RESEARCH REPORT

2

HEALTH CARE VARIATIONS

ASSESSING THE EVIDENCE

EDITED BY CHRIS HAM

KING'S FUND INSTITUTE

'No 2' in a series of research reports on current health policy issues.

ISBN 1 870607 06 6

Published by the King's Fund Institute
126 Albert Street, London NW1 7NF

Copies are available at £6.95

J7310 / DESIGN & PRINT / REDESIGN / 01-533 2631

HEALTH CARE VARIATIONS

ASSESSING THE EVIDENCE

EDITED BY CHRIS HAM

CONTENTS

TABLES AND FIGURES

Tables

Figures

CONTRIBUTORS

Nigel Buck
London School of Hygiene and Tropical Medicine

Dr Chris Ham
Policy Analyst
King's Fund Institute

Professor Bryan Jennett
University of Glasgow and King's Fund Institute

Dr Klim McPherson
Department of Community Medicine and General Practice
University of Oxford

Dr Myfanwy Morgan
Department of Community Medicine
United Medical and Dental Schools
St Thomas's Hospital

Dr John E Wennberg
Professor of Epidemiology
Dartmouth Medical School, USA

ACKNOWLEDGEMENTS

Table 1.2 is reproduced with permission of the Controller of Her Majesty's Stationery Office. Figure 6.2 is reproduced with permission of the American Medical Association from the Journal of the American Medical Association, 1987, volume 257, pp.933-936 (copyright 1987, American Medical Association).

INTRODUCTION

Chris Ham

Despite 40 years of the NHS, there are continuing geographical variations in the provision, use and outcome of health services. These variations are manifested in the greater availability of doctors and hospital beds in some areas than others;differences in the use made of these doctors and beds; and variations in the outcomes associated with the use of services. Awareness of these variations has resulted in action by policy makers on three main fronts.

First, since the early 1970s, the DHSS has sought to reduce variations in the allocation of health care resources by basing allocations on the needs of each area. In the period since 1976, need has been assessed using the formula recommended by the Resource Allocation Working Party (RAWP). Steady progress has been made in moving towards the target allocations indicated by RAWP and as a consequence the range of variation between health authorities has narrowed considerably.

Second, during the 1980s, there has been increasing interest among policy makers in variations in the use of resources. This is exemplified by the development of performance indicators by the DHSS and inter-authority comparisons by John Yates. These tools draw on routine statistics to compare the performance of health authorities. The data used focus mainly on service inputs such as staffing and beds and measures of activity such as length of stay, throughput and unit costs. Information on waiting lists and waiting times for treatment are also examined.

Third, policy makers have shown an interest more recently in the outcomes associated with health services, including geographical variations in outcomes. Two examples of this interest are the use of information on avoidable deaths, and the support given to the Confidential Enquiry into Perioperative Deaths (CEPOD). In view of ministers' expressed wish to put more emphasis on the results obtained from health services and less emphasis on resource inputs, analysis of outcomes through these and other means is likely to receive increasing attention.

Interest in health service variations has not been confined to policy makers. Organisations representing patients and service users have long expressed concern at variations in the accessibility of services. This continues to be an issue of interest to these organisations, as in the analysis done by the College of Health of variations in hospital waiting lists between areas. From a different perspective, the medical colleges and specialist associations have through their work sought to identify variations in clinical practice and to reduce or eliminate those variations deemed to be unacceptable. CEPOD is one example of this and demonstrates how the profession itself is able to take action to monitor standards.

Alongside these developments, a considerable international research effort has gone into the study of geographical variations in hospital utilisation rates. This research demonstrates the existence of wide differences in utilisation rates both between and within countries. In an attempt to explain these variations, researchers have developed a number of competing theories to test the influence of demand, supply and professional decision making on utilisation.

Against this background, this Report presents a range of material on different aspects of variations, drawing on a conference held at the King's Fund in 1987. A major focus of the Report is variations in hospital utilisation rates. There are two reasons for this. First, hospitalisation rates have a significant influence on resource use and are more important as a determinant of health service spending than variables such as length of stay, throughput or the number of tests performed on patients. If, as seems likely, the efficiency of clinical services comes under closer scrutiny within the NHS, hospitalisation rates will be a major focus of investigation. Second, the existence of variations in hospitalisation rates raises the question of whether these variations reflect the need of the populations served. A key objective of the NHS is to provide services on the basis of need. Do variations in utilisation rates indicate that the NHS is responding to differences in need between areas or do these variations in themselves give rise to inequities in service provision?

In addressing these questions, the Report draws on a diverse literature scattered through specialist journals and publications. This literature is not readily accessible to researchers or policy makers. The Report therefore offers a review and commentary on the literature in an attempt to summarise the current state of research and analysis. The Report also seeks to draw out the implications of what is known about variations in utilisation rates for policy makers and researchers.

Alongside the analysis of hospital utilisation rates, two related topics are examined. The first is variations in length of stay in surgery. Again, this is likely to be examined more closely as part of the concern to increase efficiency in the NHS. The second topic is variations in perioperative deaths between areas. With CEPOD to be extended to the whole of England, the results of the first phase of work are of relevance throughout the NHS. Both of these topics are thus central to the current debate about efficiency and effectiveness in the NHS.

Structure of the Report

The Report is organised as follows. In Chapter 1, Chris Ham reviews the literature on geographical variations in hospital utilisation rates. The chapter brings together evidence on international variations and small area variations and analyses the various theories that have been put forward to explain these variations. The lack of a convincing explanation is highlighted, although much research points to the influence of supply variables and professional decision making in accounting for variations in utilisation rates.

In Chapter 2, Klim McPherson takes the argument forward by exploring in more detail possible explanations of variations in hospitalisation rates. McPherson emphasises that different factors may be at work depending on the level of analysis. McPherson also stresses that the extent of variation differs between procedures. There are as yet no simple

explanations of variations between countries or regions, but evidence of large variations can be used to identify priorities for research into the outcomes associated with different levels of intervention.

In Chapter 3, Myfanwy Morgan examines evidence on variations in length of stay for three surgical procedures. Data are analysed for the districts in two English regions and significant differences between districts are identified. Translation of these differences into numbers of bed days indicates that about 2,000 more bed days were occupied than expected in the longest stay districts and 2,000 fewer beds days in the shorter stay districts. Morgan emphasises that supply variables and differences in the organisational and clinical practices of individual hospitals and doctors are important sources of these variations.

In Chapter 4, Nigel Buck uses data from CEPOD to identify geographical variations in death rates after surgery. The evidence from CEPOD reveals wide differences between districts in the proportion of deaths considered by expert assessors to be avoidable. There are also significant differences between regions in the organisation of clinical care. Buck emphasises the importance of enthusiasm and trust among consultants in work of this kind.

The importance of involving clinicians in reviewing variations is a key theme in Chapter 5 in which Bryan Jennett argues that surgeons would do well to collect more data on variations themselves. Jennett maintains that surgeons are most likely to act constructively on information that is seen to be clinically relevant and on matters that can be altered by their own actions. In this context, Jennett emphasises the need for audit and for more evidence on the outcomes of surgery as well as activity rates.

The need for outcomes research is also highlighted by John Wennberg in Chapter 6. As Wennberg notes, it is not possible to say whether high or low hospital utilisation rates are right. Research into outcomes is therefore needed to identify whether different utilisation rates are associated with different outcomes. According to Wennberg, the priority should be an assessment programme to evaluate alternative treatments for specific conditions such as prostatism and stable angina. Such a programme would improve clinical decision making by reducing uncertainty about the probabilities and value to patients of the outcomes of care.

Finally, in Chapter 7, the main conclusions and implications of the Report are identified. Key issues discussed in this chapter are the need for more research on outcomes, the need for greater efforts on the part of the medical profession to engage in audit, and the importance of developing guidelines and standards for use by both clinicians and policy makers.

1 A REVIEW OF THE LITERATURE

Chris Ham

1.1 Introduction

The literature on geographical variations in the utilisation of health services 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). Although research on these issues spans half a century, it is only in the last ten years that a major effort has been made to describe and explain the variations that exist.

This is illustrated in Table 1.1 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 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 chapter although other key studies will also be cited where appropriate.

TABLE 1.1 · 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 hospital utilisation 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 the way they use the facilities that are available.

Considerations of effectiveness, efficiency and equity lie behind the interest in health service variations. The literature contributes to analysis and discussion of each of these issues by raising questions such as:

- Does the high rate of hospitalisation in certain areas indicate unnecessary or inappropriate use?
- Can the higher costs associated with high rates of hospitalisation be justified in terms of improved health outcomes?
- Are differences in hospitalisation 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 professional decision-making. 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 professional decision-making but there remains a good deal of uncertainty. In the absence of convincing explanations, one of the main contributions which analysis can make is to identify key issues for debate and action by policy makers, the medical profession and researchers.

1.2 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.

To give some examples, Pearson *et al.* found that tonsillectomy and adenoidectomy was performed more than twice as often in New England as in Liverpool and four times as often as 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 Kohn and White (1976). These authors examined hospital utilisation rates in 12 areas of seven countries. The study found that standardised hospitalisation rates varied more than twofold between areas. In analysing the reasons for these variations, Kohn and White emphasised the influence of the supply of beds, differences in admission thresholds, and professional decision-making. As the authors commented:

> *It would appear that where beds are available, more patients with lower levels of perceived morbidity will have an increased chance of being admitted to hospital, whether or not this is the most appropriate or optimal source of care* (p.220).

The importance of supply as a determinant of utilisation was also highlighted in an analysis by McPherson and colleagues 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. 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 most original and important contribution of this study was the finding that variations in utilisation rates in each country followed a characteristic pattern. Independently of the method of organising and financing health care the extent of variation was similar: 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 were controversy and uncertainty among professionals about the indications for a procedure. The implication of this was that researchers should seek to reduce this uncertainty by examining outcomes associated with procedures that had highly variable rates of use.

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. Aaron and Schwartz focussed on the use made of procedures which had become possible as a result of advances in medical technology. As expected, 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.

1.3 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 utilisation 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 utilisation rates. Age-adjusted utilisation 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 utilisation. 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 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, 1984; Roos, and Roos, 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, Roos, and Henteleff, 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, clinic routines and search procedures used by specialists. 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. Failure to evaluate the effectiveness of different procedures 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 for which there was poor consensus and high variation (eg. tonsillectomy, hysterectomy) and those for which 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 his experience of feeding back information on variations to clinicians. In a number of cases, this has resulted in changes in practice.

Further light is shed on this debate by an analysis of variation in use rates in 13 regions of the United States (Chassin *et al.*, 1986). Although covering more populous areas than the other analyses discussed here, this study found large and significant differences in the utilisation 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.

In a related study, Chassin *et al.* (1987a) examined in detail the reasons for wide variations in use of three procedures. Using 1981 data, patient records were reviewed using ratings of appropriateness developed by expert panels. Differences among the 13 areas in levels of appropriateness were small. There was therefore no evidence that high use areas were using services less appropriately than low use areas. The authors suggested that further analysis was required including investigation of the possibility of underuse of services in low use areas. In commenting on these findings, Wennberg (1987d) argued that the results were also consistent with the professional uncertainty hypothesis. In other words, variations in use rates reflected a lack of professional consensus about treatment allowing doctors to treat patients in a way that was considered appropriate both by the doctor involved and by independent experts.

1.4 Variations' Studies in the United Kingdom

Many of the studies cited in this chapter concern variations in hospital utilisation rates in the United Kingdom. There have been a number of other analyses of variations in the United Kingdom including work which has examined regional variations in cataract surgery (Saderson, 1980), regional variations in outpatient attendances (Fowkes and McPake, 1986), variations in cholecystectomy rates (McPherson *et al.*, 1985a) and hysterectomy rates (Coulter and McPherson, 1986), and intra regional variations in the treatment of end stage renal failure (Dalziel and Garrett, 1987). While all of these studies report wide variations in utilisation rates, the explanations that have been put foward are partial and not always consistent. Morbidity, supply, access to services and practice style are some of the factors suggested as influences on utilisation rates, but there is considerable uncertainty about the relative importance of these factors.

In this context, a study of the use of hospital resources in Liverpool by Logan *et al.* (1972) remains an important point of reference. The starting point for this study was the existence in the Liverpool region of high numbers of acute beds in relation to the population served and high hospitalisation rates. The study sought to examine why Liverpool used more acute beds than other regions. Analysis suggested that

the supply of beds and the practice style of hospital consultants in the region were key factors. Liverpool used more beds because it had more beds and doctors had lower thresholds of admission than in other parts of the country studied. These factors were more important than any evidence about greater need in the community in accounting for higher hospitalisation rates and the slower tempo of work in Liverpool hospitals. Other relevant considerations were the mislabelling of some beds as acute instead of geriatric, and the poor organisation of surgical services with beds being distributed in a large number of hospitals, some of which were small and uneconomic. One of the main conclusions of the study was that some patients were treated in hospital when they could more appropriately have been cared for in the community.

The findings of the Liverpool study were echoed in an analysis of acute hospital services in London carried out by the London Health Planning Consortium (1979). This reported that hospitalisation rates in non-regional acute specialties were a third higher in inner London than in the home counties with outer London districts in an intermediate position. The London Health Planning Consortium recognised that these variations were likely to be influenced by morbidity, differences in environmental and social conditions, and the availability of other health services. However, the Consortium emphasised the crucial importance of supply in determining utilisation, and it demonstrated that there was a significant positive correlation between hospitalisation rates and the supply of beds. Furthermore, evidence was presented to illustrate the existence of a similar relationship between utilisation and supply in the regions of England.

1.5 Variations Between Doctors

As we have noted, explanations of geographical variations in utilisation 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 1.2.

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.

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 where the evidence indicates that the number of referrals varies from 1 per 100 consultations to 24 (Wilkin and Smith, 1986). 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

TABLE 1.2 · OPERATIONS PER CONSULTANT (WTE)

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

Source: Buttery and Snaith (1980)

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).

1.6 Summary

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

- 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.
- 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 some of these variations.
- There are significant variations between small areas in the provision and use of services, both between and within countries.
- The extent of variation appears to be similar in different countries.
- 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. (For a recent review of North American research on small area variations see Paul-Shaheen *et al.*, 1987).
- 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.
- 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.

In general, the literature raises more questions than it answers. As Chassin and colleagues concluded in their study of variations in 13 regions of the United States:

> *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 occured 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, 1985).

2 VARIATIONS IN HOSPITALISATION RATES: WHY AND HOW TO STUDY THEM

Klim McPherson

2.1 Introduction

Ever since Glover's seminal work in the 1930s and 1940s concerning differences in the incidence of tonsillectomy between education areas of Kent (Glover, 1938), the fact that there is variation in population-based rates of hospitalisation for certain conditions has been well known. The point of Glover's work was to draw attention to 'excessive' rates of tonsillectomy. A similar objective was probably behind Bunker's work which pointed to twice the incidence of many operations in the U.S. compared with the U.K. (Bunker, 1970). Such work certainly highlighted the possibility that expensive and risky procedures were being performed much more frequently in the U.S. than in the U.K.. Evidence on variations is of use mainly in identifying areas for future research, in particular, research into the causes of the variations that are observed. Once the main causes can be established, it is then possible to examine appropriateness with respect to issues of effectiveness, efficiency and equity.

The study of variations alone has little to do with audit or the policing of hospital services. It is simply a methodology for investigating which kinds of admission are subject to most variation in rates between populations, aggregated in different ways. We have already mentioned two quite separate aggregation levels, the first education areas of Kent, the second distinct countries. Level of aggregation is important in the interpretation of the variation in rates observed and an essential part of the process of understanding causes.

In this paper we assume that catchment populations are well defined so that denominators of people at risk of hospitalisation are known and relatively unambiguous. We also assume that age and sex standardisation is performed so that different rates cannot be attributable to age and sex structure of populations. This is not necessarily straightforward in practice, but it is basic to most epidemiology.

2.2 Variations Between Countries

Figure 2.1 shows the standardised rates for eight common surgical operations in five countries in the mid 1970s. It is clear that these various populations underwent surgery at quite different rates. In some cases there is a non-American consensus, but in other cases this is not so. On this evidence, hysterectomy seems to be the most variable between countries and appendicectomy the least. Table 2.1 shows the cost implications for the NHS at 1986 prices were the U.K. to perform these operations at the same rate as Norway, Canada or the U.S., simply as a matter of multiplying the unit costs. Clearly extra investment in building, resources and manpower would also be required before the U.K. could begin to emulate North American rates.

To comprehend these variations and assign causation with any certainty is difficult because there are so many differences between these countries, from organisation of health care, through payment structure and supply of resources to systematic differences in morbidity rates or patient expectations. Hence it is impossible to be certain of the reasons for these observed variations. It is clearly more expensive to have a high rate but whether or not that money is being well spent is a much more difficult question. Likewise, the converse question of whether a low rate is a sensible saving of resources is very problematic.

2.3 Why There Is Variation

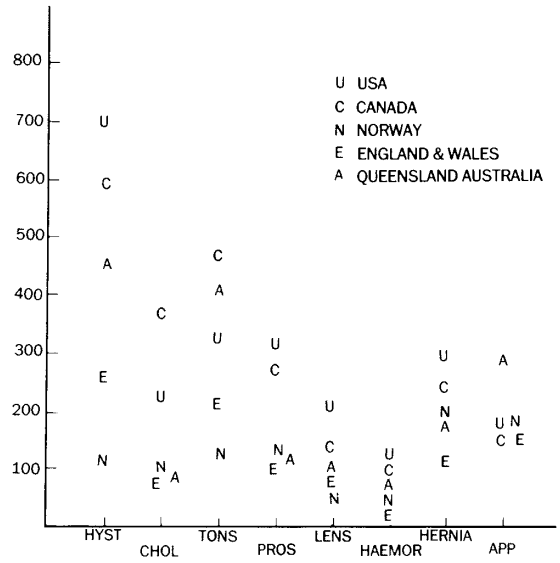
Whenever variation is observed in hospitalisation rates, there can only be a limited number of possible explanations (McPherson *et al.*, 1981). These are summarised in Table 2.2. The first is the most attractive explanation because a systematic, invariant and appropriate response to illness is what is generally expected of health care provision. Thus a threefold variation in the rate of hysterectomy between areas could, and perhaps should, be a manifestation of a threefold variation in morbidity combined with a well

TABLE 2.1 · NATIONAL HEALTH SERVICE COSTS AT DIFFERENT NATIONAL RATES

	£ Million 1986 Prices			
	Norway	E & W	Canada	USA
Hysterectomy	18	42	114	131
Cholecystectomy	31	23	63	72
Prostatectomy	30	24	65	71
Tonsillectomy	3	10	21	16
Inguinal Hernia	27	15	35	40
Operation on lens	17	26	55	75
Appendicectomy	30	33	30	28
Haemorrhoidectomy	4	2	14	13
TOTAL	160	176	455	447

FIGURE 2.1 · SURGICAL PROCEDURES

AGE STANDARDISED RATES/100,000 (AT RISK)



tested and understood consensus on the proper indications for the operation. While the former is plausible, the latter is usually less likely. For many illnesses, there is a consensus of medical opinion about what to do but for others there is an important variation in the severity of symptoms and considerable uncertainty about how best to treat them. Where there is uncertainty, then other constraints may well determine, to some degree, the decision to intervene or not. Apart from age and sex composition for which we can standardise, or mere random variation which we discuss later, the remaining headings in Table 2.2 include these other possible causes.

Clearly people cannot be admitted to full beds or be operated upon in fully booked theatres. Moreover, given budgetary limits, beds sometimes have to be kept empty for fear of overspending. Supply of facilities can in this way affect the observed rate. Similarly, individual medical opinions will differ as a function of education, experience and research interpretation. Also, patients may expect an operation or be totally averse to one except in dire circumstances. For the removal of organs, the population-based rate can be importantly affected by high rates in previous years leaving the population at risk of the operation smaller than the population of the catchment area. When this occurs rates will look lower than they truly are. For instance, in the U.K. the lifetime risk of hysterectomy is around 20 per cent and increasing, while in the U.S. it is nearer 50 per cent. Thus the population hysterectomy rate among middle aged women is lower than the true rate among women at risk, because women are not at risk of another hysterectomy, and, in the U.S. in particular, such women constitute a large proportion of the denominator.

Finally, some cultures are more inclined to regard medical intervention as a panacea while others prefer to avoid it at all costs. Of course inaccuracies or omissions in data sources have to be guarded against before one can view hospitalisation rates as different or variable. In the U.K., the increasing private sector (McPherson *et al.*, 1985b) represents a problem because private hospitals do not contribute to routine health statistics.

In summary, the above discussion suggests that there are five major possible sources of variation for standardised rates:

- Morbidity
- Random
- Supply
- Clinical
- Demand

Most of these possible causes are not directly measurable or are extremely difficult to measure. It is not therefore generally possible to investigate correlations between rates of hospitalisation and, for example, morbidity or clinical opinion or demand. Morbidity rates among populations, unconfounded by differences in the supply of services or patient demand, are extremely difficult and expensive to measure. Similarly actual clinical practice or patient demands as opposed to stated hypothetical preferences are also difficult, if not impossible to measure. Hence the usual epidemiological scheme for inferring causation from carefully measured associations is not readily accomplished in this area.

TABLE 2.2 · SOURCES OF VARIATION IN RATES OF SURGERY BASED ON GEOGRAPHICAL AREAS

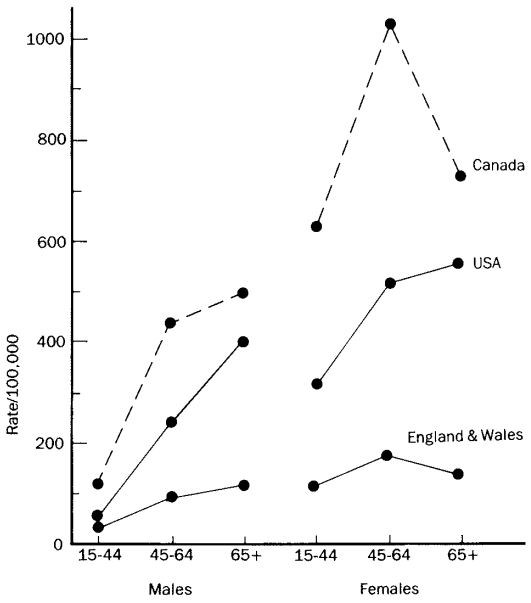
1. Age specific disease incidence
2. Age and sex composition
3. Random variation with time and place
4. Availability
   a) Manpower
   b) Hospital bed provision
   c) Funding
   d) Waiting list
   e) Methods of Payment
5. Clinical Judgement
6. Variations in patient demand or expectation
7. Rates of previous years: for organ removal
8. Prevailing custom
9. Inaccuracies in information sources
10. Systematic omissions of operations eg. private surgery, day cases

2.4 Examining Variations

Looking at cholecystectomy rates between Canada, the U.S. and the U.K. in the 1970s raises all sorts of interesting questions. Why was the rate four or five times higher in Canada than in the U.K. and among whom is it higher? Figure 2.2 shows the age and sex specific rates in the U.S., Canada and England and Wales in 1975. There are higher rates in both the U.S. and Canada compared with England and Wales for all

ages and both sexes. There is evidence here for instance that middle aged women in Canada were five times more likely to have the operation than in England and Wales and so were middle aged men. The steep fall in Canada for elderly women is worth noting. This should not be interpreted as a lower rate of intervention, for a rate among women aged 45-64 of 1,000 per 100,000 per annum is 1 per cent per year for twenty years which amounts to an 18 per cent overall risk in that period. Thus the population at risk when over 65 years is depleted and the true rate should be adjusted for the correct denominator. Inspection by age and sex in this way can give insights into the nature of variations.

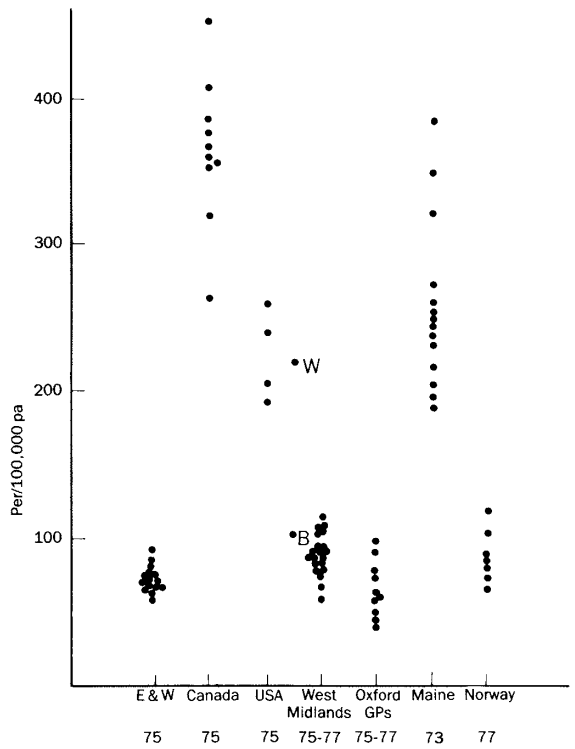
FIGURE 2.2 · CHOLECYSTECTOMY RATES BY AGE, SEX AND COUNTRY



As well as individual inspection, it is also interesting to compute the age standardised rates at different levels of population aggregation. Figure 2.3 shows cholecystectomy rates at seven such aggregations: regional health authorities in England and Wales, provinces of Canada, standard regions of the U.S., districts in the West Midlands Regional Health Authority, general practices in the Oxford District, hospital market areas of Maine and counties of Norway. What emerges is a lower prevailing rate in England and Wales and Norway than in North America and the observation that the lowest Canadian province (Newfoundland) has a rate two and a half times higher than the highest English region (Oxford). In the U.S. it is possible to estimate the rates for whites and blacks separately and the rates amongst blacks (B) in the U.S. look similar to the rates in England and Wales. Most importantly the variation in cholecystectomy rates in England and Wales implies a pretty tight consensus (McPherson *et al.*, 1984).

The important methodological point from these data is that the plausible sources of variation in rates at different levels of aggregation are different, and hence

FIGURE 2.3 · CHOLECYSTECTOMY



give indirect insights into their causes. If we examine the variation in, for example, cholecystectomy rates, between areas, there are a few discrete levels of aggregation. In the U.K. these are tabulated in Table

TABLE 2.3 · POSSIBLE LEVELS OF AGGREGATION (IN UK)

1. Between GPs in a single district
2. Between districts in a single region
3. Between regions in a single country
4. Between countries

2.3 and three of these are shown in Figure 2.3. Similar data are shown for hysterectomy, hernia repair and prostatectomy in Figures 2.4 to 2.6 respectively. When examining variation between regional health authorities, it is clear that supply, demand and morbidity are, in principle, plausible causes of this variation. This is because the populations of regions could exhibit specific regional characteristics and they could plausibly suffer different levels of morbidity because of characteristic life styles and exposures. Moreover, of course, the supply of medical services is largely determined by history, and supply is known to vary between regions. Because regional health authorities involve large populations, rates will be based on large numbers and random variation, except for very rare procedures, will be neglible. Also for the same reason, because the rates of admission will be the

FIGURE 2.4 · HYSTERECTOMY

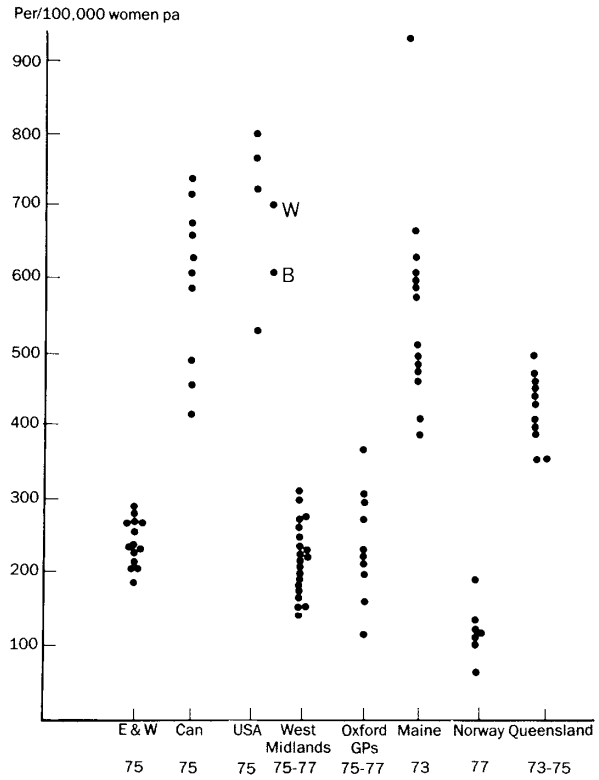


FIGURE 2.6 · PROSTATECTOMY

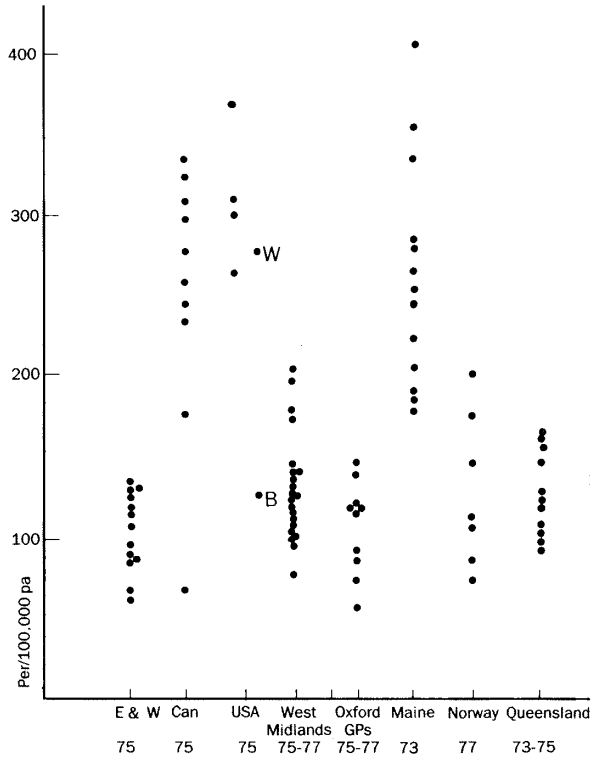
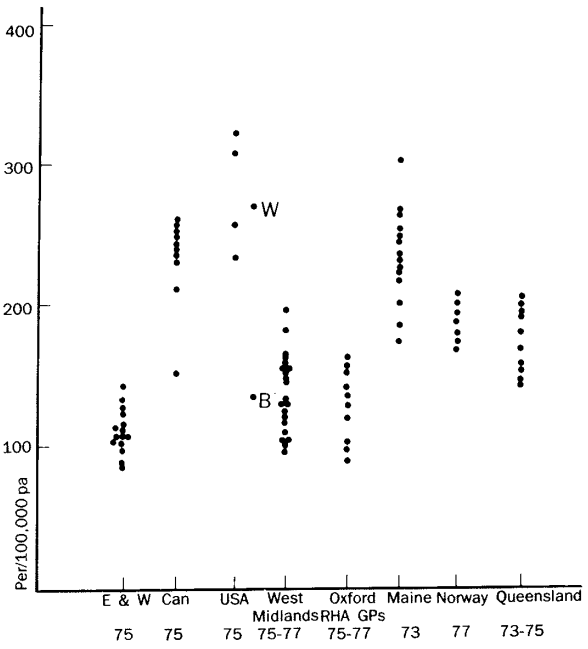


FIGURE 2.5 · HERNIA REPAIR



consequence of averaging many (possibly different) clinical opinions, such a source of variation is an implausible dominant cause of variation between regions. Clearly, however, if there were characteristic regional opinions or teaching this would not be so. Indeed, when examining variation between countries such a postulate becomes much more likely in spite of the averaging effect because such systematic differences in clinical opinion can be more easily sustained between countries.

Looking at inter regional variation, it is possible to compare the extent of this variation. Relative to the total variation observed for cholecystectomy, the variation between regional health authorities is extremely small (Figure 2.3). For hysterectomy, the rates are not quite so invariant, for prostatectomy they are more variable and for hernia repair they vary most of all. All these comparisons are made relative to the total variation observed and since hernia repair does not vary a great deal between countries (see Figure 2.1) this comparison may be different when we come to compare absolute variation. Clearly, though, it makes some sort of clinical sense in that indications are more tightly defined for hernia repair than for cholecystectomy, with prostatectomy and hysterectomy being intermediate. Thus cholecystectomy could be justified with silent stones for prophylactic purposes whereas an inguinal hernia is in general either present or absent, and if absent then no

surgery is indicated. If this holds true for other procedures, international variation could in part be a manifestation of differing characteristic clinical opinion.

In absolute terms, it is interesting to compare the range of rates between regions in the U.K.. This shows that prostatectomy is the most variable with a three-fold difference between the highest regional health authority and the lowest. The remaining three operations show a less than twofold variation. However, the causes of this variation are complex and difficult to disentangle. It turns out, for instance, that hernia and prostatectomy rates correlate quite highly with the difference between the regional allocation for funding for hospital and community services and RAWP targets. This might indicate that supply is the dominant determinant for the rates of these operations, but clearly such an inference is extremely problematic, because many other things could equally well influence this.

It therefore seems sensible to change the population aggregation level in order to make the examination of variation more straightforward. If we look at the variation between districts in a single region or between GPs in a single district then the plausible causes of variation are more restricted. This is because we are comparing populations who live and work, on the whole, in neighbouring communities. Massive differences in morbidity rates are less likely for these populations. On the other hand, since these communities are each served by relatively few specialists (sometimes only one or two), clinical opinion could have a significant impact on the decision to hospitalise. However, since the level of aggregation is lower, the number of events are fewer, and the random component in the variation of rates is larger. In Figures 2.3 to 2.6 data for the U.K. at this or lower levels of aggregation have been collected over three years to make the rates statistically more stable.

At this level of aggregation, it seems that prostatectomy shows threefold variation, hysterectomy and hernia repair show a little more than twofold variation, and cholecystectomy shows less than twofold variation. If the hypothesis about clinical opinion being the major source of variation is true, and if there is little important variation in the underlying morbidity rates, then prostatectomy is associated with a larger implied uncertainty concerning its indications than cholecystectomy. This is important if we want to identify causes of hopsitalisation for which greater uncertainty exists.

It is interesting that variation between GPs appears to be greater than between districts, with the exception of hernia repair. On the face of it, this might imply that GP uncertainty about referral is more important than specialists' decisions to admit. However, the random component of variation in these data is much larger than for variation between districts because the numbers are smaller. On the other hand, the supply component must be negligible because all the GPs are referring their patients to the same consultants in the same hospital. Interestingly, cholecystectomy and hysterectomy show the largest variation at this level of aggregation.

A summary of the important plausible sources of variation at different levels of aggregation is shown in Table 2.4. This indicates that there are no simple explanations of variations between regions or between countries. As yet there are too many plausible and unmeasurable causes.

TABLE 2.4 · PLAUSIBLE SOURCES OF VARIATION IN STANDARDISED RATES AT DIFFERENT LEVELS OF AGGREGATION

Variation Between	Morbidity	Clinical	Availability 'Supply'	Patient Demand	Errors in Data Sources	Random
GPs in one district	S	L	O	S	M	L
Districts in one region	S	L	L	S	S	M
Regions in one country	L	S	L	M	S	S
Countries	L	L	L	L	S	O

L = Large effect
M = Medium effect
S = Small effect
O = No effect

relative to others in same row

2.5 Measuring Variation

From the point of view of understanding the nature of health care provision, it is important to be able to compare the magnitude of the variation in the rates of surgery or other medical events between neighbouring areas or practices. Variation is traditionally measured by a sample estimate of variance of a set of rates, or equivalently by the standard deviation. But these measures of variability have important limitations for the comparison of variations among systems of care — such as comparisons of variation in tonsillectomy rates between hospital districts in the U.K. and hospital market areas in the United States — as well as for comparing the relative variation among the same population set. If one system or procedure has a prevailing rate that is substantially higher than the second, then the variation among small areas, measured by the standard deviation of the rates, is likely to be higher in the first case than in the second, simply because the prevailing rate is higher. Further, random variation, as a component of total variation, will be lower when the medical event is more common. Moreover, as the population size of the areas is itself variable (see Table 2.5) the random component will assume a varying proportion of the total variation in rates when comparisons are made between systems of care.

The measures of variation we need must therefore not be affected by artefacts of different prevailing rates or by large observed variations attributable to small numbers. We propose a simple method to measure and compare the systematic variation in standardised rates of events between defined populations. The model adopted is a simple development of proportional hazards or multiplicative relationships which enables comparisons that take account of the effect of different population denominators and numerators on the

TABLE 2.5 · POPULATION DENOMINATORS

		Sample	
1. Canadian Provinces	114,000 — 7 million	complete	1 yr
2. England & Wales RHAs	2 — 5 million	1:10	1 yr
3. USA (geographic regions)	37 — 67 million	incomplete	1 yr
4. W Midlands Districts	~ 200,000	complete	3 yrs
5. GPs	7,000 — 17,000	complete	3 yrs
6. Maine, USA	20,000 — 250,000	complete	1 yr
7. Norway	100,000 — 300,000	complete	1 yr
8. Queensland	3,000 — 730,000	complete	3 yrs

estimate of variance. The method is described in some detail in McPherson *et al.* (1982), and essentially one finishes up with a measure of extra-Poisson variance — that is systematic variation. If we look for instance at the prostatectomy rates in Figure 2.6, the variation between GPs in Oxford looks larger than the variation between districts in Oxford or regions in England and Wales. In fact when we adjust for the random component of variation in rates it appears that all the variation we see between GPs for prostatectomy is random and there is no systematic variation left.

Such a measure enables us to assert that variations between hospital districts for these operations are more determined by the operation than the country (or system of health care delivery) where they are observed. Some interesting exceptions are hernia repair (for which variation is high in England and Wales) and hysterectomy for which variation is high in the U.S. and Norway. If we look at Figure 2.1 it can be seen that the population rates for hysterectomy are high in the U.S. and low in Norway, and for hernia low in England and Wales. Thus high variation can equally occur when rates are low and high.

Since the main determinants of variation at the hospital service area level of aggregation are clinical judgement, large variations might be a manifestation of clinical uncertainty concerning the optimal indications. Hence such evidence provides a priority rating for research effort into the outcomes associated with different levels of intervention.

3 VARIATIONS IN LENGTH OF STAY AT DISTRICT LEVEL

Myfanwy Morgan

3.1 Introduction

An important trend in hospital use has been the decline in average lengths of hospital stay. For example, in England and Wales the mean length of stay for general surgical patients declined from 8.9 days in 1974 to 7.1 days in 1984. This reduction in length of stay has occurred across all age groups and for most operative procedures and medical diagnoses. However, there are questions of the extent to which length of stay varies between geographical areas, the causes of these variations, and their implications for resource use.

We examined the variations in length of stay at district level for three surgical procedures: inguinal hernia repair, appendicectomy and cholecystectomy (Morgan *et al.*, 1987). These are frequently performed procedures which encompass a range of age groups and include both planned and emergency admissions. An analysis by McPherson *et al.* (1982) indicates that the variation in age/sex standardised admission rates, both between and within countries, for these procedures is much smaller than for other common operations, including tonsillectomy, hysterectomy and prostatectomy. This suggests that differences in length of stay will make an important contribution to the total geographical variation in hospital use for the three study procedures.

The analysis was based on data recorded in the Hospital Activity Analysis (HAA) for the 15 districts in South East Thames RHA (SETRHA) (1982) and the 16 districts in Northern RHA (NRHA) (1984). In each region patients were classified by their district of treatment. To increase the homogeneity of the cases studied, two types of exclusions were made. Firstly, admissions for which a second procedure was recorded were excluded, with the exceptions of appendicectomy with laparotomy and cholecystectomy with cholangiography. These exclusions were most common for cholecystectomy: 24 per cent of cholecystectomy cases were excluded in each region, compared with about 8 per cent of appendicectomies. Secondly, inguinal hernia repairs were excluded if recorded as a day case.

3.2 Length of Stay Distributions

The mean length of stay distributions for the two regions were very similar at ages under 75 years. They also displayed the usual pattern of increasing length of stay with rising age (Table 3.1). For example, in SETRHA (1982) the mean length of stay for appendicectomy rose from 5.0 days in the under 25 age group to 10.0 days for patients aged 65-74. For inguinal hernia repair the mean length of stay increased from 3.0 to 6.0 days and for cholecystectomy from 9.3 to 12.7 days.

The strong relationship between age and length of stay required that differences in age were allowed for in making comparisons between districts. An age adjusted score was therefore calculated which measured the distance of each patient from the regional mean for his age group for that operative procedure.

Analysis of variance showed there were significant differences between districts in their mean age adjusted score for each operation. Districts were therefore ranked for each operation according to their length of stay score and the three districts with the lowest length of stay score (shortest stay) and the three districts with the highest score (longest stay) were selected for more detailed study. In general, the districts classified as shortest or longest stay varied for different procedures with only one district being classified as shortest stay for all three procedures and two districts classified as longest stay (Table 3.2). However, there was a tendency for teaching districts to have longer stays, especially in SETRHA. Two SETRHA teaching districts ranked as longest stay for one procedure and one district for three procedures, while the one teaching district in NRHA ranked as longest stay for one procedure.

Comparison of the two regions showed that the length of stay distributions of their three shortest stay districts were very similar, as were the length of stay distributions of their three longest stay districts. However, within each region there were marked differences between the longest and shortest stay districts. As Figure 3.1 shows, the principal difference

TABLE 3.1 · MEAN LENGTH OF STAY BY AGE AND OPERATIVE PROCEDURE

Age groups	South East Thames RHA, 1982			Northern RHA, 1984		
	Appen-dicectomy	Inguinal hernia	Chole-cystectomy	Appen-dicectomy	Inguinal hernia	Chole-cystectomy
Under 25	5.0	3.0	9.3	4.9	2.7	8.5
25-44	5.5	4.8	9.6	5.5	4.4	9.2
45-64	7.2	5.5	10.7	7.5	4.7	10.5
65-74	10.0	6.0	12.7	10.0	6.0	12.9
75 & over	11.6	7.9	14.5	15.3	8.2	16.5
All ages (No. cases)	5.5 (4407)	5.4 (4687)	11.1 (1879)	5.5 (3694)	4.9 (3705)	11.0 (2287)

TABLE 3.2 · NUMBERS OF DISTRICTS RANKED AS SHORTEST AND LONGEST STAY FOR THE THREE SURGICAL PROCEDURES

	SETRHA (1982)	NRHA (1984)
Shortest stay		
0 procedure	10	11
1 procedure	1	2
2 procedures	4	2
3 procedures	—	1
Longest stay		
0 procedure	8	10
1 procedure	6	4
2 procedures	—	1
3 procedures	1	1

was that a large number of patients in the longest stay districts were spending one or two additional days in hospital. As a result, in SETRHA shortest stay districts 62 per cent of appendicectomy admissions aged under 45 years and 30 per cent aged 45-64 years had lengths of stay of 4 days or less, compared with 32 per cent and 14 per cent respectively in the longest stay districts (Figure 3.1). Similar differences in lengths of stay between the shortest and longest stay districts occurred for inguinal hernia repair and cholecystectomy. This held for patients aged 65 years and over as well as for the younger age groups (Figures 3.2 and 3.3).

Although the length of stay distributions in the longest stay districts showed an overall shift towards greater lengths of stay, even in these districts the proportion of very long stay patients was quite small. For example, lengths of stay of 29 days and over comprised 3.5 per cent of cholecystectomy admissions in the three SETRHA longest stay districts and 1.4 per cent in the three shortest, with the proportions being very similar in NRHA.

Another way of viewing the variations in hospital use is to compare the observed lengths of stay with the regional average, albeit recognising that this does not necessarily represent good practice. In both regions the observed length of stay for each procedure in the longest stay districts was at least 13 per cent above their regional average and at least 13 per cent below in the shortest stay districts. The range was considerably greater for inguinal hernia repair in NRHA (Table 3.3). Translation of these differences into numbers of bed days indicates that about 2000 *more* bed days were occupied than expected for the three procedures in the longest stay districts in each region and 2000 *fewer* bed days in the shortest stay districts.

Differences in the length of stay distributions between districts were partly due to differences in the

FIGURE 3.1 · LENGTH OF STAY IN SHORTEST AND LONGEST STAY DISTRICTS *(SETRHA, 1982)*

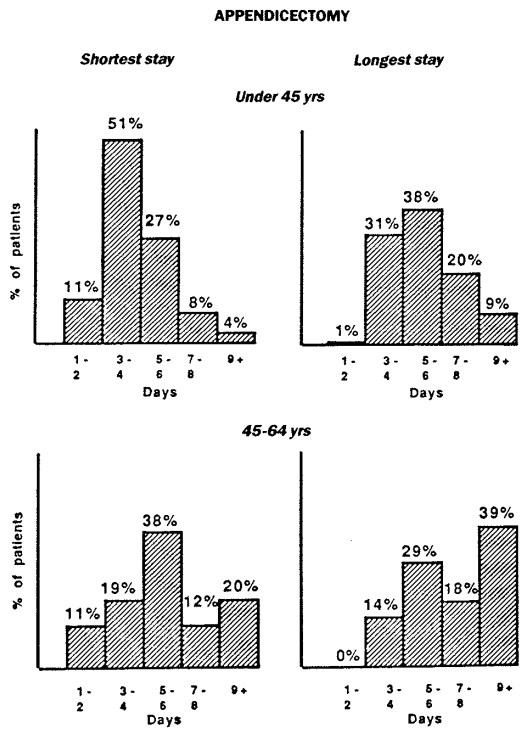


FIGURE 3.2 · LENGTH OF STAY IN SHORTEST AND LONGEST STAY DISTRICTS *(SETRHA, 1982)*

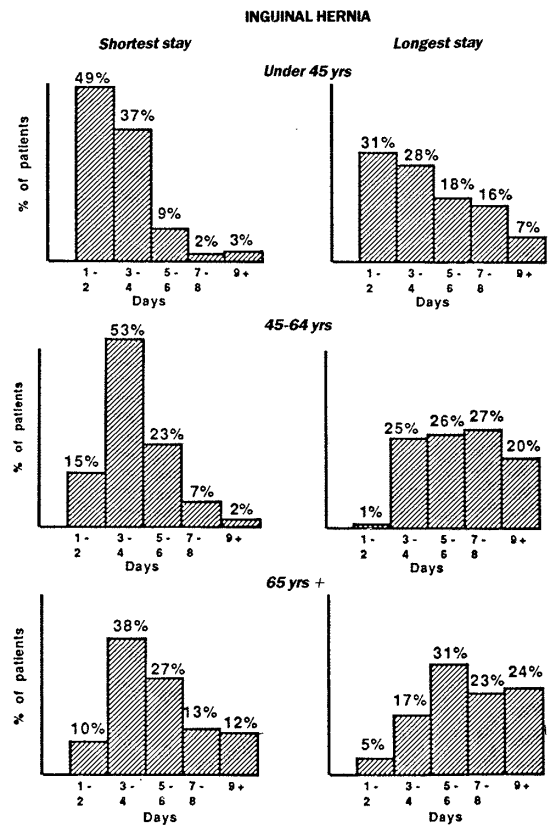
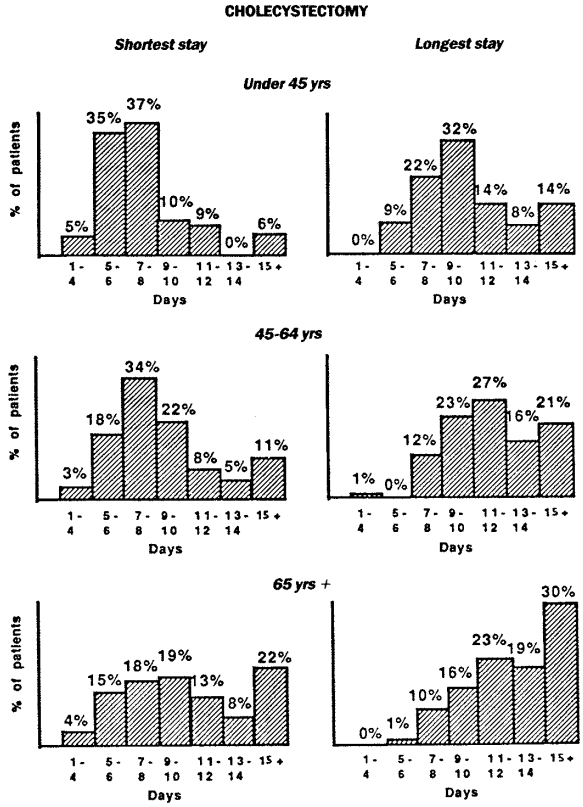


FIGURE 3.3 · LENGTH OF STAY IN SHORTEST AND LONGEST STAY DISTRICTS *(SETRHA, 1982)*



length of the pre-operative period. For example, only 3 per cent of hernia admissions in the three shortest stay districts in SETRHA had a pre-operative stay of 2 days or more compared with 19 per cent in the three longest stay districts. The per centages for cholecystectomy admissions were 16 per cent in the shortest stay districts and 50 per cent in the longest stay. However, despite these variations in the pre-operative length of stay, the greatest contribution to the overall differences in hospital use was made by the post-operative period.

3.3 Causes and Implications of the Observed Variations

Several types of explanation need to be considered in seeking to explain the observed variations in length of stay between districts. These include artefactual explanations in terms of differences in recording or case mix; differences in patients' needs for nursing care and in the supply of beds and other facilities; and differences in hospital organisation and clinical practices.

Artefactual explanations

One possibility is that the observed variations in length of stay reflect differences in case mix. However, the proportion of cases excluded due to a second procedure was almost identical in the longest and shortest stay districts, suggesting this was not a major source of bias. Cases of strangulated hernias and drainage of appendix abcess, which form two sub-groups that are normally associated with higher complication rates and greater lengths of stay, were also found to each account for less than 1 per cent of admissions. In addition, there was no evidence of the concentration of these conditions in the longest stay districts. Thus while possible differences in the severity of cases treated cannot be entirely excluded, the available evidence suggests this is unlikely to have introduced a major bias.

Another possible explanation of the observed variations is that they reflect inaccuracies in the recording and coding of HAA data (Butts and Williams, 1982). Problems of the completeness and accuracy of HAA data is known to vary for different items. Data on for example waiting list times, which is particularly subject to variation, have therefore not been included. It is however likely that the recording of common surgical procedures is far less subject to error than more complex conditions and diagnoses, while only a large and consistent bias would seriously affect the results obtained.

Supply and demand

The supply of acute hospital beds and medical manpower are important factors explaining variations between districts in total acute hospital admission rates and lengths of stay (London Health Planning Consortium, 1979). In the present study, a comparison was made between the ranking of districts on an

TABLE 3.3 · OBSERVED LENGTH OF STAY FOR THE LONGEST AND SHORTEST STAY DISTRICTS AS A PERCENTAGE OF THE REGIONAL AVERAGE (ALL AGES)

	Appendicectomy	Inguinal hernia	Cholecystectomy	Bed days involved (No.)
3 longest stay districts				
South East Thames RHA	114	116	115	+1980
Northern RHA	113	125	115	+2185
3 shortest stay districts				
South East Thames RHA	87	82	86	−2176
Northern RHA	83	75	85	−1979

average length of stay score (based on the average of their ranks for the three procedures) and the numbers of beds per general surgeon (WTE). Although a fairly crude analysis, this showed the expected pattern in non-teaching districts of a relatively plentiful supply of beds, and hence less pressure on their use, being associated with a lower throughput and greater length of stay. However, in teaching districts the length of stay was greater than expected in relation to the numbers of beds per WTE consultant. This may reflect the requirements of teaching and the reduced case load of some academic appointments. Teaching hospitals also tend to be concentrated in deprived inner city areas whose patients may be viewed as having greater social needs for hospital care.

Another supply factor influencing length of hospital stay is the availability of alternative residential facilities and community services. In particular, the opportunity to discharge patients requiring continued nursing care to convalescent hospitals and units has a direct effect in determining 'need' for an acute hospital bed. In SETRHA (1982) only 2.7 per cent of admissions for the three study procedures were discharged to other hospitals or convalescent units, and 1.9 per cent of admissions in NRHA (1984). However, although only a small per centage of patients in each region were transferred to other hospitals and units, there was a high transfer rate in one district in each region. One district in SETRHA (classified as shortest stay for both appendicectomy and cholecystectomy), discharged 19 per cent of hernias, 20 per cent of appendicectomies and 31 per cent of cholecystectomy admissions to other hospital or convalescent units, with a correspondingly smaller proportion discharged directly home. Similarly, one district in NRHA discharged 8 per cent of hernias, 6 per cent of appendicectomies and 27 per cent of cholecystectomy admissions to another hospital or unit. Transfers occurred among all age groups, but were most frequent among elderly patients. The HAA treats each admission and discharge as a separate period of hospital care, while its coverage does not extend to convalescent hospitals. The transfer of patients to these facilities would thus significantly reduce the recorded length of stay in the two districts with high transfer rates.

On the demand side, the social conditions of the population and especially the presence of deprivation, is regarded as increasing patients' needs for hospital services, due both to their greater morbidity and their unfavourable home circumstances (poor housing conditions, living alone, etc). However, although district populations can be compared in terms of their social characteristics and measures of deprivation developed, the extent to which this is reflected in hospital utilisation rates will depend on clinical assessments of need and on the supply of beds and other facilities. It is thus difficult to determine the effects of the social conditions of the population on lengths of stay and needs for health care using aggregate data (Morgan, Mays, Holland, 1987). The small number of patient reviews that have been undertaken provide more direct evidence on the role of social factors on rates of hospital use. They indicate that patients' home circumstances often only have a marked effect on their length of stay if they also have a high level of dependency and other medical problems. These effects are thus concentrated among elderly patients (Coid and Crome, 1982). This suggests that although the social circumstances of patients may have contributed to the observed variations in lengths of stay between districts, this is unlikely to account for the consistent differences in lengths of stay among all age groups.

Clinical and organisational practices

The practice style of individual clinicians has been identified by Wennberg and colleagues as being of central importance in explaining the large geographical variations in hospital admission rates for conditions whose management is surrounded by considerable medical uncertainty (Wennberg *et al.*, 1987b). These variations in practice style also have implications for the length of patients' hospital stay. They include differences in the extent to which pre-operative tests and investigations are conducted on an in-patient basis, general norms regarding the appropriate length and management of the post-operative period, as well as the effects of different surgical techniques and types of anaesthesia which have made a major contribution to the overall decline in length of stay (Sloan and Valvona, 1986).

Interacting with, and sometimes determining, the variations in clinical management are the organisational characteristics of the hospital and individual ward (Eastaugh, 1980). These include the routines for admission and bed allocation procedures. For example, where a flexible pool of beds is available, rather than a fixed quota of beds for individual firms or specialties, this may discourage the discharge of patients before others are admitted. Other determinants of length of stay are the extent of discharge planning and social work support, the availability of theatre space and staff, and the time taken for laboratory tests and x-ray investigations. Griffith, Waters and Acheson (1979) in one of the few studies of the determinants of variations in the length of post-operative stay found that for inguinal hernia repair the mean post-operative stay was similar for consultants at any one of the eight hospitals studied but was significantly different for consultants who operated at more than one hospital. This led them to conclude that for this operative procedure the characteristics of the hospital appear to exercise a greater influence in determining mean post-operative stay than does the individual consultant. However, the situation may vary for other procedures and medical conditions where less consensus exists as to their appropriate management.

3.4 Conclusions

This analysis of routinely available HAA data has identified substantial variations in lengths of hospital stay between districts for three common surgical procedures. These differences occurred among all age groups and mainly arose from large numbers of patients each spending one or two more days in hospital in the longest stay districts.

Probably of particular importance in explaining the observed variations are differences on the supply side,

in terms of the availability of acute hospital beds. In addition, in one district in each region the opportunity of transferring patients to other residential facilities reduced the patients' length of stay recorded in the HAA. However, the considerable variation between districts in their length of stay scores for particular procedures, suggests that differences in the organisational and clinical practices of individual hospitals and surgical firms may also form an important source of variation. Further studies are required to determine the precise effects of these factors in contributing to the observed variations. However, even in the absence of such information, the provision to districts of comparative length of stay profiles for selected procedures could provide useful benchmarks for monitoring performance which avoids the case mix problem of specialty based performance indicators. This may encourage clinicians to review their own practices. In particular, the evidence of lengths of stay which are markedly higher than the regional average across all age groups suggests that in these districts there is scope for reducing the length of hospitalisation. However, there are also questions of whether patients in the shortest stay districts are being discharged inappropriately early, due possibly to pressures on beds. As a result, are such patients experiencing difficulties in coping at home and placing heavy demands on general practitioner and community services, and are short lengths of stay associated with higher complication and readmission rates?

Underlying the concern that some patients are spending 'unnecessarily' long in hospital are questions of the appropriateness of resource use. Assuming that a shorter length of stay has no detrimental effects on patients this is generally seen as economically desirable, since it will reduce the cost per case of hospital treatment. If this is associated with a stable admission rate and reduction in beds it will also serve to reduce total hospital costs. However, if the reduction in length of stay is accompanied by increased numbers of admissions and a higher throughput, greater demands will be placed on hospital resources, in terms of the medical and nursing staff, theatre time and other facilities. This more intensive use of hospital beds will generally increase total costs (Jonsson and Lindgren, 1980). Reducing lengths of stay, although producing a more 'efficient' use of hospital beds, may therefore have major financial implications for a centrally funded health service such as the National Health Service if accompanied by a greater throughput of patients.

4 REGIONAL AND DISTRICT VARIATIONS IN PERIOPERATIVE DEATHS

Nigel Buck

4.1 Introduction

The Confidential Enquiry into Perioperative Deaths (CEPOD) studied three regions during 1986. The regions included in the Enquiry were the Northern Region, the South Western Region and the North East Thames Region. Over a 12 month period, deaths within 30 days of an operation were reported to CEPOD via a network of local reporters at each hospital. These reports were followed by a questionnaire seeking information about the surgical and anaesthetic care provided for each patient. Each questionnaire was then assessed in a blind fashion by relevant experts. Full details of the methodology used are contained in the CEPOD report (Buck, Devlin and Lunn, 1987).

CEPOD received 4,034 reports of deaths from the three regions. Table 4.1 shows that a similar number of deaths occurred in each region. Using data gathered from the Hospital Activity Analysis, the table also shows that the total number of operations carried out during the same period was 219,402 in North East Thames, 192,421 in the Northern Region and 143,435 in the South Western Region.

TABLE 4.1 · REPORTED DEATHS AND OPERATIONS

	CEPOD Reported deaths	HAA Operations
Region		
Northern	1359	192,421
South Western	1307	143,435
North East Thames	1368	219,402

4.2 The Enthusiasm of Consultants

There were differences between regions in the enthusiasm and ability of consultants to participate in the Enquiry. This is illustrated in Table 4.2. Overall, 69 per cent of surgical forms were returned, but the range of response was from 58.5 per cent in North East Thames to 78.9 per cent in Northern. Similarly, 72.6 per cent of anaesthetic forms were returned, but the range of response was from 63.2 per cent in North East Thames to 79.5 per cent in Northern.

TABLE 4.2 · RESPONSE RATES (%)

	CEPOD Cases Consultants Withdrawn	CEPOD Surgical Forms Returned	CEPOD Anaesthetic Forms Returned
Region			
Northern	9.1	78.9	79.5
South Western	14.0	69.7	75.1
North East Thames	15.5	58.5	63.2
TOTAL	12.9	69.0	72.6

Table 4.2 also shows that 12.9 per cent of cases were lost to the Enquiry through the non-participation of consultants and North East Thames was again the region most affected.

The differences revealed by CEPOD were even greater at district level. This is illustrated in Tables 4.3, 4.4 and 4.5. These tables show for the three regions variations in the number of reported deaths and in the number of cases where consultants withdrew. As the tables illustrate, despite the positive response to the Enquiry in general, there were pockets of resistance in a few centres. Three districts (District 13 in North East Thames, District 12 in South Western and District 11 in Northern) accounted for half of the cases lost to CEPOD.

TABLE 4.3 · NORTH EAST THAMES PROFILE

District/ Hospital Number	CEPOD Reported Cases	CEPOD Cases Consultants Withdrawn
1	37	4
2	14	2
3	73	35
4	66	3
5	38	5
6	82	0
7	40	5
8	4	0
9	5	0
10	47	0
11	15	3
12	109	0
13	167	98
14	57	0
15	54	0
16	154	10
17	133	33
18	83	0
19	109	15
20	76	0
21	5	0
22	0	0

4.3 The Findings

One finding of the Enquiry was that there were marked differences between districts in the proportion of deaths considered by the assessors to be avoidable. This is illustrated in Figure 4.1. A more detailed analysis is presented for each region in Figure 4.2. The variation between regions was much smaller and this is accounted for by the way in which high and low values are averaged when they are combined into an overall figure.

There were significant differences between regions in how clinical care was organised. Table 4.6 shows that in 72.4 per cent of cases in the Northern region the most senior surgeon to take the history before the operation was a consultant compared with 59.4 per cent in South Western and 53.6 per cent in North East

TABLE 4.4 · SOUTH WESTERN REGION PROFILE

District/ Hospital Number	CEPOD Reported Cases	CEPOD Cases Consultants Withdrawn
1	46	0
2	100	0
3	108	4
4	138	11
5	14	0
6	161	3
7	180	0
8	101	30
9	105	0
10	141	24
11	67	46
12	146	65

TABLE 4.5 · NORTHERN REGION PROFILE

District/ Hospital Number	CEPOD Reported Cases	CEPOD Cases Consultants Withdrawn
1	9	0
2	192	10
3	57	0
4	52	0
5	57	17
6	122	0
7	37	0
8	79	0
9	97	1
10	38	4
11	69	42
12	91	2
13	120	0
14	96	4
15	15	0
16	60	0
17	47	1
18	60	34
19	61	9

Thames. There were similar differences in the most senior surgeon to examine the patient before the operation, the grade of the surgeon who started the operation, the grade of the surgeon who undertook the operation and the quality of notes.

4.4 Conclusions

CEPOD has demonstrated the feasibility and value of peer audit on a large scale. For any similar studies in the future, it is vital that the enthusiasm and trust of consultants is maintained. It is likely that the next stage of CEPOD will involve all health regions in England and Wales. A five year rolling programme is being devised. In the first year the focus is likely to be on all deaths among children under the age of 10. There will also be an analysis of cases where the patient did not die. Subsequently, attention is likely to concentrate on particular types of operation and deaths which occur out of hours.

FIGURE 4.1 · PERCENTAGE OF DEATHS ASSESSED AS CONTAINING "AVOIDABLE" ELEMENTS (SURGEON ASSESSORS' OPINION)

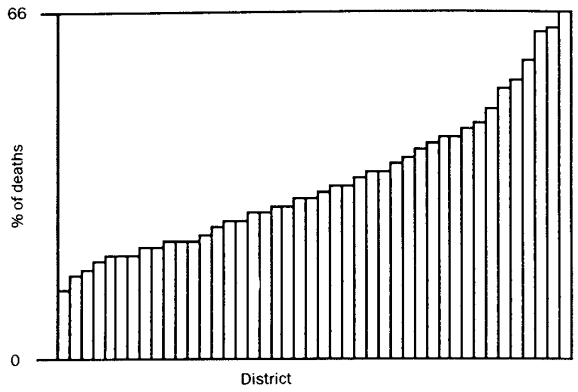


FIGURE 4.2 · AVOIDABLE DEATHS BY REGION AND DISTRICT

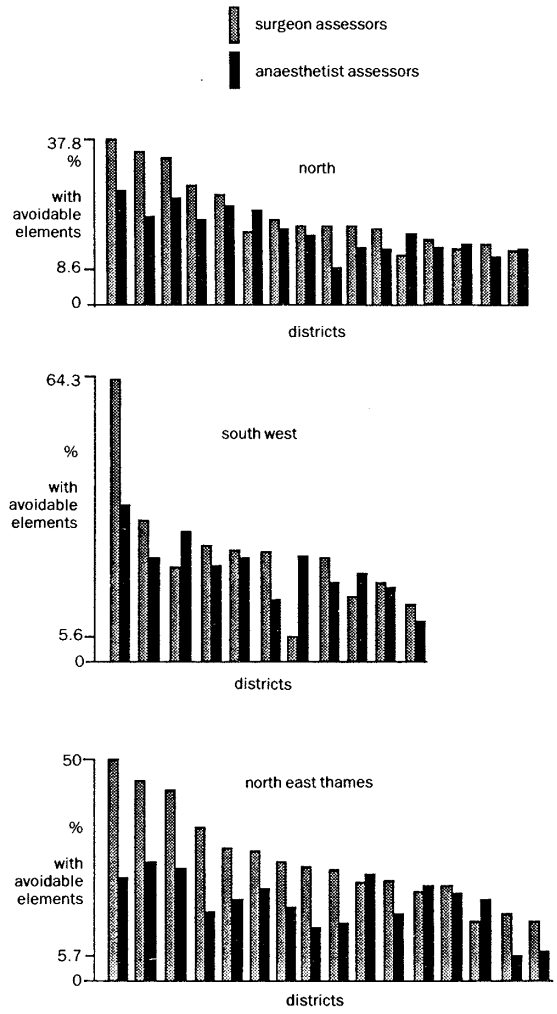


TABLE 4.6 · MOST SENIOR SURGEON WHO TOOK HISTORY BEFORE THE OPERATION (% by region)

Grade of surgeon	SW	N	NET	TOTAL
Senior House Officer	7.0	4.8	3.0	5.0
Registrar	21.7	12.5	31.7	21.0
Senior Registrar	5.6	6.4	9.0	6.9
Consultant	59.4	72.4	53.6	62.8
Associate Specialist	23.0	0.0	0.0	0.0
Student	0.0	0.0	0.0	0.0
House Surgeon	2.1	1.5	1.1	1.6
Other	0.0	0.0	0.0	0.0
Not answered	1.9	0.9	0.7	1.1

5 VARIATIONS DATA FROM SURGEONS, FOR SURGEONS

Bryan Jennett

5.1 Introduction

Data about variations in population-based rates of surgery are of limited interest to surgeons. Measures of process such as the length of the waiting list or of stay in hospital are often mistrusted because the data sources are intuitively considered suspect. Surgeons would do well to collect more data themselves rather than responding defensively to inadequate data from others. Surgeons are more likely to react constructively to data that have been produced by other surgeons and that deal with matters that both seem clinically relevant and could be altered by their own actions. For surgeons are men of action who are used to being turned to by colleagues and patients in the expectation of action; and most surgeons are naturally interested in the outcome of their actions. The benefits of successful surgery are obvious — the relief of symptoms, life saved and years of quality life added. Yet there is a suspicion that surgery is sometimes unnecessary or unjustified. There is also the certainty that surgery is sometimes unsuccessful, in that it fails to cure, or it leaves the patient more disabled than he was, or dead.

Evidence about variations in surgical practice should arouse the curiosity of surgeons and stimulate them to find out more about what they are doing. They may do so by establishing prospective audit within a unit, a hospital or a district; and by initiating reviews of practice on a national scale by expert groups with the aim of assessing the appropriate place for various procedures and practices, based on synthesis of the best data available. This may identify areas of activity where more formal assessment procedures are required in order to clarify uncertainties. This in turn should lead to more explicit definition of the indications for surgery and to declared goals for outcome, both of which can be expressed in written guidelines.

Audit at unit level is likely to discover variations between different places in several aspects of surgical activity. One is in the balance of work, by type of disease, proportion of emergency versus elective admissions and operations, proportion of patients coming from the district or from more distant places, and the proportion of routine cases as distinct from those reflecting the special interests and skills of local surgeons. Other variables are efficiency factors, such as length of stay and intensity of use of facilities; and variations in the procedures used and the outcomes achieved for specific conditions. Reviews on a broader base may indicate variations not only between places but over time and between different professionals. The big gap in variations data is in outcome, and when this is filled it should provide a clue to whether high or low rates are appropriate. If comparisons in outcome after surgery are to be valid, however, assessment will need to be more systematic than it often is. It should consider not only survival but the degree and duration of improvement, including the rate of relapse due to the waning of the placebo effect of surgery or from recurrence of disease.

5.2 Evidence from Surgeons

Surgeons in several specialties in the UK have already published data on various aspects of surgical practice that could be considered as contributions to the literature on variations — although this word does not appear in the titles. Earlam (1984) used Hospital Activity Analysis to review patterns of management of oesophageal cancer in North East Thames. The variations in practice that this disclosed have led to the setting up of a trial by the Medical Research Council. A higher mortality for prostatectomy in district as compared with teaching hospitals has been reported more than once from such routine statistics (Lee *et al.*, 1957; Anon, 1980). Prospectively collected registers provide only limited information but more than routine statistics and they are usually more reliable. The UK Cardiac Surgery Register, begun by cardiothoracic surgeons themselves in 1977, has revealed wide regional variations between the rates for various types of operation (English *et al.*, 1984). It was part of the agreement needed to persuade all surgeons to cooperate in this endeavour that mortality and complication rates associated with individual units would not be published. It revealed however that mortality ranges from less than 2 per cent to more than 20 per cent . Wide regional variations in the provision of facilities for the treatment of renal failure have been revealed by data collected by nephrologists as part of the activity of the European Dialysis and Transplant Register (Dowie, 1984). Success rates after renal transplantation have been shown to vary widely between centres (Taylor *et al.*, 1985). It was review of breast cancer registers in the US that first led to the suspicion that radical mastectomy was an ineffective procedure which in turn led to formal trials that confirmed this.

Surveys of current practice can provide an insight into the lack of consensus in action even for common conditions. A questionnaire about the treatment of breast cancer by 454 general surgeons in UK revealed wide variations not only in the kind of surgery done but also in the use of radiotherapy and chemotherapy (Gazet *et al.*, 1985). Some commentators regarded these variations in practice as not only to be expected but to be welcomed as evidence that clinical freedom was alive and well in the NHS. It does seem inherently improbable however that patients receiving such widely divergent care are all getting optimal care. Routine surgical audit may be regionally based, as in the Lothians where every general surgical unit has a BBC micro (Gruer *et al.*, 1986). District audit may reveal differences between adjacent teaching and district general hospitals; the higher post-operative mortality in the DGH in one district led to the establishment of an intensive care unit (Gilmore *et al.*, 1980).

Prospectively collected data bases according to strict protocols, with clear definitions and good follow-up, provide powerful information about variations in disease patterns, management and outcome. This may emerge as a by-product of a randomised trial, as

happened for breast cancer. Or it may have been set up as a prospective audit of surgery, as in the bowel cancer project that was based on collaboration between 84 surgeons. This showed a six times variation in the breakdown of the anastomosis, a complication associated with a trebling of mortality and a doubling of the length of stay (Fielding *et al.*, 1980). The reason for this variation appeared to be the surgeon in charge of the case, who determined not only what happened during the operation but also the pre-operative and post-operative care. If all the surgeons had obtained the results of the best surgeons, it was estimated that 450 deaths a year would have been avoided and there would have been a saving of £4million a year.

The international data base of severe head injuries began 19 years ago in Glasgow and now has more than 3,000 cases from Glasgow, the Netherlands and California. There were wide variations in the rate of use of several interventions between these three places, the only feature that did not vary being the mortality rate (Jennett *et al.*, 1980). Such a finding suggests that variations in management did not influence outcome, giving scope for considerable savings of effort and expense without losing benefit. The potential for using variations for the assessment of technologies has as yet been largely unrealised — yet it may prove particularly useful for surgery. It calls for correlating variations in the application of different therapeutic processes to patients who have been matched for prognosis and then analysing variations in outcome in relation to differences in process.

Wennberg's challenge to surgeons, culled from his various papers, appears to be that the clinical hypotheses underlying their work are weak, implicit, often untested and largely individualistic (Wennberg, 1984). He suspects that the decision rules of surgeons are based largely on traditional teaching and on schools of thought, tempered by the idiosyncracies of the individual surgeon. He points out that surgeons frequently disagree about the objectives of treatment, about the facts of surgical practice and about the interpretation of these. For these various reasons, surgeons forfeit their right to rely on the 'rational agency hypothesis' proposed by the economist Arrow. This states that matters associated with decisions about clinical management are too complex for patients to understand, and they should be left to doctors who 'know'. However variation studies indicate that surgeons often do not know enough either about the probabilities of various outcomes, or about patients' preferences for different outcomes. Consequently surgery may often be inappropriate, a term that is preferable to unnecessary. Moreover appropriateness should apply to the decision to operate, not judged retrospectively once the outcome is known. Not every surgical death or complication indicates an inappropriate decision to operate nor an incompetently carried out procedure. It is a question of calculated risks, of balancing the probabilities of benefits and burdens, that justifies embarking on a surgical procedure.

5.3 The Role of Guidelines

If variations data are expected to influence the activities of the participants in clinical decision making, rather than being regarded as a spectator sport for epidemiologists, it is necessary to consider how to alter practice. The most obvious means would be to produce clinical guidelines. Such guidelines have been produced for the management of head injured adults by a national group of neurosurgeons convened by the King's Fund; where these have been adopted admission rates have fallen and mortality reduced. Under the fetching acronym CRUG (Clinical Resource Use Group), the Scottish Home and Health Department is currently addressing this matter. Endorsing a health authority paper rationalising the use of gastrointestinal endoscopy, the Chief Medical Officer commended the guidelines that emerged. In his letter to all other health authorities he commented:

> *The endoscopy group accepted, albeit with some reluctance, the need for a 'preferred protocol', thus grasping a particularly difficult issue.*

The value of guidelines is that they encourage explicit discussion and the declaration of both the goals of surgical treatment and the methods best suited to achieving these. Moreover they enable consultants to influence practice by junior staff and they provide a standard for audit. Such outcome audit may indicate a need to alter the content of guidelines and this is one answer to the criticism of those who suspect that guidelines may fossilise or freeze practice and inhibit innovation. In fact it is variation in the use of established technologies rather than the problem of innovation that has to be taken in hand if surgeons are to fulfil the expectations of their patients and of the colleagues who have referred them for a surgical opinion (Jennett, 1985).

6 PRACTICE VARIATIONS AND THE NEED FOR OUTCOMES RESEARCH

John E Wennberg

6.1 Introduction

How can we take steps to reduce professional uncertainty about the scientifically and ethically best way to practise medicine? Professional uncertainty about appropriate practice exists because we simply have not given enough attention to the evaluation of the theory behind everyday clinical practice. Too often, physicians simply do not know what works best to improve the health of their patients. Systematic studies to evaluate outcomes, similar in purpose to those we now routinely apply to the evaluation of new drugs, are needed to test the various reasons for doing surgery, for performing diagnostic tests and for using expensive modalities of care such as hospitals and intensive care units. The usefulness of drugs prescribed for non-approved uses also needs to be brought under scrutiny.

Practice variation studies reveal the large difference in costs and utilisation resulting from professional uncertainty. Policy makers are increasingly attracted by the spectacle of practice variations, and many succumb to the temptation to assert that expenditures and utilisation over the minimum represent unnecessary care or discretionary spending. The problem is that no one really knows which rate of utilisation is right. Outcome studies are thus needed to restore the prerogative of the medical model, that is to make it possible for resource allocations to be based on need and expected outcomes. Outcome studies are also needed because without knowledge of the effects of care, the 'true demand' for health care cannot be known. Without such knowledge, patients cannot express their preferences for treatment, and debates about unnecessary service, underservice or rationing must be held in ignorance of the true impact of health care on health.

6.2 Practice Style and Small Area Variations

A more accurate term than 'small area variations' would be a longer phrase such as 'studies of variation in use rates among populations that receive all or most of their care from one or only a few discrete provider organisations'. Such a term would serve as a constant reminder of the unique feature of small area studies that permits the effects of the decisions of a small number of physicians to be observed in the rates of service of defined population groups — and thus to make transparent the importance of differences in practice styles for the cost and utilisation of care. The classic study of small area variations is Glover's 1938 study of the impact of school health officer's practice style (or 'medical opinion', as he phrased it) on the tonsillectomy rates among British school children living in different districts:

> *Comparisons of some of the rates in different areas . . . revealed striking contrasts in areas apparently somewhat similarly circumstanced . . . the operation rate in Margate was eight times that in Ramsgate; that of Enfield was six times that of Wood Green and four times that of Finchley; that of Bath five times that of Bristol; that of Guildford four times that of Reigate; that of Salisbury three times that of Winchester* (Glover, 1938, p.1224).

In a later paper Glover added:

> *Great variations in local incidence . . . appear to depend almost entirely upon medical opinion in the individual area* (Glover, 1948, p.1).

Other examples are studies comparing the utilisation rates among people enrolled in health maintenance organisations, lists of patients belonging to general practices or among the citizens living in different communities and receiving most of their care from local resources such as the residents of Oxford, Cambridge, Boston or New Haven.

When utilisation is measured among populations so defined, the typical small area variation phenomena are observed. Comparatively little variation in rates is observed for conditions where illness rates determine utilisation, while much greater variation is seen for conditions where practice style plays a role in determining utilisation. For example, hospitalisation rates for hip fracture and major bowel surgery for cancer show little variation among small areas. For these conditions, physicians agree on the diagnosis and the need for hospitalisation or surgery and differences in rates reflect differences in the rate for broken bones or incidence of cancer. By contrast, hospitalisation rates for ankle and forearm fractures vary substantially, as physicians in some areas elect to treat a greater proportion of these cases in the hospital than their colleagues in other areas who tend to prefer ambulatory treatment. Operations, such as hysterectomies, tonsillectomies, prostatectomies, and bypass surgery show similar high variation patterns of utilisation, largely because of differences among physicians in their opinion about appropriate practice. The fact that most causes of hospitalisation and most major surgery present the high variation profile indicative of professional uncertainty is what all the fuss is about.

The case for the importance of practice style does not rest on the pattern of variation alone. Several arguments lead the same way:

- No significant relationships have been found between measures of population need for health care and rates for high variation services. Survey studies of residents of communities with twofold differences in per capita costs for hospitals and up to sixfold differences in rates for individual operations show that illness rates, insurance coverage and a whole gamut of factors that predict patient demand for physician services are not related to differences in service use among small areas. Population age structure does not correlate with utilisation (in the USA). It is not that need does not vary — it obviously does — but rather that per capita utilisation rates and indicators of need are not correlated very much (e.g. 5 per cent or less of total variation) (Wennberg, 1985 and 1987a).

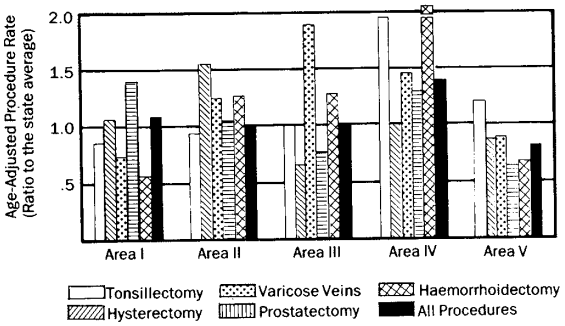
- Physician migration studies show the sometimes extraordinary impact that one or two physicians can have on utilisation. The classic example is Glover's account of the effects of Dr. Garrow's appointment as school medical officer on tonsillectomy rates for children in his district (Glover, 1938).
- Studies of physician response to feedback of information on small area variations show that practice style (rates of service) can change when physicians learn about their differences (Lembcke 1959; Dyck *et al.*, 1977; Wennberg and Gittlesohn, 1973; American Medical Association, 1986).
- The surgical signature phenomenon stands as a critique of the theory that health services are now rationed on the basis of competing needs and points to the importance of physician preference in the allocation of health resources (Figure 6.1).
- The professional literature documents wide differences in clinical decision rules (which in the case of Bloor's work shows a correlation with small area rates) (Bloor and Venters, 1978).
- A critique of the professional literature reveals large gaps in the knowledge base concerning the efficiency of hospitalisation, common diagnostic and surgical procedures as well as for drugs when used in non-approved ways. The exception is the scientific basis for the use of new drugs which has been systematically promoted by public policy over the last twenty years (Cochrane, 1972; Bunker *et al.*, 1977; Wennberg *et al.*, 1980). When new assessment techniques are applied, much more of the conventional wisdom may be shown to be in error (Wennberg *et al.*, 1987c; Fowler *et al.*, in press; Wennberg *et al.*, in press; Barry *et al*, in press; Miao, 1977).
- Physicians themselves recognise that practice variations stem from unresolved controversies concerning correct theory and they are willing to engage in studies to test their differences and improve the scientific basis of medicine (Chassin *et al.*, 1987a; Chassin *et al.*, 1987b; Kosecoff *et al.*, 1987; Wennberg, 1987b).

6.3 On the Implications of Professional Uncertainty for Equity and Efficiency

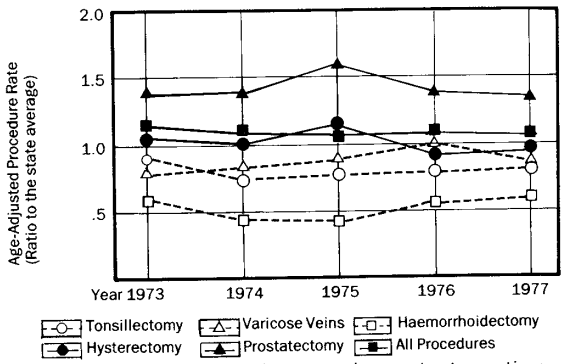
Borrowing from the American historian Frederick Jackson Turner, I have suggested that public policy for health in the West rests on a misplaced belief in 'manifest efficacy'. Despite the evidence to the contrary, the public assumes that clinical decision making is substantially based on science and that physicians can allocate available resources rationally among competing needs. Consistent with this assumption, policy makers have assumed that equity in health care can be achieved by making resources available on the basis of indicators of need such as the age or illness rates of population groups.

In the UK, equity is fostered by a public policy that seeks need-adjusted equalised per capita expenditures, presumably under the assumption that benefits will be allocated equally among individual patients according to a priority based on efficiency of treatment and

FIGURE 6.1 · THE SURGICAL SIGNATURE



The numbers on the vertical axis are the ratio of the state average rate to the area rate. The figure gives data for the five most populous hospital areas in Maine. It shows that the rates at which specific procedures are performed within an area vary markedly and to a large degree are independent of the total operation rate. Area II and Area III have the same total operation rate, but Area II exceeds in hysterectomies doing 56% more than the state average while Area III exceeds in varicose veins. In each of the five areas a different pocedure is performed most often; in four of the five areas, the least performed procedure is different. The numbers of surgeons and their specialty distribution do not vary to the same degree. It appears that variations in physician supply represent a less important contribution to small area variations in rates of high variation surgical procedures than differences in the opinions about the proper indications for surgery.



The surgical signature tends to remain constant over time. The trend lines give the rates of Area I for a five year period.

severity of illness. Yet small area variations, perhaps most dramatically the surgical signature phenomenon, show that even in the face of equalised per capita spending or equal utilisation rates for all surgery, the distribution of specific benefits to specific patients is, to a remarkable degree, quite arbitrary. The pattern of allocation of specific services raises a series of 'which rate is right' questions which frustrates the notion that equity is synonymous with equalised per capita expenditures. One example of the 'which rate is right' question runs as follows: is it more equitable to spend X per cent of gross national product (GNP) on tonsillectomies and remove thirty per cent of tonsils by age 15 or to spend .2X and remove six per cent? The policy analyst, in facing this and many similar

questions, can be no more certain about answers to equity issues than the physicians are certain about answers to efficacy questions.

The belief in manifest efficacy also affects the policies and politics of 'efficiency'. Efficiency involves efforts to assure that resources are economically allocated towards their intended ends — with little question of the ends which are assumed to be beneficial. Hence the emphasis is on the average lengths of stay and unit price, rather than underlying variations in the rates of service. The irony is that the most important factor in accounting for differences in cost per capita among equally needy populations is usually the per capita volume of services rather than variations in average length of stay or average price. For example, in demographically similar Boston and New Haven, the per capita rate of investment of health services in the local citizenry varies by a factor of almost two. The equivalent of 16 per cent of the GNP is invested in the health of the residents of Boston, while for New Havenites the figure is about 9 per cent. These differences occur largely because the rates of services such as admission to hospitals are much higher for Bostonians. This is a typical finding among small areas in the USA — and given the known variations in surgery rates among UK health districts, one which is very likely true also in the UK, at least for surgical admissions (Wennberg *et al.*, 1987b).

6.4 On Priorities and Programmes to Improve the Scientific Basis of Medicine

Some 25 high variation surgical procedures and 40 medical conditions can be identified as priority candidates for outcome assessment. Together, these conditions comprise over two-thirds of hospitalisation; the development of systematic information about the effects of alternative treatments for these conditions would greatly improve the scientific basis of surgery and general medicine and provide the physician, patient, and policy maker with information needed to improve decision making. In the United States, the government is now considering an assessment programme to establish teams of research and clinical scientists who will have long term responsibility for evaluating various theories on appropriate practice in these priority areas. The teams would be organised to assess alternative treatments for specific conditions — such as prostatism, stable angina and knee and hip conditions that sometimes warrant joint replacements. They would be expected to undertake assessments of current practice patterns, keep track of new developments, and design needed prospective clinical trials.

It is technically feasible to do these assessments. The past twenty years or so have seen remarkable progress in statistical and epidemiological techniques for measuring patient symptoms and evaluating the quality of life. They have also brought equally remarkable progress in computer and information science, new approaches for evaluating patient preferences and the adaptation of clinical decision theory to medical decision making. It is now possible to speak of a new set of disciplines which together constitute the evaluative clinical sciences.

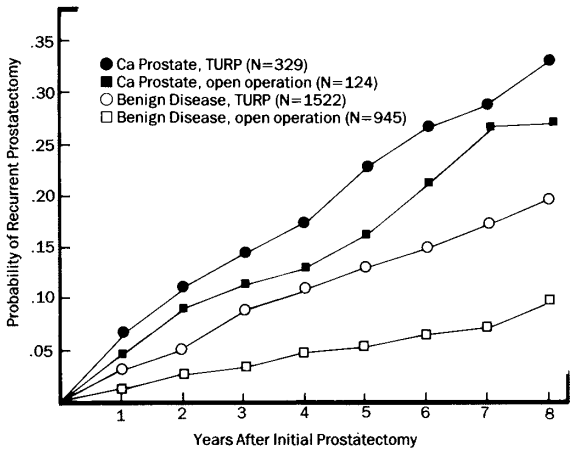
They offer the promise of a scientific programme that can greatly improve clinical decision making by decreasing uncertainty about the probabilities and the value to patients of the outcomes of care. They also offer new ways of communicating information to physicians and patients that can greatly increase understanding about the consequences of medical choices and thus help patients make the decisions they truly want.

We recently undertook an assessment of prostatectomy that takes advantage of some of the new developments in the evaluative sciences. (Wennberg *et al.*, 1987c; Fowler *et al.*, in press; Wennberg *et al.*, in press; Barry *et al.*, in press). The study of prostatectomy was possible because of the Maine Medical Assessment Programme, a programme of the Maine Medical Association organised to inform practising physicians about small area variations and to help them evaluate the role and significance of practice style in these variations. The Maine physicians decided to take a close look at the wide variations in the rate of prostatectomy among neighbouring Maine communities. The evaluation began with a detailed review of the surgical literature and with consultations with urologists practising in high and low rate areas to achieve a detailed understanding of the theoretical reasons for using prostatectomy and to obtain the best available estimates for the probabilities of the various outcomes associated with the watchful waiting and the surgical strategies for treating benign prostatic hypertrophy. We then sought to improve upon the estimates for outcomes we found in the literature by using Medicare claims data to measure the post-operative death rates and the incidence of complications following the operation.

Because of the long period of patient follow up and the size and representative nature of the patient population in the claims database (virtually all patients 65 and older) the estimates based on this data source were more accurate than those available in the published literature (see Figure 6.2). The estimates were integrated into a decision analysis to show that life expectancy is actually a little longer for those who do not undergo early operation.

The failure to demonstrate an extension in life expectancy means that, for patients without chronic obstruction, the operation is justified only under the quality of life theory for surgery. Therefore, patient preferences provide the rationale for the choice between surgery and watchful waiting in the individual situation. Information on the probabilities for symptom relief, improvements in the quality of life and data on the importance of symptoms for patients is thus of critical importance for rational decision making. Our literature review revealed a virtual absence of information about the probabilities for these 'soft' outcomes following prostatectomy. To obtain this critical information, we developed a patient questionnaire incorporating many of the methodological advances for measuring symptoms, functional status, and quality of life, and interviewed patients before their surgery and at three, six and

FIGURE 6.2 · CUMULATIVE PROBABILITY FOR RECURRENT PROSTATECTOMY BY TYPE OF PROSTATECTOMY



The figure gives the eight year cumulative probability for one or more secondary prostatectomies by diagnosis and type of primary prostatectomy. Patients who received an open operation had a lower probability for receiving a second operation. For those with no evidence of malignancy during the first operation, the relative risk of recurrence was 2.0. In the Cox regression model, the only significant co-variable was cancer of the prostate. Age, size of hospital, teaching status of hospital and all patient illness co-variables were not significantly associated with the probability of undergoing a secondary operation.

Source: Wennberg (1987 c).

twelve months after their prostatectomy. We were thus able to fill in many of the missing gaps in the estimates of the probabilities for symptoms[1] reduction and improvement in the quality of life, and obtain estimates for incontinence, impotence, and other complications not easily documented by claims data.

The patient interview results emphasise the complexity of clinical decision making when the objective is to improve the quality of life and show the importance of establishing which treatment patients truly want. While we found a rough correlation between the level of symptoms and the degree to which patients were bothered by their symptoms, we found no data that permitted us to predict accurately how a particular patient would feel about his symptoms. For example, some severely symptomatic patients were not bothered much by their symptoms. Presumably, some of these patients, if informed that they truly have an option and are told about the risks and benefits of surgery as well as those of watchful waiting, might have a different attitude towards surgery than those who are bothered a lot by their symptoms. Such findings emphasise the need for an active role for the patient in the decision to undergo this operation.

The prostatectomy assessment also offers a more specific diagnosis of the practice style related causes of small area variation for prostatectomy than was previously possible. For this operation, the origins of the practice style factor were an unjustified belief in the preventive theory (i.e. doing the procedure at an early time to avoid problems later). By providing more accurate estimates for the probabilities for outcomes as well as formal tests of the underlying surgical theory, such assessments should be broadly useful for improving the scientific basis of surgical practice. While not intended to replace randomised clinical trials when the latter are clearly indicated, the non-experimental study designs we used should be helpful in determining whether a randomised clinical trial is scientifically and ethically justified and can be useful in designing an optimum clinical trial. In situations where the outcome events are rare or so far in the future that randomised clinical trials are difficult or impossible or where sham surgery is needed to control for the placebo effect of surgery, they may offer the only reasonable means for testing surgical theory.

The work on prostatectomy was made possible because of the willingness of practising physicians and researchers to work together to test the underlying theories of clinical practice. I like Chris Ham's suggestion that the UK concentrate on establishing laboratories in certain health districts to begin to tackle the problems of variation in utilisation and to undertake research into outcomes (Chapter 7). I would like to add to his suggestion the notion that these studies should be done in an international context. Variations in theory and practice are greater in the international context and the tools for assessment that may be available in one country (e.g. claims data) may help resolve issues that emerge in another. Moreover, medicine is an international scientific enterprise and clinical scientists are part of a world-wide community. I very much like the proposal of the Copenhagen Collaborating Centre for the Study of Regional Variations in Health Care to organise an international network among clinicians and research scientists to study and ferret out the clinical uncertainties and undertake the necessary outcome studies. Ideally, this network should include population laboratories comprising communities that:

- are demographically and organisationally sufficiently similar to assure that local practice styles are reflected in per capita rates
- are medically sophisticated so that variations do not represent lags in the diffusion of technology or ignorance of appropriate care because opinions are not informed
- have health service researchers or teams of researchers interested in assessment of utilisation and outcome
- have clinicians who are interested in forming teams to investigate specific diseases to clarify the basic uncertainties
- have routine databases for monitoring variations and for longitudinal follow up of patient outcome.

1. *The initial loss of life due to the operation is not regained by avoiding the risks for surgery among those who may need the operation in the future to treat significant chronic obstruction.*

7 CONCLUSIONS AND IMPLICATIONS

Chris Ham

The research reviewed in this Report demonstrates the need for caution in interpreting evidence on variations in hospital utilisation rates. Although the supply of services and professional decision making appear to be important influences on utilisation rates, in the absence of data on the outcomes associated with different treatments it is difficult to determine correct rates of use. This applies both to hospitalisation rates and to lengths of stay.

To the policy maker concerned to promote efficiency and effectiveness in the use of health service resources, this conclusion provides a warning against hasty action based on incomplete evidence. At a time when the pressure to use NHS resources efficiently is greater than ever before, the natural response to data on variations is to press those health authorities who are apparently using resources inefficiently to improve their performance. This has already happened in the application of performance indicators and ministers have indicated that further action is likely to be taken to reduce variations in performance between authorities. Yet until more evidence on the consequences of different treatment patterns is available, it cannot be assumed that health authorities whose lengths of stay and hospitalisation rates are high are those whose performance should be adjusted downwards.

Equally, it cannot be assumed that the performance of these authorities is always appropriate. Indeed, to return to the work of Logan *et al.* (1972) discussed in Chapter 1, research in the Liverpool region in the late 1960s pointed to a clear association between the supply of services, hospitalisation rates and length of stay. Furthermore, through detailed analysis of services in the region the research demonstrated the scope for using resources more efficiently and substituting care in the community for hospitalisation. Twenty years on, the potential for realising efficiency gains of this kind is probably more limited as the most obvious examples of inefficiency have been eliminated in response to resource constraints. Nevertheless, the principle of analysing the use of resources on a local basis remains relevant.

This was certainly the view of the DHSS in its review of nationally-available statistics on acute services published in 1981 (DHSS, 1981). As the DHSS noted, any serious attempt to evaluate variations in resource use should focus on the district level. In part, this is a task for managers and clinicians making use of performance indicators, and in part it calls for original research. Such research would build on Holland's preliminary analysis of district variations in utilisation rates (Holland 1986) and would develop further the analysis by Wennberg and colleagues in the United States of differences in New Haven and Boston in health service expenditure, utilisation rates and provision (Wennberg *et al.*, 1987b). Initially, one or two pilot studies could be undertaken, followed by more comprehensive analysis if this seemed justified. The aim would be to establish the relationship, if any, between variations in provision and utilisation on the one hand, and health outcomes on the other.

Research is also needed into those high variation procedures where there is evidence of professional uncertainty about appropriate use. In his contribution to this Report, Wennberg has made the case for a programme of technology assessment designed to evaluate alternative treatments for specific conditions and the outcomes associated with these treatments. As Wennberg notes, the tools of evaluation are now available. What is required is the political will and the resources to support such a programme.

The example of prostatectomy cited by Wennberg illustrates one approach to evaluation and the Rand Programme of Health Services Research illustrates another. The Rand Programme seeks to establish the missing clinical links between data on variations and data on appropriateness (Brook *et al.*, 1984) by bringing together groups of experts to list and rank indications for treatment for specific procedures (see for example Solomon *et al.*, 1986). An expert consensus is then established, supported by a literature review, and in a number of cases this has been applied retrospectively to establish levels of inappropriate use. The results suggest that as many as one- third of certain common medical or surgical procedures are performed inappropriately in the United States. To date, the Rand approach has been used mainly in the United States, although it has recently been extended to the United Kingdom. The findings show that the indicators of appropriateness used by UK doctors are more rigorous than their American counterparts and this may account for the lower levels of utilisation of certain procedures in the UK (Brook *et al.*, 1988).

A further approach is to use consensus conferences to develop guidelines for the provision and use of medical and surgical procedures. Consensus conferences were initiated in the United States in 1977 and aim to evaluate specific technologies in a public forum making use of a panel hearing evidence from experts. The first United Kingdom consensus conference was held in 1984 on the subject of coronary artery by-pass grafting, and subsequent topics have included treatment of primary breast cancer, the role of asylum in the care and treatment of people with mental illness and genetic screening.

Consensus techniques are intended principally to influence professional attitudes and practices but they may also provide more information to the public, thereby stimulating informed choice on the part of service users (see Wennberg and Gittlesohn, 1982). The influence that users can have on clinical practice is well demonstrated by experience in Switzerland where the rate of hysterectomies fell following publication of information on variations in the use of hysterectomies in the press (Domenighetti, 1986). In the United Kingdom, the interest shown by women in the results of the consensus conference on treatment for breast cancer indicates that within the NHS there may well be scope for developing pressure by users for the more appropriate use of medical care.

Alongside efforts to inform and educate the public, a strong case can be made for greater involvement by clinicians in reviewing evidence on variations and acting on the results. Wennberg's example of the participation by the Maine Medical Association in

analysing variations demonstrates what can be done, as, in a different context, does experience with CEPOD in the United Kingdom. As Jennett emphasises in his contribution to this Report, clinicians are more likely to engage in audit if the information that is collected is seen to be clinically relevant and includes data on outcome as well as activities. A dual strategy is required to achieve this, involving leadership from the top of the profession by the medical colleges and specialist associations, and commitment from those directly involved in service provision.

There are a number of signs that the leaders of the profession are aware of the importance of these issues and are prepared to act. As an example, the President of the Royal College of Physicians has expressed interest in the audit of variations and the development of guidelines for clinical practice (Timmins, 1988; Hoffenberg, 1987). Similarly, the Royal College of Surgeons and the Association of Anaesthetists are to continue their collaboration in the development of CEPOD. At the local level, as Jennett points out, there have been various experiments with audit, but the overall impression created is that interest in audit among clinicians is uneven. Further progress may therefore depend on the bodies responsible for education and training making their approval for accreditation dependent on the existence of satisfactory arrangements for audit.

If audit is to be effective, it needs to take place within the context of clear guidelines for clinical practice. These guidelines can be derived from various sources including consensus conferences, expert consensus techniques, and standards formulated by the medical colleges and specialist associations. Guidelines can be used both to inform clinical practice and to evaluate practice. Used in this way, guidelines provide a series of benchmarks against which retrospective reviews can be undertaken. As Jennett points out, guidelines are already in use in a number of areas, but they need to be developed and deployed more widely if high standards are to be maintained.

Finally, a more speculative possibility is to develop standards for use by policy makers and managers. These standards would not involve indications of appropriateness for individual procedures as these are more properly the province of the medical profession. Rather, they would focus on norms of utilisation for particular populations. As examples, norms might be promulgated for the number of operations to be performed for a given population, the use of outpatient services, and GP referral rates. Given the diversity of community characteristics, these norms might be expressed as a range rather than a specific figure (see Paul-Shaheen *et al.*, 1987). The aim of these norms would be to enable managers and policy makers to audit the performance of their services, thereby complementing medical audit. Given uncertainty about appropriate levels of utilisation, the establishment of norms will not be easy, but the process itself may stimulate research to establish the missing links between data on variations and data on appropriateness.

In summary, the main conclusions and implications of this Report are:

- Evidence on variations in hospital utilisation rates needs to be interpreted with caution. In the absence of data on the outcomes associated with different treatments, it is difficult to determine correct rates of use. This applies both to hospitalisation rates and to lengths of stay.
- Future work should focus in particular on the analysis of variations in the use of resources at district level. This is a task for managers and clinicians making use of performance indicators, and for researchers. Initially, one or two pilot studies of the relationship between provision, utilisation and outcomes should be undertaken followed by more comprehensive analysis if this seems justified.
- Further research is needed to evaluate alternative treatments for specific conditions and the outcomes associated with these treatments. Priority should be given to the assessment of those high variation procedures where there is evidence of professional uncertainty about appropriate use.
- Clinicians should be more closely involved through medical audit in reviewing evidence on variations and acting on the results. This calls for leadership from the top of the profession and commitment from those directly involved in service delivery.
- Audit needs to take place within the context of clear guidelines for clinical practice. These guidelines can be derived from various sources including consensus conferences, expert consensus techniques and standards formulated by the medical colleges and specialist associations. Guidelines should be used to inform clinical practice and to evaluate practice.
- Norms of service utilisation should be developed for use by policy makers and managers. These norms might be expressed as a range rather than a single figure and they would be used to audit the performance of services.

REFERENCES

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

American Medical Association (1986), *Confronting Regional Variations: The Maine Approach* Chicago, Department of Health Care Review, Division of Health Policy and Programme Evaluation

Anon (1980) 'Second best prostatectomy' *BMJ*, 280, 590

M.J. Barry *et al.* 'Watchful waiting vs immediate transurethral resection for symptomatic prostatism: the importance of patients' preferences'. JAMA, In press

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

R. Brook *et al.* (1984), 'Geographic Variations in the Use of Services: Do They Have Any Clinical Significance'. *Health Affairs*, 3:2, 63-73

R. Brook *et al.* (1988) 'Diagnosis and treatment of coronary disease: comparison of doctors' attitudes in the USA and the UK'. *The Lancet*, i, 750-53.

N. Buck *et al.* (1987), *The Report of a Confidential Enquiry into Perioperative Deaths*. The Nuffield Provincial Hospitals Trust and the King's Fund, London

J.P. Bunker (1970), 'Surgical Manpower. A Comparison of Operations and Surgeons in the United States and in England and Wales'. *NEJM*, Vol 282, 135-144

J.P. Bunker *et al.* (1977), *Costs, Risks and Benefits of Surgery*. Oxford University Press, New York

R.B. Buttery and A.H. Snaith (1980), 'Surgical provision, waiting times and waiting lists'. *Health Trends*, 12, 57-61

M.S. Butts and D.R.R. Williams (1982), 'Accuracy of hospital activity analysis data' *BMJ*, 285, 506-7

C. Cantley and D. Hunter (1985), 'People Processing: Towards a Typology of Selected General Practitioner Referral and Admission Practices in Care of Elderly People'. *Ageing and Society*, 5, 267-288

M.R. Chassin *et al.* (1986), 'Variations in the Use of Medical and Surgical Services by the Medicare Population'. *NEJM*, 314, 285-290

M.R. Chassin, *et al.* (1987a), 'Does inappropriate use explain geographic variations in the use of health care services'. *JAMA*, 258, 2533-37

M.R. Chassin *et al.* (1987b), 'How coronary angiography is used: clinical determinants of appropriateness'. *JAMA*, 258, 2543-47

A.L Cochrane (1972), *Effectiveness and efficiency: random reflections on Health Services*, Nuffield Provincial Hospital Trust, London

J. Coid and P. Crome (1986), 'Bed blocking in Bromley'. *BMJ*, 293, 1253-56

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

A. Coulter and K. McPherson (1986), 'The Hysterectomy Debate'. *Journal of Social Affairs*, 2:4, 379-396

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

M. Dalziel and C. Garrett (1987), 'Intraregional variations in treatment of end stage renal failure' *BMJ*, 294, 1382-3

P. Day and R. Klein (1986), 'Controlling the gatekeepers: the accountability of general practitioners'. *Journal of the Royal College of General Practitioners*, 129-130

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

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

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

R. Dowie (1984), 'Deployment of resources in treatment of end-stage renal failure in England and Wales' *BMJ*, 288, 988-91

F.J. Dyck *et al.* (1977), 'Effect of surveillance on the number of hysterectomies in the province of Saskatchewan' *NEJM*, 296, 1326-1328

R. Earlam (1984), 'Oesophageal cancer treatment in North-East Thames Region, 1981: medical audit using HAA data'. *BMJ*, 288, 1892-94

S.R. Eastaugh (1980), 'Organistational determinants of surgical lengths of stay', *Inquiry*, 17, 85-96.

T.A.H. English *et al.* (1984), 'The UK Cardiac Surgical Register 1977-82' *BMJ*, 289, 1205-08

L.P. Fielding *et al.* (1980), 'Anastomotic integrity after operations for large-bowel cancer: a multi-centre study'. *BMJ*, 281, 411-14

F.G. R. Fowkes and B. I. McPake (1986), 'Regional Variations in Outpatient Activity in England and Wales'. *Community Medicine*, 8:4, 286-291

F.J. Fowler *et al.* 'Symptom status and quality of life following prostatectomy'. *JAMA*, In press

J.C. Gazet *et al.* (1985), 'Survey of treatment of primary breast cancer in Great Britain' *BMJ*, 290, 1793-95

O.J.A. Gilmore *et al.* (1980), 'Surgical audit: comparison of the workload and results of two hospitals in the same district'. *BMJ*, 281, 1050-52

A. M. Gittlesohn and J.E. Wennberg (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

J. A. Glover (1938), 'The Incidence of Tonsillectomy in School Children'. *Proceedings of the Royal Society of Medicine*, xxxi, 1219-36

J.A. Glover (1948), 'The paediatric approach to tonsillectomy' *Archives of Disease in Childhood*, 1-6

M. Griffith *et al.* (1979), 'Variation in hospital stay after inguinal herniorrhaphy' *BMJ*, 279, 787-89

R. Gruer *et al.* (1986), 'Audit of surgical audit' *The Lancet*, i, 23-26

R. Hoffenberg (1987), *Clinical Freedom* The Nuffield Provincial Hospitals Trust, London

W. W. Holland (1986), 'The RAWP Review: Pious Hopes'. *The Lancet*, ii, 1087-90

B. Jennett *et al.* (1980), 'Treatment for severe head injury' *J Nuerol Neurosurg Psychiat*, 411-14

B. Jennett (1985), 'Benefits and burdens of surgery' *Br J Surg*, 72, 939-41

B. Jonsson and B. Lindgren (1980), 'Five comon fallacies in estimating the economic gains of early discharge' *Social Science and Medicine*, 14C, 27-33

R. Kohn and K. L. White, (1976), *Health Care: An International Study*, Oxford University Press, London

J. Kosecoff *et al.* (1987), 'Obtaining clinical data on the appropriateness of medical care in community practice', *JAMA*, 258, 2538-2542

J.A.H. Lee *et al.* (1957), 'Fatality from three common surgical conditions in teaching and non-teaching hospitals' *The Lancet*, 785-90

P.A Lembcke (1959), 'A scientific method for medical auditing'. *Hospitals*, 33, 65-71

C.E. Lewis (1969), 'Variations in the Incidence of Surgery'. *NEJM*, 281, 880-884

R. Logan, *et al.* (1972), *Dynamics of Medical Care* Memoir No. 14, London School of Hygiene and Tropical Medicine

London Health Planning Consortium (1979), *Acute Hospital Services in London* HMSO

K. McPherson *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*, 15A, 273-288

K. McPherson *et al.* (1982), 'Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England and Norway'. *NEJM*, 307, 1310-14

K. McPherson *et al.* (1985a), 'Do cholecystectomy rates correlate with geographic variations in the prevalance of gallstones?'. *Journal of Epidemiology and Community Health*, 39:2, 179-182

K. McPherson *et al.* (1985b), 'Increasing use of private practice by patients in Oxford requiring common elective surgical operations' *BMJ*, 291, 797-99

L.L. Miao (1977), 'Gastric freezing: an example of the evaluation of medical therapy by randomised clinical trials'. In Bunker J P, Barnes B A, Mosteller, F (eds) *Costs, Risks and Benefits of Surgery* Oxford University Press, New York, Chapter 12, 198-211

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

M. Morgan *et al.* (1987), 'Can hospital use be a measure of need for health care?' *Journal of Epidemiology and Community Health*, 41, 269-74

F. D. Moore (1985), 'Small Area Variations Studies: Illuminating or Misleading?' *Health Affairs*, 4:1, 96-101

P. Paul-Shaheen, *et al.* (1987), 'Small area analysis: a review and analysis of the North American Literature' *Journal of Health Politics, Policy and Law*, 12:4, 741-809

R.J.C. Pearson *et al.* (1968), 'Hospital Caseloads in Liverpool, New England and Uppsala'. *The Lancet*, 559-566

N. Roos (1984), 'Hysterectomy: Variations in Rates Across Small Areas and Across Physicians' Practices'. *American Journal of Public Health*, 74:4, 327-335

N. Roos and L. Roos (1982), 'Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population'. *Medical Care*, xx, 945-58

N. Roos, L. Roos and P. Henteleff (1977), 'Elective Surgical Rates — Do High Rates Mean Lower Standards?'. *NEJM*, 360-365

H.F. Sanderson (1980), 'Regional variation in cataract extraction rates and their relationship with resource supply and need'. *Journal of the Royal Society of Medicine*, 73, 492-496

F. Sloan and J. Valvona (1986), 'Why has hospital length of stay declined: an evaulation of competing theories'. *Social Science and Medicine*, 22, 63-73

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

R.M.R Taylor *et al* (1985), 'Renal transplantation in the UK and Ireland — the centre effect' *The Lancet*, i, 789-802

N. Timmins (1988), 'Doctors to monitor quality of treatment' *The Independent* 15 March

E. Vayda (1973), 'Comparison of Surgical Rates in Canada and in England and Wales'. *NEJM*, 289, 1224-1229

J.E. Wennberg *et al.* (1980), 'The need for assessing the outcome of common medical practices. *Annual Review of Public Health*, 1, 257-295

J.E. Wennberg and A. Gittlesohn (1982), 'Variations in Medical Care Among Small Areas'. *Scientific American*, 246, 100-112

J.E. Wennberg and A. Gittlesohn (1982), 'Variation in Medical Health Care Delivery'. *Science*, 182, 1102-1108

J.E. Wennberg (1984), 'Dealing With Medical Practice Variations: A Proposal for Action'. *Health Affairs*, 3:2, 6-32

J.E. Wennberg (1985), 'On patient need, equity, supplier-induced demand, and the need to assess the outcome of common medical practices'. *Medical Care*, 23, 512-520

J.E. Wennberg (1987a), 'Population illness rates do not explain population hospitalisation rates' *Medical Care*, 25, 354-359

J.E. Wennberg *et al.* (1987b), 'Are Hospital Services Rationed in New Haven or Over-Utilised in Boston?' *The Lancet*, 1185-9

J.E. Wennberg *et al.* (1987c), 'Use of claims data systems to evaluate health care outcomes'. *JAMA*, 257, 933-936

J.E. Wennberg (1987d), 'The paradox of appropriate care' *JAMA*, 258, 2568-9

J. E. Wennberg *et al.* 'An assessment of prostatectomy for benign urinary tract obstruction' *JAMA*. In press

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

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

7310/DESIGN & PRINT/REDESIGN/01-533 2631

KING'S FUND INSTITUTE

The Institute is an independent centre for health policy analysis which was established by the King's Fund in 1986. Its principal objective is to provide a balanced and incisive analysis of important and persistent health policy issues and to promote informed public debate about them.

Assessing the performance of health care systems is one of the Institute's central concerns. Many of its projects focus on trying to determine whether health care systems achieve their objectives. The Institute is also concerned with health policy questions which go wider than health services proper. These centre on the scope of public health policy and on social and economic determinants of health.

The Institute's approach is based on the belief that there is a gap between those who undertake research and those responsible for making policy. We aim to bridge this by establishing good relations with the scientific community, and by gearing our work towards making the most effective use of existing data. One of our key objectives is to undertake informed analyses and channel them to politicians, civil servants, health managers and professionals, authority members and community representatives.

The Institute adopts a multidisciplinary approach and seeks to make timely and relevant contributions to policy debates. A high priority is placed on carefully researched and argued reports. These range from short policy briefings to more substantial and reflective policy analyses.

The Institute is independent of all sectional interests. Although non-partisan it is not neutral and it is prepared to launch and support controversial proposals.

Other publications from the King's Fund Institute

Healthy Public Policy: A Role for the HEA Briefing Paper No 1, Virginia Beardshaw. £2.00.

Public Expenditure and the NHS: Trends and Prospects Briefing Paper No 2, Ray Robinson and Ken Judge. £3.95.

Promoting Innovation in Community Care: From Small Scale Developments to Mainstream Provision Briefing Paper No 3, David Hunter. £3.00.

Community Physicians and Community Medicine Report No 1, Sarah Harvey and Ken Judge. £4.95.

Health Finance: Assessing the Options Briefing Paper No 4, Ray Robinson et al. £4.95.

Griffiths and Community Care: Meeting the Challenge Briefing Paper No 5, David Hunter and Ken Judge. £3.95.

Further details about the Institute and its publications can be obtained from:

Su Bellingham
King's Fund Institute
126 Albert Street
London NW1 7NF
Telephone: 01-485 9589