability of doctors and beds, could fully account for health care variations. Rather, he contended that "the practice style factor" (p.7) of individual clinicians was an important determinant. It was this that shaped whether patients were managed medically or underwent surgery. Practice style also affected the kinds of investigations ordered and decisions such as whether care should be provided on an inpatient or day patient basis. In support of his argument, and to return to an earlier point, Wennberg noted that the pattern of variation was similar in quite different health care systems. The common factor between these systems was that doctors shared the same scientific uncertainties concerning the value of certain procedures.

These procedures can be divided into those where there was little consensus and high variation (eg. tonsillectomy, hysterectomy) and those where there was a consensus and little variation (eg inguinal herniorrhaphy). Most procedures exhibit high variation. As Wennberg noted, the most direct evidence for the importance of practice style comes from experiments in which information on variations is fed back to clinicians, and this then leads to a change in practice.

The last study of small area variations to be considered here is an analysis of 13 regions of the United States (Chassin et al, 1986). This study found large and significant differences in the use of services



provided by all medical and surgical specialties. The authors noted that their findings are partly consistent with the view that the degree of variation for a particular procedure is linked to the degree of consensus concerning the indications for its use. However, some of the results did not support this view, nor was there evidence to suggest that doctors in high use areas performed procedures less appropriately than those in low use areas. The authors emphasised the uncertainty surrounding the reasons for variations:

"The available data do not allow us to explain the wide variations we have observed. In addition, we cannot establish the 'correct' use rates from these data. For any given procedure, geographical differences may reflect substantial inappropriate overuse in the high use areas with very little inappropriate use in the low use areas. On the other hand, variations may have occurred because physicians in the low use areas were not providing enough services to those who needed them, whereas those in the high use areas were meeting legitimate medical needs in an appropriate manner. A third possibility is that the rates of use of procedures were appropriate in both high and low use areas and that the differences in rates resulted from differences in the incidence of diseases. Finally, some combination of all three possibilities may have been responsible for our findings" (Chassin et al., 1986, p.289).

This conclusion is a useful reminder that interpretation of the evidence on variations remains highly contestable, a point reinforced by other commentators (eg Moore, 1984). As Smits (1986) has noted, two particular issues which merit further attention are, first, the extent to which

variations reflect the uneven dissemination of innovative procedures, and second, the relationship between epidemiology and use rates.

### Variations Between Doctors

As we noted in the previous section, explanations of geographical variations in use rates have focussed increasingly on variations between individual doctors. The existence of variations between doctors has been documented both in relation to hospital care and general practice. Thus, in the UK, Buttery and Snaith (1980) have reported the existence of wide differences between surgeons in the number of operations performed. The data compiled by Buttery and Snaith, relating to 1977, are displayed in Table 5.

TABLE 5

Region	General surgery		ENT Surgery	Ortho- paedics	Ophthal- mology	Urology	Gynaecology	All six specialties
	Alone	Plus Urology						
Northern	777	770	891	412	362	706	753	662
Yorkshire	1553	1608	1034	472	392	1849	679	954
Trent	1268	1230	1083	629	421	911	1320	1006
East Anglia	1051	1059	844	562	270	1106	924	784
Wessex	1128	1061	740	761	282	747	932	813
Oxford	1042	1017	1201	885	316	NA	996	934
South Western	1185	1163	762	715	335	986	1164	894
West Midlands	936	931	934	686	444	888	982	841
Mersey	951	937	1211	550	326	778	1031	845
North Western	1268	1284	1157	609	420	1376	1274	1032
All Provincial Regions	1110	1109	978	614	364	1107	998	881
NW Thames	1002	999	678	804	379	970	854	837
NE Thames	NA	NA	NA	NA	NA	NA	NA	NA
SE Thames	1315	1290	1046	892	368	1176	1238	1057
SW Thames	1589	1615	1023	1247	448	1816	1320	1288

*Health Trends*, 1980, Volume 12.

In the six surgical specialties examined, the regional average number of operations per surgeon varied from 662 to 1,288. As the table shows, the extent of variation is often greater in individual specialties. More recently Yates and colleagues have compared the workload of orthopaedic surgeons (registrars, senior registrars and consultants). The range of operations per surgeon was from less than 150 to over 750 per senior doctor per year. Yates et al (1985) argued that variations on this scale could not be explained by variations in case mix or compensating workload in other areas. They concluded "there is undoubtedly a proportion of surgeons in this country who could undertake more operating" (p5).

In the case of GPs there is a considerable volume of data in the UK on variations between GPs in terms of prescribing habits, investigation rates, and home visits (see for example Metcalfe, 1985; Crombie, 1984). An issue of continuing interest has been variations in referral rates. Dowie's (1984) analysis of GP referrals to medical outpatient departments identified three sets of factors relevant to referral decisions: professional attributes such as medical knowledge and judgement; personal style, such as interaction with the patient; and knowledge of the health care system. Wilkin and Smith (1986), in a review of the literature on GP referrals to consultants, noted the importance of Dowie's research, but argued that most of the variations that exist remain unexplained. On the basis of research conducted in Manchester, these authors concluded that neither patient characteristics nor the characteristics of GPs and their practices could adequately account for the variations observed, and they called for further research "looking both at those patients who are referred and those who are not, how those decisions are arrived at, what

are the outcomes for patients and the costs both for services and for the community" (p37).

A study which goes some way to meeting these objectives is the work done by Cantley and Hunter (1985) on GP referrals of elderly people. This study examined decisions made by GPs in two Scottish towns, and it identified a range of considerations used by GPs in deciding on appropriate methods of treatment. The particular focus of the study was whether patients were referred to a specialist geriatric unit or were treated in local GP hospital beds. Key variables included clinical considerations, social considerations, GPs' perceptions and expectations of services, resources, constraints and pressures, service management and professional interests.

Apart from the work of Cantley and Hunter and Dowie, there appears to be little analysis of the way in which practice style operates in primary care. This is surprising in view of the role of GPs as gatekeepers and rationers of scarce health service resources (Day and Klein, 1986). In this context, it is worth noting that the DHSS is undertaking a pilot study of referral rates in North Lincolnshire to establish why rates vary. One possibility is that this will lead to a quota system involving a limit on the number of referrals GPs are permitted to make. This would clearly have significant implications for doctors.

#### Variations' Studies in the United Kingdom

Many of the studies cited in this paper concern variations in use rates in the United Kingdom. In addition to these studies, other work in the United Kingdom has comprised Sanderson's study of regional variations in

cataract extraction rates (Sanderson, 1980), Fowkes and McPake's analysis of regional variations in outpatient attendances (Fowkes and McPake, 1986), McPherson et al's research on variations in cholecystectomy rates (McPherson et al, 1985), Coulter and McPherson's analysis of hysterectomy rates (Coulter and McPherson, 1986), and the DHSS review of geographical variations in acute services (DHSS, 1981).

In his study, Sanderson (1980) found large variations in cataract extraction rates between English regions. These were partly associated with the proportion of elderly people in the population and with bed supply, but were only weakly correlated with manpower supply. Sanderson concluded that there were no clear cut explanations of the variations that existed, and he noted "local factors are important variables in the resource supply/utilisation equation and a single generalised statement about variables in surgical rates cannot be made. The crucial question in terms of health service policy, however, is whether the variations in cataract extraction reflect inequalities in the opportunities for care. It seems likely from the relationship of bed supply to operation rates that this is so..." (p496).

Fowkes and McPake (1986) identified large variations between English regions in the use of outpatient services. Among half the specialties examined total outpatient attendances per 1000 population varied by threefold or more. In most other specialties the range was around two fold. In addition, there were considerable variations in the ratio of total to new outpatient attendances and in attendances per clinic session. Fowkes and McPake concluded that available data did not permit an explanation of these variations and they called for further research.

The aim of this would be to develop guidelines for clinical practice and optimum rates of activity for use as performance indicators by managers and planners.

McPherson et al's study of cholecystectomy (1985) examined variations in operation rates in seven English districts. In this study, an independent measure of morbidity was available (the prevalence of gallstones at necropsy), and it was found to be positively and strongly associated with cholecystectomy rates. However, the authors went on to note that variations in use rate between countries could not be explained by variations in morbidity. Rather, they suggested that international variations reflected the application of different clinical thresholds in decisions on whether or not to operate. McPherson et al also speculated that financial arrangements in North America encouraged intervention in a way that did not occur in England.

Coulter and McPherson (1986), in a review of the evidence on variations in hysterectomy rates, found no support for the hypothesis that patient demand was a key factor in the decision on whether to undertake surgery. They also discounted the prevalence of gynaecological morbidity as an explanation, arguing that "it is unlikely that relevant morbidity differences could account for these large variations" (p384). Coulter and McPherson emphasised instead the influence of supply factors and practice style, and they echoed McPherson et al's study of cholecystectomy in suggesting that international differences probably reflected differences in the mode of organising and financing health services.

The DHSS review of the acute hospital sector presented a range of data on variations between English regions in the provision and use of in-patient and out-patient services. The data revealed the existence of wide variations in discharges and deaths and the use of outpatient services, as well as in length of stay, waiting lists, available beds, doctors and occupancy rates. As the review pointed out, the region which had consistently high lengths of stay (Mersey) also had the most beds per 1000 population, while the region with consistently short stays (Oxford) had the least beds per 1000. Referring to the work of Buttery and Snaith (1980), the review drew attention to the lack of correlation between waiting list size and the level of surgical provision and to the relatively constant waiting time for operations whatever the level of provision. In their own analysis, Buttery and Snaith observed that differences in the provision of surgical services were greater than differences in financial provision, and they noted "this suggests that medical policies exert a greater influence on health services provision than financial policies" (p59).

As this discussion has indicated, work on geographical variations in the UK has encompassed a range of aspects of health care provision besides use rates. Three broad categories of investigation can be identified:

- a) The use of statistical indices to compare the performance of health authorities, as exemplified by the development of performance indicators by the DHSS and inter-authority comparisons by John Yates. The data used typically focusses on service inputs such as staffing and beds and measures of efficiency such as length of stay, throughput and costs per case. Information on waiting lists and waiting times is

also examined (see for example College of Health, 1985). Performance indicators and inter-authority comparisons are used in the planning and management of health services to identify aspects of service provision which require further investigation.

- b) Analysis of variations in the allocation of health care resources, as exemplified by RAWP. The principal concern here has been to establish a means of measuring the need for health care of the population in order to achieve greater justice in the distribution of resources. The formula does not take into account how resources are used;
- c) Research into geographical variations in mortality and morbidity, as exemplified by work on avoidable deaths at St. Thomas's Hospital. This work examines variations between health authorities in mortality from conditions amenable to medical intervention (for example, Charlton et al, 1983). In this way, it is hoped to develop indicators of outcome against which the use of resources can be evaluated. Other work in this category includes Peter Townsend's research on small area variations in mortality and morbidity in Bristol and in the Northern Region, spatial analysis of mortality from particular causes (eg. cancer), and the Confidential Enquiry into Perioperative Deaths.

It can therefore be seen that considerable effort has been devoted to describing and analysing health care variations in the United Kingdom. Despite this, two main conclusions can be drawn from existing work. First, many of the studies mentioned have been carried out by researchers working independently of one another. Thus, research on variations in use rates has typically been conducted by researchers based in academic departments, while work on the RAWP formula has in the main been the responsibility of analysts based in government departments. Again,



research on geographical variations in mortality has not been integrated with work on performance indicators. Only in one or two cases have attempts been made to bring different kinds of data together and there would seem to be scope for further work in this direction (see below).

Second, notwithstanding the availability of information on a national basis (eg through HIPE), there has been little systematic analysis of variations in use rates. Most of the work done appears to have involved ad hoc studies rather than routine analysis of existing data sets. There would seem to be merit in exploiting these data sets more fully as a way of monitoring trends and patterns in use rate variations. To explore these issues further, we now examine the implications of the variations literature for researchers, the medical profession and policy makers. In so doing, an attempt is made to identify priorities for future work.

#### Issues for Debate

A useful starting point in considering priorities for future work is Wennberg's plan for dealing with variations (Wennberg, 1984). This has three parts. First, Wennberg argues that there is a need to monitor performance in small areas on service inputs, use rates and outcomes. The resulting information should be fed back to clinicians and decision makers in order to influence and change clinical practice, as is the case in the Maine Medical Assessment Programme. Second, Wennberg contends that greater efforts should be put into assessing the effectiveness of services and measuring outcomes. Indeed, at the Copenhagen conference on variations held in November 1986, Wennberg argued that the outcome problem is the major challenge facing variations researchers. This can be tackled through literature reviews, consensus techniques and original research, in

some cases making use of existing administrative data bases (eg. MEDICARE). Third, there is a need to reduce unnecessary or inappropriate use of hospital services, principally through more concerted efforts on the part of the medical profession.

There are echoes here of work done by Robert Brook and his colleagues as part of the RAND programme of health services research. One element in this programme is designed to describe more systematically the pattern of health care variations that exist in the USA (see Chassin et al, 1986). Another element focusses on establishing the missing clinical links between data on variations and data on appropriateness (Brook et al, 1984). Thus, in a series of studies, groups of experts have been brought together to list and rank indications for treatment for specific procedures (see for example, Solomon et al, 1986). An expert consensus has then been established, supported by a literature review, and in some cases this has been applied retrospectively to establish levels of inappropriate use. To date the RAND consensus technique has been used mainly in the USA, although there have been two (as yet unpublished) studies conducted in the UK.

These initiatives are consistent with the proposals from Schacht and Pemberton (1985) for the greater use of review committees to establish the circumstances under which treatment should and should not be provided. They are also congruent with efforts made in a number of countries to use consensus techniques of varying kinds to review controversial areas of medical practice. The results of these reviews are intended principally to influence professional opinion, but they may also be used to provide more information to the public in the hope of stimulating informed

choice on the part of service users (see Wennberg and Gittlesohn, 1982). The influence that users can have is well demonstrated by experience in Switzerland where the rate of hysterectomies fell following publication of information of variations in the use of hysterectomies in the press (Domenighetti, 1986).

It would seem, therefore, that the literature on health care variations has a number of implications for those involved in this field. These are:

- a) The need for systematic monitoring of variations in service inputs, use rates and outcomes;
- b) The importance of feeding back information gathered in (a) to clinicians, policy makers and the public;
- c) The need to investigate the outcomes associated with different treatments;
- d) The need for greater efforts on the part of the medical profession to engage in clinical audit;
- e) The value of the more widespread use of consensus techniques of varying kinds in order to develop guidelines and appropriateness indications;
- f) The importance of further analysing the reasons for variations, including the relationship between epidemiology and use rates, and the relative importance of demand factors, supply factors and practice style;.
- g) As part of f), the need to complement statistical analyses of large data sets with research into the component parts of practice style to determine how treatment decisions are made by clinicians.
- h) The possibility of developing standards for use by policy makers and

managers, for example concerning GP referral rates, use of outpatient services, and the number of operations to be performed for a given population in particular specialties.

In terms of future research, there would also seem to be value in a local study integrating data on variations in inputs, provision, use rates and outcomes. As the DHSS argued in its review of acute services, any serious attempt to evaluate variations in activity and resource usage should focus on the district level, starting with one or two pilot studies (DHSS, 1981). This is already done in part through the use of performance indicators, but more detailed analysis is needed. There are several possible approaches, but one option is to take two districts in a similar RAWP position and identify for each district variations in use rates and outcomes. Such a study would build on Holland's preliminary analysis of this issue (Holland, 1986) and would develop further the research by Wennberg and colleagues in the United States on differences in New Haven and Boston in health service expenditure, use rates and provision (Wennberg et al, 1987).

#### Summary

Analysis of the literature on geographical variations in health care points to the following conclusions:

1. There is a large and growing literature on variations, much of it concerned with variations in the use of common surgical procedures. A few procedures have been examined extensively (eg. tonsillectomy and adenoidectomy), others have been relatively neglected.
2. There are significant international variations in the provision and use of services. Differences in methods of organising and financing

health services appear to be important in explaining these variations.

3. There are significant variations between small areas in the provision and use of services, both between and within countries.
4. The extent of variation appears to be similar in different countries.
5. Many possible explanations have been put forward for small area variations, including the characteristics of the population served, the availability and supply of services and the practice style of clinicians. There is little consensus on these explanations, although much research points to the importance of supply variables and the practice style factor.
6. There is little agreement in the literature on the correct or appropriate use of services and there is a continuing debate on whether high rates signify unnecessary usage or low rates signify under provision. If in the USA the implication of much of the work done is that some services are overprovided, in the UK the reverse often holds.
7. One of the difficulties in resolving these issues is that there are few data on the outcomes associated with different treatments, on the pattern of morbidity by area, or on appropriate indications for use.
8. In general, the literature raises more questions than it answers. These questions provide fertile territory for health policy analysis.

Aaron H J and Schwartz W B (1984). The Painful Prescription. The Brookings Institution, Washington DC.

Bloor M and Venters G (1978). An Epidemiological and Sociological Study of Variations in the Incidence of Operations on the Tonsils and Adenoids. University of Aberdeen, Institute of Medical Sociology, Occasional Paper No2.

Brook R et al (1984). 'Geographic Variations in the Use of Services: Do They Have Any Clinical Significance'. Health Affairs, Vol 3, No2, pp63-73.

Bunker J P (1970). 'Surgical Manpower. A Comparison of Operations and Surgeons in the United States and in England and Wales'. NEJM, Vol 282, 15 January 1970, pp135-144.

Buttery R B and Snaith A H (1980). 'Surgical Provision, waiting times and waiting lists'. Health Trends, Vol 12 pp57-61.

Cantley C and Hunter D (1985). 'People Processing: Towards a Typology of Selected General Practitioner Referral and Admission Practices in Care of Elderly People'. Ageing and Society, Vol 5 pp267-288.

Charlton J R H et al (1983). 'Geographical Variation in Mortality From Conditions Amenable to Medical Intervention in England and Wales'. The Lancet 26 March, pp691-696.

Chassin M R et al (1986). 'Variations in the Use of Medical and Surgical Services by the Medicare Population'. NEJM, Vol 314, pp285-290.

College of Health (1985). Guide to Hospital Waiting Lists 1985. London.

Copenhagen Collaborating Centre (1985). CCC Bibliography on Regional Variations in Health Care. Copenhagen.

Coulter A and McPherson K (1986). 'The Hysterectomy Debate'. Journal of Social Affairs, Vol 2, No4, pp379-396.

Crombie D L (1984). Social Class and Health Status: Inequality or Difference. Occasional Paper 25, the Royal College of General Practitioners.

Day P and Klein R (1986). 'Controlling the gatekeepers: the accountability of general practitioners'. Journal of the Royal College of General Practitioners, March pp129-130.

DHSS (1981). Report of a Study of the Acute Hospital Sector, London.

Domenighetti G et al (1986). 'Reducing Hysterectomies: The Mass Media'. Paper given at Copenhagen conference on regional variations, November 1986.

Dowie R (1983). General Practitioners and Consultants: a study of outpatient referrals, King Edward's Hospital Fund for London, London.

Fowkes F G R and McPake B I (1986). 'Regional Variations in Outpatient Activity in England and Wales'. Community Medicine, Vol 8, No4, pp286-291.

Gittlesohn A M and Wennberg J E (1977). 'On the incidence of tonsillectomy and other common surgical procedures'. In Bunker J P, Barnes B A and Mosteller F (eds) Costs, Risks and Benefits of Surgery (Oxford University Press, New York).

Glover J A (1938). 'The Incidence of Tonsillectomy in School Children'. Proceedings of the Royal Society of Medicine, Vol xxx, pp1219-36.

Holland W W (1986). 'The RAWP Review: Pious Hopes'. The Lancet, 8 November, pp1087-90.

Lewis C E (1969). 'Variations in the Incidence of Surgery'. NEJM, Vol 281, 16 October 1969, pp880-884.

Mc Pherson K et al (1981). 'Regional Variations in the Use of Common Surgical Procedures: Within and Between England and Wales, Canada and the United States of America'. Social Science and Medicine, Vol 15A, pp273-288.

McPherson K et al (1982). 'Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England and Norway'. NEJM, Vol 307, 18 November, pp1310-14.

McPherson K et al (1985). 'Do cholecystectomy rates correlate with geographic variations in the prevalence of gallstones?'. Journal of Epidemiology and Community Health, Vol 39, No2, pp179-182.

Metcalfe D (1985). "No Excuses". pp184-204 in Gray D J Pereira (ed). The Medical Annual 1985, (John Wright, Bristol).

Moore F D (1985) 'Small Area Variations Studies: Illuminating or Misleading?'. Health Affairs, Vol 4, No1, pp96-101.

Pearson R J C et al (1968). 'Hospital Caseloads in Liverpool, New England and Uppsala'. The Lancet, 7 September, pp559-566.

Roos N (1984). 'Hysterectomy: Variations in Rates Across Small Areas and Across Physicians' Practices'. American Journal of Public Health, Vol 74, No4, April, pp327-335.

Roos N and Roos L (1982). 'Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population'. Medical Care, Vol xx, September, pp945-58.

Roos N, Roos L and Henteleff P (1977). 'Elective Surgical Rates - Do High Rates Mean Lower Standards?'. NEJM, 18 August, pp360-365.

Sanderson H F (1980). 'Regional variation in cataract extraction rates and their relationship with resource supply and need'. Journal of the Royal Society of Medicine, Vol 73, July, pp492-496.

Schacht P J and Pemberton A (1985). 'What is Unnecessary Surgery? Who Shall Decide? Issues of Consumer Sovereignty, Conflict and Self-Regulation'. Social Science and Medicine, Vol 20, No3, pp199-206.

Smits H L (1986). 'Medical Practice Variations Revisited'. Health Affairs, Vol 5, No3, pp91-96.

Solomon D et al (1986). Indications for Selected Medical and Surgical Procedures - A Literature Review and Ratings of Appropriateness. Cholecystectomy. RAND Corporation, Santa Monica.

Vayda E (1973). 'Comparison of Surgical Rates in Canada and in England and Wales'. NEJM, Vol 289, 6 December, pp1224-1229.

Wennberg J et al (1987). 'Are Hospital Services Rationed in New Haven or Over-Utilised in Boston?'. The Lancet, 23 May pp1185-9.

Wennberg J (1984). 'Dealing With Medical Practice Variations: A Proposal for Action'. Health Affairs, Vol 3, No2, pp6-32.

Wennberg J and Gittlesohn A (1973). 'Small Area Variations in Health Care Delivery'. Science, Vol 182, 14 December, pp1102-1108.

Wennberg J and Gittlesohn A (1982). 'Variations in Medical Care Among Small Areas'. Scientific American, Vol 246, April, pp100-112.

Wilkin D and Smith T (1986). Variation in GP Referrals to Consultants. Centre for Primary Care Research, University of Manchester.

Yates, J et al (1985). 'Comparing Performance'. NAHA News, November.





