R. Maxwell

The Assessment and Use of Health Care Technology

Chris Ham and Bryan Jennett

August 1987

King's Fund Institute

CONTENTS

INTRODUCTION

This report presents the results of our review of experience of health care technology assessment in the UK and overseas. The report has been prepared principally to inform the programme of work on technology assessment being developed at the King's Fund Institute. It is therefore a benchmark exercise, seeking to establish the state of the art of technology assessment as a basis for determining the contribution the Institute is able to make in this field. Almost inevitably in such a large field, the review is selective, but it is hoped that the main milestones and landmarks have been identified.

At this stage, the report has been written for internal discussion, in particular for the meeting of the Institute's Advisory Committee in October 1987. With some amendment, it is intended that the report will be published by the Institute as a contribution to debate and discussion about health care technology assessment in the UK.

We are grateful to all of those people who have commented on earlier drafts of the report and who have contributed to the development of our thinking. Particular thanks are due to Barbara Stocking.

Chris Ham

Bryan Jennett

CHAPTER 1

WHAT IS TECHNOLOGY ASSESSMENT ?

In a widely-quoted statement the Office of Technology Assessment of the United States Congress defines technology as "The drugs, medical devices and surgical procedures used in medical care, and the organisational and supportive systems within which such care is provided." (Office of Technology Assessment, 1982, pp200-201). The assessment of medical technology involves the examination of the properties of a technology from a number of perspectives including its safety, effectiveness, efficiency and acceptability (Institute of Medicine, 1985, p.2). Put another way, technology assessment entails the evaluation of medical interventions in order to establish whether they are safe, help to extend life or improve the quality of life, are cost-effective and are viewed positively by those to whom they are administered.

On the basis of this broad definition, various areas of technology assessment can be identified. One major area concerns the use of drugs where there are established procedures for testing new drugs and monitoring usage after their introduction. Second, examples of technology assessment can be found in the area of preventive medicine, including the evaluation of immunisation programmes and of breast cancer screening. Third, of increasing importance is the application of technology assessment to diagnostic tools, most notably in the case of CT scanning and magnetic resonance imaging. Fourth, technology

assessment has been used to evaluate medical treatments, including heart surgery, intensive care units and kidney transplants.

In each of these areas, a number of assessment criteria have been employed. The principal purpose in many cases has been to establish the clinical effectiveness of an intervention or procedure. That is, assessment has been grounded in a concern to identify whether the intervention produces or assists in producing a positive health outcome for patients. In other cases, the motivation has been based on economic considerations and has sought to establish the cost-effectiveness of an intervention. In still other cases, the major focus has been on the social acceptability of the intervention and the response of patients. Some forms of technology assessment combine elements of each approach.

Technology assessment has embraced a wide variety of methods. These range from formal randomised controlled trials through economic evaluations to consensus conferences, data synthesis and the development of clinical guidelines and protocols. A recent analysis by the Institute of Medicine in the United States identified a plethora of approaches and a large number of organisations that were involved in sponsoring and undertaking technology assessment programmes (Institute of Medicine, 1985). Despite the breadth of activity taking place, the analysis echoed a review conducted by the Office of Technology Assessment (1982) in concluding that a more systematic and coordinated approach was required. Similar conclusions have been reached by observers of the British and Swedish scenes (Jennett, 1986; Swedish Ministry of Health and Social Affairs, 1986) and there is an emerging international consensus that there should be a greater investment in technology assessment

activities. Why is this, and where has the interest in technology assessment stemmed from?

Interest in technology assessment has arisen from sources both within the medical profession and external to it. Doctors have long been involved in informally monitoring and evaluating the impact of their interventions and in reporting the results to a wider audience. In this sense, technology assessment is as old as medicine itself, even though the term technology assessment has only recently entered the vocabulary of health policy. The scientific tradition on which medicine is based places a high value on formulating hypotheses, testing these hypotheses through empirical experimentation, and analysing the results to establish the effect of medical activity. It has been through this kind of scientific evaluation that some of the most important advances in medical science have been made, not least in the area of drugs and vaccines.

Yet as the pace of technological change has quickened, many procedures have been introduced in the absence of any assessment of their effectiveness. While some of these procedures have rapidly demonstrated their benefit and have been widely adopted, others have been of questionable value. Examples of beneficial procedures are hip replacements, the implantation of heart pacemakers and cataract surgery. Examples of questionable procedures that had a few years of popularity but were then shown to be of no value are gastric freezing for duodenal ulcer, mammery artery ligation for coronary artery disease, and bypass surgery for the prevention of stroke (Jennett, 1986).

Much more common are those procedures which are appropriate for some patients but not others. For example, coronary artery bypass grafting is of benefit only for selected patients with heart disease and for others offers no advantages over medical management (Williams, 1985; Hampton, 1983). Similar questions have been raised about the increasing use of caesarian sections, radical surgery for breast cancer, intensive care units and hysterectomies. The existence of significant variations in the provision and use of services between countries, areas and clinicians suggests that there may be other procedures which are not used in an optimum manner. Where such variations are not determined primarily by differences in morbidity, there is prima facie evidence to indicate that levels of provision and use should be reviewed.

In the light of this kind of evidence, writers such as Cochrane (1972), Dollery (1978) and Jennett (1986) have called for a more rigorous approach to assessing the costs and benefits of contemporary medical practices. This involves analysing not only the claims made for new and emerging technologies but also reviewing the evidence on established procedures. In part this means assessing expensive and often complex high technology medicine such as CT scanners, and in part it entails evaluating the use of mature technologies which although inexpensive in themselves may have a significant impact on resource allocation because of the volume of their use. An example of the latter is the use of pre-operative chest x-rays, an area in which there has been considerable success in reducing unnecessary usage (Fowkes, 1985).

While Cochrane, Dollery and Jennett have been at the forefront of the movement for technology assessment within the medical profession, they have found support from policy analysts, politicians, health economists and others. Two

policy analysis studies of particular relevance are Stocking and Morrison's work on the introduction of whole body CT scanners, and the Council for Science and Society's assessment of procedures for making decisions in the NHS on the use of expensive medical techniques (Stocking and Morrison, 1978; CSS, 1982). Both concluded that existing methods of assessing medical technology and managing its introduction are inadequate and that new organisations are needed to provide information for policy makers.

Among politicians, the most cogent commentary on technology assessment was made more than a decade ago by David Owen when he was the Minister of State for Health in the 1974-79 Labour government (Owen, 1976). Drawing on evidence of variations in practice among GPs and hospital consultants, Owen called on doctors to examine much more carefully the impact of their decisions on resource allocation. Given a situation of almost limitless demand for health care and finite resources, Owen argued that the medical profession had to face "the practical economic facts of life" (p.1008). In this context, it is worth noting that the Priorities document (DHSS, 1976), published when Owen was a Minister in the DHSS, placed a particular emphasis on the need to increase the efficiency of clinical services. The document included an illustrative list of reports on innovations in clinical practice in the hope that "this...will encourage further scrutiny by the professions of the resources used by different regimes" (p.28).

Owen's interest in technology assessment stemmed principally from a concern to achieve better value for money at a time of increasing resource constraints. This concern has been taken forward more recently through initiatives such as clinical and management budgeting, performance indicators and diagnostic

related groups. Health economists have played a part in the development of these initiatives and in the application of cost-benefit and cost-effectiveness analysis to health services. These tools, together with the use of QALYS for comparing the benefits of different treatments, offer a range of methodologies for the economic evaluation of medical care. A recent review of work in this field enumerated advances made over the last decade, but bemoaned the lack of impact of economic evaluation on policymakers (Drummond, 1987). There are echoes here of Jennett's discussion of technology use in which he notes that publishing the results of technology assessment is no guarantee that medical practices will change. The clinical freedom of the medical profession means that it is ultimately the repsonsibility of individual clinicians to decide what is best for patients. Influencing clinical decisions requires a range of strategies and incentives, and examining the nature of these strategies and incentives is an issue of considerable importance.

A group of external critics not yet mentioned are patients and the public as users of services. One area in which the influence of this group has been felt strongly is the maternity services where there has been a movement against the increasing use of technologies and of more interventionist approaches (electronic foetal monitoring, induction, and caesarians). Patients' views have also been prominent in discussions of treatment of terminal illness, breast cancer and end-stage renal failure. In the last of these cases, important ethical issues have arisen over the rationing of dialysis treatment, and organisations representing patients have argued for technology to be made more widely available. A key point follows, namely that technology assessment is as much concerned with increasing the use of

effective treatments as with decreasing or eliminating the use of ineffective treatments. What these examples also indicate is that the social acceptability of medical technology needs to be considered alongside clinical effectiveness and economic efficiency.

As the above discussion has indicated, there are many perspectives on medical technology. It can be suggested that technology assessment is an important activity because of the failure to evaluate new procedures before their widespread adoption, the need to allocate resources efficiently in an era of financial constraints, and the concern of patients and the public with the personal and social impact of medical technology. We have noted that technology assessment has been applied patchily in practice, and even where it has been undertaken results of assessments have not always been acted upon. One of the obstacles to the use of information derived from assessments is clinical freedom. The autonomy enjoyed by the medical profession helps to explain the over-enthusiastic adoption of some procedures, the uneven development of others, and the difficulties encountered in limiting or halting the use of yet others. Also significant is the ambiguity of much of the existing information about medical technologies and the professional uncertainty which surrounds therapeutic interventions. It is these factors which create a climate in which clinical freedom is able to flourish.

The rest of this paper explores these issues further and seeks to identify the contribution that the King's Fund Institute can make to technology assessment. Chapter 2 examines the evidence on variations in the use of health services and the relevance of the literature on variations to technology assessment. Chapter 3 describes the scope of technology assessment work in the United

Kingdom. Chapter 4 looks further afield to Europe and North America to assess whether there are any lessons to be learnt from overseas experience. Chapter 5 describes the different methods of assessment that exist, while Chapter 6 explores the influence of technology assessment on clinical practice. Finally, Chapter 7 discusses the options available for technology assessment work in the future, and it outlines an initial programme of work for the King's Fund Institute.

REFERENCES

Cochrane A (1972) Effectiveness and Efficiency, Nuffield Provincial Hospitals Trust, London

CSS (1982) Expensive Medical Techniques, London

DHSS (1976) Priorities for Health and Personal Social Services in England, HMSO, London

Dollery C (1986) The End of an Age of Optimism, Nuffield Provincial Hospitals Trust, London

Drummond M (1987) 'Economic Evaluation and the Rational Diffusion and Use of Health Technology' Health Policy, Vol 7, pp.309-324.

Fowkes F G R (1985) Strategies for changing the use of diagnostic radiology, King's Fund, London

Hampton J (1983) 'The end of clinical freedom', BMJ, Vol 287, No 6401, pp. 1237-8

IOM (1985) Assessing Medical Technologies, National Academy Press, Washington DC

Jennett B (1986) High Technology Medicine: Benefits and Burdens, Oxford University Press, Oxford. New edition

OTA (1982) Strategies for Medical Technology Assessment, US Government Printing Office, Washington DC

Owen D (1976) 'Clinical freedom and professional freedom', Lancet pp. 1006-1008

Stocking B and Morrison S (1978) The Image and the Reality, Nuffield Provincial Hospitals Trust, London

Swedish Ministry of Health and Social Affairs (1986) The Assessment of Medical Technology - A Proposal for National Coordination

Williams A (1985) 'Economics of coronary artery bypass grafting', BMJ, Vol 291, pp. 326-329

Bennett, B (1985) High Pressure Welding
University Press, Oxford

OTA (1985) Biomaterials for Medical Devices
Printing Office, Washington D

Owen D (1979) Ultrasonic Cleaning
pp. 1009-1060

Stockley B and Wenham S
Physical Properties

CHAPTER 2

VARIATIONS IN THE USE OF HEALTH SERVICES

We noted in Chapter 1 that evidence on variations in the provision and use of health services has been used to raise questions about the effectiveness and efficiency of medical practice. From the perspective of technology assessment, this evidence is important in highlighting areas of medical practice where there may be lack of agreement between doctors on the benefits of a particular procedure and its indications for use. A growing volume of literature has sought to account for the variations that exist and to test the significance of different factors. This chapter summarises some of the key issues which emerge from the literature and identifies the implication for policy and priorities for research. The analysis demonstrates that study of health care variations can make a useful contribution to a programme of work on technology assessment, not least in identifying technologies that appear to need detailed investigation.

Background

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 geographical variations in health spans half a century, it is only in the last ten years that a major effort has been made to describe and explain these variations.

This is illustrated in Table 1 which shows the year of publication of the references listed in the bibliography prepared by the Copenhagen Collaborating Centre for the Study of Regional Variations in Health Care. Analysis of the publications identified by the 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

YEAR OF PUBLICATION OF CITATIONS IN THE CCC BIBLIOGRAPHY

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

As already noted, one of the seminal papers on health care variations examined the rate at which tonsils were removed in different parts of England. Study of variations in the *use* of services as measured by the number of procedures carried out for a given population is the main theme of much of the later research that has been conducted, and it is principally this research which is examined here. However, analysis has been extended in some cases to include variations in the *provision or supply* of services (beds, doctors etc) and variations in the *efficiency* with which services are provided (length of stay, costs per case). Furthermore, researchers have used a range of population units in studying variations. These include small areas within one country, large areas within one country, small and large areas between countries, and countries as a whole. In examining variations at different levels of aggregation, a number of researchers have turned their attention to the practice style of individual doctors as a possible source of the variations that exist. This has given rise to an important minor theme in the literature, namely variations between clinicians in the way they use 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 use of services in certain areas indicate unnecessary or inappropriate use?
- Can the higher costs associated with high rates of use be justified in terms of improved health outcomes?
- Are differences in use related to the need for care of the population concerned?

As will become apparent, it is possible to offer some answers to these questions, but much work remains to be done. Although considerable progress has been made in *describing* variations, attempts at explanation have so far proved inconclusive or at least not susceptible to clear-cut conclusions. Research on causality has focussed on demand-side factors such as population characteristics and morbidity, supply-side factors such as the provision of doctors and beds, and on the practice style of clinicians. A number of studies have speculated on the importance of these factors; others have used multivariate analysis and related forms of statistical investigation to test their significance. Despite the growth of interest in variations studies and the increasing sophistication of the methodologies used, there is a continuing debate about the relative importance of different variables. Much work points to the importance of the supply of services and practice style, but there remains a good deal of uncertainty. In the absence of convincing explanations, one of the main contributions which analysis can make is to identify key issues for debate and action by policy makers, the medical profession and researchers. We discuss this more fully in the final part of the chapter.

International Variations

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

Liverpool, Uppsala and New England. These differences are illustrated in Table 2.

TABLE 2

—DISCHARGE-RATES PER 10,000 POPULATION FOR SELECTED COMMON OPERATIONS, BY SEX

Operation	H.I.P.E. codes	Males			Females		
		Liverpool	Uppsala	New England	Liverpool	Uppsala	New England
Tonsillectomy and/or adenoidectomy ..	261–3	29	15	68	23	19	72
Inguinal herniorrhaphy	402	26	35	49	3	7	5
Appendicectomy ..	441	23	31	20	26	27	14
Cholecystectomy ..	521	3	21	10	7	49	27
For peptic ulcer* ..	422–3 427, 431 433–4	10	9	7	3	3	3
D. & C.	732	..	..	..	25	76	65
Total abdominal hysterectomy ..	722	..	..	..	14	6	28
For prolapse* ..	724, 743–4	..	..	..	14	7	11
Mastectomy	381–3	..	..	..	10	8	25
Prostatectomy ..	672–7	7	14	18	..	..	..
Extraction of lens ..	173	3	5	9	5	5	10

* Corrected to exclude operations for other conditions.

Source: Pearson et al. (1968)

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

of the differences may be related to variations in incidence of a condition, but many are more likely to be caused by differences in the systems of medical care" (Pearson et al, 1968, p.563).

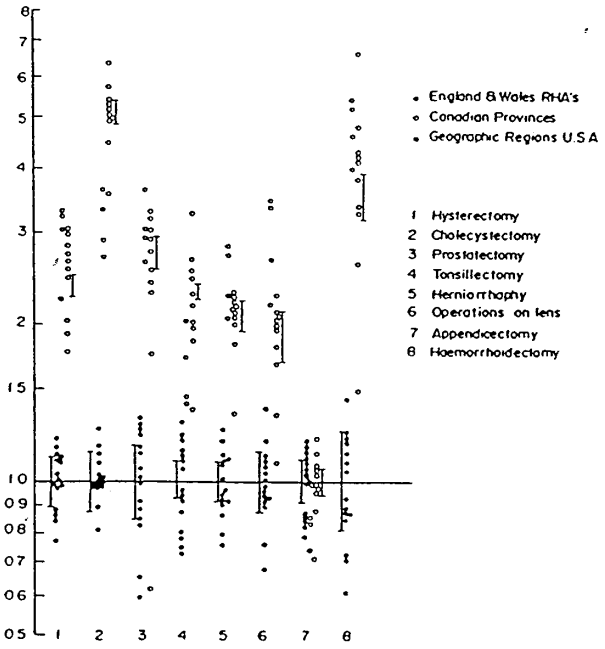
This conclusion was supported by Bunker's comparison of surgical services in the United States and England and Wales (Bunker, 1970). Bunker found that there were twice as many surgeons in proportion to the population in the United States as in England and Wales, and they performed twice as many operations.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

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

that surgical rates in Canada were 1.8 times greater for men and 1.6 times greater for women than in England and Wales. The age standardised and sex specific rates for particular operations were two or more times higher in Canada than in England and Wales. In seeking to explain these variations, Vayda argued that the key factors were the more conservative treatment styles in England and Wales, the greater availability of surgeons and beds in Canada, and the impact of financial incentives to operate in Canada. Differences in disease prevalence, as measured by mortality rates, were not found to be important. Vayda concluded that it was difficult to establish whether surgical rates were too high in Canada or too low in England and Wales. Accordingly, he called for further work through controlled trials to establish the benefits of surgical and non-surgical treatment for common diseases, and he suggested that the medical profession should initiate or expand audit programmes to establish appropriate indications for surgery (Vayda, 1973). In other words, Vayda used data on variations to argue for a greater investment in technology assessment.

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

TABLE 3



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

Source: McPherson et al. (1981)

McPherson et al noted that a major problem in interpreting these data was "the lack of any comparative morbidity rates, whose variation could, in principal, explain all the observed variation in the rates for the operations we have described here" (p.280). A further problem was the absence of an agreed set of indications for surgery which would enable a correct rate to be established. Notwithstanding these difficulties, on the basis of a series of multiple regression analyses, McPherson et al concluded that variations in

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

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

TABLE 4

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

Coefficient of Variation (%)	Hernia Repair	Appendectomy	Cholecystectomy	Prostatectomy	Hysterectomy	Hemorrhoidectomy	Tonsillectomy	All Seven Procedures
New England	0.11	26	18	30	22	30	36	14
Norway	0.20	16	18	33	31	47	48	11
West Midlands	0.20	16	16	24	20	35	31	12
Range (high/low)								
New England	1.7	2.3	1.9	2.2	2.2	4.8	4.2	1.69
Norway	1.3	1.6	1.5	2.2	3.0	2.9	4.7	1.34
West Midlands	2.0	2.0	1.5	2.1	2.1	4.6	3.3	1.55
Systematic Component * ($\times 100\ \sigma^2$)								
New England	0.6	1.7	1.7	5.0	4.8	12.7	12.2	2.08
Norway	0.2	2.4	1.9	9.3	10.4	14.7	27.5	1.28
West Midlands	4.4	2.9	2.1	6.2	3.7	12.2	18.5	1.33

*See Appendix.

Source: McPherson et al. (1982)

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

had become possible as a result of advances in medical technology. 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. Overall, Aaron and Schwartz concluded that the rationing of services was more difficult in Britain because of the lower overall investment in health services.

Small Area Variations

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

This finding is supported by other studies. For example, Wennberg and Gittlesohn (1973), in an analysis of health care variations in thirteen areas

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

In a specific study of tonsillectomy and adenoidectomy in Vermont, Gittlesohn and Wennberg reported that age-adjusted rates varied between areas from 4 to 41 per thousand children per year. It was estimated that for the entire state, 22 per cent of children would have their tonsils and adenoids removed by their twentieth birthday, but the risk of removal varied from 9 per cent to 60 per cent between areas. Directly adjacent communities had rates of 11 per cent, 19 per cent, 20 per cent, 27 per cent and 60 per cent, and the authors concluded, "it is unlikely that the differential tonsillectomy rates can be related to variations in the incidence of tonsillitis, recurrent sore throat, or otitis media. Rather, the major source of variation appears to be in differing attitudes by physicians as to indications for the procedure" (Gittlesohn and Wennberg, 1977, p.95).

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

In this context, a study by Bloor and Venters of small area variations in tonsillectomy and adenoidectomy in Scotland is relevant (Bloor and Venters, 1978). The surgical rate within one region varied between areas from 6.2 to 15.8 operations per thousand children per year. These differences in part reflected variations between general practitioners in their rate of referral to specialists, but independently of this the practice style of specialists had a key influence on the number of operations performed. The authors noted

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

The importance of practice style and professional uncertainty emerge as key themes in Wennberg's later work. A review of small area variations in six states in New England identified the importance of supply factors in accounting for variations, but concluded that the availability of beds and surgeons could not furnish a complete explanation of all the variations that existed (Wennberg and Gittlesohn, 1982). Rather, the judgements and preferences of doctors were a key factor. Furthermore, Wennberg and Gittlesohn coined the phrase "surgical signature" (p.106) to describe the phenomenon of high surgical rates for particular operations in individual areas. Failure to undertake technology assessments 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.

The last study of small area variations to be considered here is an analysis of 13 regions of the United States (Chassin et al, 1986). Although covering more populous areas than the other analyses discussed here, this study found large and significant differences in the use of services provided by all medical and surgical specialties. The authors noted that their findings are partly consistent with the view that the degree of variation for a particular

procedure is linked to the degree of consensus concerning the indications for its use. However, some of the results did not support this view, nor was there evidence to suggest that doctors in high use areas performed procedures less appropriately than those in low use areas. The authors emphasised the uncertainty surrounding the reasons for variations:

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

This conclusion is a useful reminder that interpretation of the evidence on variations remains highly contestable, a point reinforced by other commentators (eg Moore, 1984). As Smits (1986) has noted, two particular issues which merit further attention are, first, the extent to which variations reflect the uneven dissemination of innovative procedures, and second, the relationship between epidemiology and use rates.

Variations Between Doctors

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

TABLE 5 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

Health Trends, 1980, Volume 12.

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

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 of the variations that exist remain unexplained. On the basis of research conducted in Manchester, these authors concluded that neither patient characteristics nor the characteristics of GPs and their practices could adequately account for the variations observed, and they called for further research "looking both at those patients who are referred and those who are not, how those decisions are arrived at, what are the outcomes for patients and the costs both for services and for the community" (p37).

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

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

Variations' Studies in the United Kingdom

Many of the studies cited in this paper concern variations in use rates in the United Kingdom. In addition to these studies, other work in the United Kingdom includes Sanderson's study of regional variations in cataract extraction rates (Sanderson, 1980), Dalziel and Garrett's research on intraregional variations in the treatment of end stage renal failure (Dalziel and Garrett, 1987). Fowkes and McPake's analysis of regional variations in

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

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

The issue of equity in treatment also arose in the study by Dalziel and Garrett (1987) of intraregional variations in the treatment of end stage renal failure. This study reported that the further a patient lived from a dialysis centre, the less likelihood there was of a patient receiving treatment. The authors speculated that one of the reasons for this was that the further GPs and hospital consultants were from the dialysis centre, the less likely they were to have current information about methods of treatment. Dalziel and Garrett called for further investigation of intraregional variations as a means of identifying implications for the NHS.

for these large variations" (p384). Coulter and McPherson emphasised instead the influence of supply factors and practice style, and they echoed McPherson et al's study of cholecystectomy in suggesting that international differences probably reflected differences in the mode of organising and financing health services.

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

To explore the issues raised by this work further, we now examine the implications of the variations literature for researchers, the medical profession and policy makers. In so doing, an attempt is made to identify priorities for future work on technology assessment.

Issues for the Future

A useful starting point in considering priorities for future work is Wennberg's plan for dealing with variations (Wennberg, 1984). This has three parts. First, Wennberg argues that there is a need to monitor small area variations in service inputs, use rates and outcomes. The resulting information should be fed back to clinicians and decision makers in order to influence and change clinical practice, as is the case in the Maine Medical Assessment Programme. Second, Wennberg contends that greater efforts should be put into assessing the effectiveness of services and procedures and measuring outcomes. Indeed, at the Copenhagen conference on variations held in November 1986, Wennberg argued that the outcome problem is the major challenge facing variations researchers, in particular clinicians. This can be tackled through literature reviews, consensus techniques and original research, in some cases making use of existing administrative data bases (eg. MEDICARE). Third, there is a need to reduce unnecessary or inappropriate use of hospital services, principally through more concerted efforts on the part of the medical profession, for example through audit.

Some work of this kind is being done by Robert Brook and his colleagues as part of the RAND programme of health services research. One element in this programme is designed to describe more systematically the pattern of health care variations that exist in the USA (see Chassin et al, 1986). Another element focusses on establishing the missing clinical links between data on variations and data on appropriateness (Brook et al, 1984). Thus, in a series of studies, groups of experts have been brought together to list and rank indications for treatment for specific procedures (see for example, Solomon et al, 1986). An expert consensus has then been established, supported by a

literature review, and in some cases this has been applied retrospectively to establish levels of inappropriate use. To date the RAND consensus technique has been used mainly in the USA, although there have been two (as yet unpublished) studies conducted by RAND in the UK (see also Chapter 5).

These initiatives are consistent with the proposals from Schacht and Pemberton (1985) for the greater use of review committees to establish the circumstances under which treatment should and should not be provided. They are also congruent with efforts made in a number of countries to use consensus techniques of varying kinds to review controversial areas of medical practice. The results of these reviews are intended principally to influence professional opinion, but they may also be used to provide more information to the public in the hope of stimulating informed choice on the part of service users (see Wennberg and Gittlesohn, 1982). The influence that users can have is well demonstrated by experience in Switzerland where the rate of hysterectomies fell following publication of information of variations in the use of hysterectomies in the press (Domenighetti, 1986).

It would seem, therefore, that the literature on health care variations has a number of implications for those involved in this field and for the Institute's work on technology assessment. These are:

a) The need for systematic monitoring of variations in service inputs, use rates and outcomes;

b) The importance of feeding back information gathered to clinicians, policy makers and the public;

c) The need to investigate the outcomes associated with different treatments;

d) The need for greater efforts on the part of the medical profession to engage in clinical audit;

e) The value of the more widespread use of consensus techniques of varying kinds in order to develop clincial guidelines and appropriateness indications;

f) The importance of further analysing the reasons for variations, including the relationship between epidemiology and use rates, and the relative importance of demand factors, supply factors and practice style;

g) As part of f), the need to complement statistical analyses of large data sets with research into the component parts of practice style to determine how treatment decisions are made by clinicians;

h) The possiblity of developing standards for use by policy makers and managers, for example concerning GP referral rates, use of outpatient services, and the number of operations to be performed for a given population in particular specialties.

There would seem to be considerable value in linking the work of the Institute with that taking place in Health Services Development on consensus conferences and changing clinical practice (see especially points c, d and e above). In terms of future research, there would also seem to be value in a local study integrating data on variations in inputs, provision, use rates and outcomes. As the DHSS argued in its review of acute services, any serious attempt to evaluate variations in activity and resource usage should focus on the district level, starting with one or two pilot studies (DHSS, 1981). This is already done in part through the use of performance indicators, but more detailed analysis is needed. There are several possible approaches, but one

d) The need for greater efficacy of the

engage in clinical audit;

e) The value of the more widely used

kinds in order to develop clinical

indications;

f) The importance of regular monitoring

involving the relationship between

relative importance of clinical

styles;

g) As part of (f), the more

date only with reported

incorporate new findings

h) The practicing of

management

outpatient

practice style factor.

6. There is little agreement in the literature on the correct or appropriate use of services and there is a continuing debate on whether high rates signify unnecessary usage or low rates signify under provision. If in the USA the implication of much of the work done is that some services are overprovided, in the UK the reverse often holds.
7. One of the difficulties in resolving these issues is that there are few data on the outcomes associated with different treatments, on the pattern of morbidity by area, or on appropriate indications for use.
8. In general, the literature raises more questions than it answers. These questions provide fertile territory for health policy analysis.

REFERENCES

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

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

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

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

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

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

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

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

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

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

Dalziel M and Garrett C (1987) 'Intraregional variations in treatment of end stage renal failure' BMJ, Vol 294, 30 May, pp. 1382-3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

CHAPTER 3

TECHNOLOGY ASSESSMENT IN THE UNITED KINGDOM

The lack of adequate mechanisms for evaluating technology and managing its introduction in the UK has been highlighted by a number of commentators. This emerged clearly from the analysis by Stocking and Morrison (1978) of the impact of whole body computed tomographic scanners. Stocking and Morrison traced the way in which scanners were developed in the medical equipment industry through their diffusion in the NHS to the point at which evidence about the impact of scanners became available. They showed that there was no clear policy for managing the introduction of body scanners, nor was there evidence of their effectiveness for imaging different organs before scanners were introduced. Furthermore, when evidence did emerge, it suggested that scanners were of less benefit than had been originally anticipated from experience of head scanners. On the basis of this example, Stocking and Morrison argued that there should be a greater investment in the assessment of new technologies, that the introduction of such technologies should be managed more effectively, and that a new organisation should be established to carry out assessments in order to inform policy decisions.

Similar conclusions were reached by the Council for Science and Society's Report, Expensive Medical Techniques (Council for Science and Society, 1982). This report drew attention to the lack of evaluation of new techniques, their too rapid diffusion, and the over-enthusiasm which often accompanied the introduction of innovative procedures. The report also argued that the

and Jennett offer a consistent view on the current state of technology assessment in the UK. To explore this further, this chapter considers in more detail what existing institutions and organisations are doing. In broad terms, the following organisations and activities are relevant to our discussion: DHSS, health authorities, consensus conferences, industry, economic appraisals, MRC/clinical trials, specialist academic units, and drug licensing and surveillance. Professional associations also play some part in technology assessment but their role is considered in chapter 6.

DHSS

Much of the interest shown by the DHSS in medical technology concerns the use of equipment. The Scientific and Technical Branch (STB) of the DHSS has responsibility for assessing the technical performance, safety and mechanical reliability of medical equipment but its work does not encompass clinical effectiveness. STB provides funding to support 3 sorts of activity: equipment evaluation, research and development and pump-priming to assist the evaluation of new equipment. A recent analysis conducted by the Cabinet Office (1986) indicated that in 1985/86 some £2m was spent on equipment evaluation. The analysis reported that evaluations are initiated by DHSS technical officers. The equipment concerned is bought by the DHSS and handed over to the evaluating body, normally a health authority. The results are reported in the Department's bulletin, Health Equipment Information, and these results are thought to be important in influencing purchasing decisions by health authorities.

A further £2m was allocated to research and development in 1985/86. This was distributed to universities (40%), hospitals (34%) and industry (26%). Two

thirds of the budget went to three major service areas: aid for the disabled, pathology and radiology. In addition, DHSS provided £1.2m a year to support the Bio-Engineering Centre at Roehampton. This Centre is involved in the development of artificial limbs. The research and development budget is overseen by a Research Liaison Group comprising civil servants and external advisers.

Pump-priming funds of about £0.5m a year are used by STB to support the purchase of new equipment at an early stage of production. The DHSS buys the equipment, gives it to a health authority, and an assessment is then prepared. The pump-priming programme is similar to the arrangements for equipment evaluation, except that the initiative usually comes from the manufacturer.

More generally, DHSS plays some part in assessing new medical technologies other than equipment and in funding their development. As Stocking (1987) notes the usual procedure will be for the Chief Medical Officer to seek advice on a technology from the Standing Medical Advisory Committee or one of its sub-committees. If the advice is accepted, guidance will be issued to health authorities. In some circumstances, special studies are commissioned, as in the economic evaluation of heart transplants (Buxton et al, 1985); in other circumstances, internal analysis will be undertaken, often through multidiscplinary teams drawing on medical, economic and policy expertise. The Department also provides central funding to support the introduction of some new technologies. For example, in announcing allocations to health authorities for 1987/88, the Secretary of State for Social Services set aside £40m for a number of supra-regional specialised services including liver transplantation, specialised liver services, endoprosthetic services for

primary bone tumors, neonatal and infant cardiac surgery, heart transplantation and services for the treatment of end stage renal failure in children. These allocations were separate from the normal budgets of health authorities and were specifically earmarked for spending on the services concerned.

One point to note about the DHSS is that technology assessment is an activity in which a number of different branches and divisions are involved. These include the Office of the Chief Scientist, the Economic Advisers' Officer, the Scientific and Technical Branch, and the policy divisions. To ensure proper liaison between these groups, the Chief Scientist's Coordinating Group on Health Technology Assessment has recently been formed within the Department with the following terms of reference:

> "To consider and recommend action in the field of assessment and evaluation of new and established medical procedures and techniques. Clinical, technical and economic factors will usually be considered (for example medical need, product or procedure safety and financial implications could form a basic analysis). The remit specifically includes medical treatments and drug therapies where these are based on products licensed by the Committee on Safety of Medicine. Appropriate techniques of disciplined enquiry would include formal clinical trials and quick/soft (QS) studies etc; short, medium and long term studies would all be of interest" (DHSS, personal communication).

Health Authorities

There is no set procedure followed by health authorities in reaching decisions

on the use of medical technologies. The initiative in deciding whether to purchase a piece of equipment or to develop a new, specialised service usually comes from clinicians. In some cases the initiative will proceed within existing budgetary allocations, in other cases additional funding will be required, particularly if an expensive item of equipment is involved. If extra resources have to be found, the proposal will be discussed by the doctors supporting it, their medical colleagues and local managers. Priorities will be established, and, if the proposal is accepted, the money may be found from the district budget, regional sources or through fund raising.

Expensive items of equipment will usually be funded by RHAs and at this level the regional scientific officers perform a key role. Acting in close collaboration with their colleagues in the DHSS, these officers have specific responsibilities for advising on the acquisition of medical equipment. The Cabinet Office analysis of medical equipment noted that regional scientific officers have become increasingly authoritative in recent years and have frequently challenged clinicians' recommendations at regional levels. Since the abolition of the Health Service Supply Council in 1985, certain regions have been given designated responsibility for particular classes of equipment.

Also important, both in relation to equipment and to technology generally, are the community physicians employed by health authorities. Health services evaluation is one of the designated functions of community physicians, although in practice the effectiveness with which this function is performed varies between districts. Community physicians rely for much of the time on

guidance offered by medical advisory committees at district and regional levels and on the views of key individuals who are both knowledgeable and at the leading edge of new thinking in their field (Council for Science and Society, 1982). They may also receive support from community physicians based in academic departments. A significant recent initiative in this field is the establishment of a health care evaluation project between the Frenchay Health Authority and the Department of Epidemiology and Community Medicine at the University of Wales Medical School at Cardiff. This involves the creation of a joint post of senior lecturer in health care evaluation to carry out technology assessment work in the Health Authority.

Few health authorities employ health economists to assist in technology assessment but some do buy in the expertise of economists based in academic departments. The Economic Advisers' Office in the DHSS is involved in work on technology assessment but this has been on a modest scale to date. Commenting on this issue, the Cabinet Office report noted:

> "There is... a need for evaluation of the economics of new techniques and new equipment. Often this will need to be based on practical trials, which it would obviously be inefficient for every health authority to conduct independently. This might be done by DHSS itself, but there are alternative possibilities, for example MRC through its new Health Service Research Committee or an independent institute for health economics" (Cabinet Office, 1986, p35).

As this comment suggests, there is considerable scope for introducing more

rigorous procedures for the assessment of medical technologies in health authorities. Although in theory the development of health services is carefully planned and managed, in practice new technologies are often introduced as a result of skilful lobbying by the clinicians concerned and not because they are part of agreed policies. A recent example in one district was a decision to buy a whole body CT scanner (Ham, 1986), but the Council for Science and Society noted that expensive medical techniques often "turn up first as a cuckoo in an unsuspecting district's nest" (1982, p18).

Among the reasons for this are:

- the replacement of a piece of diagnostic equipment with a new item with unforeseen consequences for workload and expenditure;
- the appointment of a new consultant with a special interest which has significant resource consequences;
- the launching of a fund raising campaign for a new piece of machinery and the associated public pressure to provide a new service.

The introduction of general management into the NHS combined with tight cash limits has created a strong countervailing force, but it remains the case that professional pressures for the adoption of new techniques are often irresistible. This was certainly the view of the Griffiths report which noted that "clinical evaluation of particular practices is by no means common and economic evaluation of those practices extremely rare" (Griffiths Report, 1983, p10). In the light of the Cabinet Office report, the DHSS has taken the initiative to develop guidance for option appraisal for medical equipment, and this may go some way to introducing more rigorous methods of assessment.

Consensus Conferences

Consensus development conferences were initiated in the United States by the National Institute of Health in 1977 (Jacoby, 1985). The aim of these conferences is to assess specific technologies in a public forum making use of a panel hearing evidence from experts. Over 50 consensus conferences have been staged in the United States. The first UK consensus conference was held under the auspices of the King's Fund in November 1984 on the subject of coronary artery bypass grafting. This was followed in October 1986 by a second conference on the treatment of primary breast cancer. A third conference on the role of asylum in the care and treatment of people with mental illness was held in April 1987. The series now goes under the name of the King's Fund Forum.

Stocking (1985) has noted 2 important differences between the approach used in the United States and the UK. First, the composition of the consensus panel in the United States ensures that panel members are all knowledgeable about a subject, although not all are medically qualified. As such the American conferences approximate to a system of peer review. In contrast, in the United Kingdom the approach is more akin to a jury system, panelists not usually being experts in the field. The second difference concerns the questions examined and the evidence used by the panel. In the United States the focus is exclusively on the scientific issues, whereas in the United Kingdom, issues of costs and the implications for services are also considered.

Consensus conferences in the UK extend over two to three days. During this time, experts give evidence and are questioned, and the audience is invited to

participate in the debate. Taking account of all views expressed, the panel produces a consensus statement giving their views on the main issues and making recommendations for the use of a technology. In the case of coronary artery bypass grafting , four questions provided a framework for the panel's work. These were:

(1) What are the pros and cons of coronary artery bypass surgery (compared with alternatives) for various types of patients (including age and sex), in terms of survival and quality of life?

(2) What are the indications for various investigations for coronary artery disease?

(3) What size are the potential pools of patients for investigation and for coronary artery bypass surgery, taking account of alternative therapies? Are these estimates likely to change substantially over the next five to ten years?

(4) What would be the cost and implications for service organisation of increased provision for investigation and therapy?

The panel concluded that a strong case had been made for CABG as the most effective treatment in the case of intractable angina. Accordingly it recommended that a lead should be given nationally to increase provision, aiming for a rate of 300 operations per million population, compared with the prevailing rate of 169 operations per million in the Thames regions and 47 operations per million in the UK as a whole. The consensus statement was

published in the British Medical Journal and widely disseminated in the NHS.

An evaluation by the organisers of the CABG conference reported that the event had been important and should be repeated on other issues. Among specific criticisms identified by panelists, speakers and the audience, the most significant concerned the question of bias in the evidence presented, in particular in favour of CABG. There was also criticism of the limited amount of time for audience participation, and of the decision to ask the panel to work overnight to prepare its statement. These points were taken into account in the planning of subsequent events. An evaluation of the impact of the consensus statement was also put in hand.

As we discuss in the next chapter, several European countries have also experimented with consensus conferences, as has Canada. Here we may note that critics have identified a number of weaknesses in the consensus conference method, particularly as developed by the NIH in the United States (Lomas, 1986). These include:

- proceedings can be dominated by persuasive individuals with strong views;
- conferences tend to produce recommendations that are too general;
- the method avoids issues where good data are lacking and does not generate new data;
- consensus statements rely too much on compromise;
- consensus statements make little impact on practice.

Against this, it should be noted that consensus development conferences have

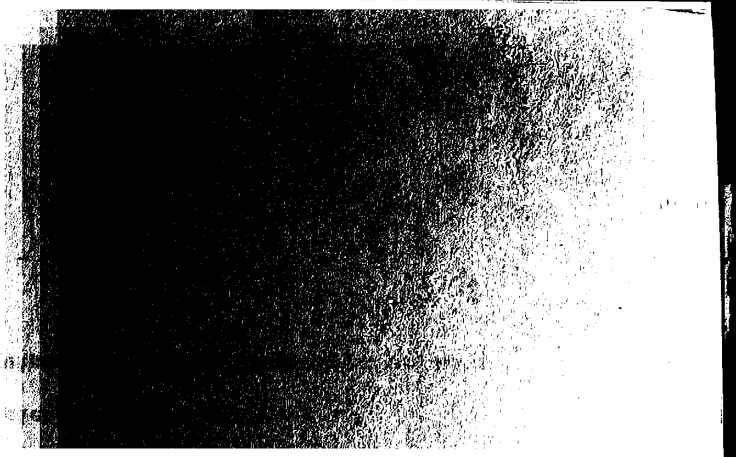
the potential to enable research to be brought together into practical policy recommendations, they enable conflicting views to be reconciled, they draw on professional and public experience, and they obtain evidence from front line practitioners.

Industry

The Cabinet Office review of the medical equipment industry provided a useful analysis of the role of the industry and the relationship between the industry and the NHS. The review noted that the industry has an annual turnover of £1billion and makes an important contribution to employment and exports. The companies making up the industry are diverse, some being mainly involved in the manufacture of medical equipment, others having this as only one part of their overall activities. The great majority of establishments (79%) were small, employing fewer than 20 employees. The review reported that there was evidence to suggest that the industry was falling behind in areas of high technology and surrendering the lead to overseas competition.

To reverse this trend the review called for greater support for the industry from the NHS and DHSS. This included increasing government support for research and development for medical equipment, encouraging collaboration between the industry and the health care professions, and increasing capital equipment provision in RHAs and DHAS. As the review noted in an important section:

> "the NHS is inherently conservative in genuinely new medical technologies, expecting their efficiency to be clearly demonstrated before it will buy equipment in significant



quantities. This can hardly be faulted on medical grounds, but it does stand in contrast to the USA which is , if anything, over receptive to innovation...it will need a deliberate effort to create market openings in the NHS for new technologies" (p40).

The main message of the review was that the medical equipment industry needed support from the home market, in particular the NHS, to enable it to tackle overseas markets successfully.

The review provided a nice example of the dilemmas facing government in the area of technology assessment. On the one hand, government wishes to support the medical equipment industry because of the contribution it is able to make to the economy. On the other hand, government wants to secure the orderly assessment and introduction of equipment and technology to ensure that health service expenditure is allocated optimally. The result, as Stocking has noted, is that "some countries seem to have one foot on the accelerator and one foot on the brake as far as diffusion of technologies is concerned" (Stocking, 1987).

Economic Appraisal

Health economists have been at the forefront of technology assessment activities in the United Kingdom. A review of economic appraisals of health technology carried out by Drummond and Hutton (1985) listed 71 studies published between 1971 and 1985. The vast majority of these were conducted on the independent initiative of researchers and they covered a wide range of technologies including CT scanners, coronary care, neonatal intensive care,

open heart surgery and long term care for priority groups. Drummond and Hutton also noted the role of the DHSS in encouraging economic appraisals, in part through work carried out in house but more particularly through commissioning work conducted by others.

Drummond and Hutton's review pointed out the variable quality of many of the appraisals undertaken, and the limited impact of much of the work done. They attributed this lack of impact to the gap that often existed between the researchers and policy makers and the fact that few appraisals were originated by policy makers. A further contributory factor was the publication of the results of appraisals in academic journals not read by policy makers. In addition, health authorities were not required to perform economic appraisals of new technology in the same way as they were required to do option appraisals of capital developments.

The principal capacity for undertaking economic appraisal exists in academic centres. Of these, the most significant are those at York University and Aberdeen University. There are in addition a number of other smaller centres, such as those that exist at Brunel University, St Thomas' Hospital Medical School, and the Health Services Management Centre, Birmingham University. The Office of Health Economics has also done considerable work in this field, although in the eyes of some its well known association with the pharmaceutical industry casts doubt on its independence. From the Institute's point of view, one of the clear implications of the work done by health economists is the importance of linking technology assessment to the needs of policy makers and presenting results in a form which is accessible and likely to be read. Timely, readable reports, may have more impact than volumes of

carefully conducted research presented after decisions have been taken. An example which leads some support to this point is the study by Buxton and others of heart transplants (Buxton et al, 1985), funded by DHSS and resulting in a continuation and expansion of the transplant programme.

Specialist Units

A good example of a specialist unit involved in technology assessment is the National Perinatal Epidemiology Unit at Oxford University. This was established in 1978 with the following terms of reference: "to conduct epidemiological research in the perinatal field with a view to providing information which can promote effective use of resources in the perinatal services". The Unit had a core staff of eight during 1985 and a number of project staff. Three quarters of the Unit's funding is provided by DHSS, the remainder coming from research councils and foundations and health authorities.

As the Unit's title indicates, its major concern is with perinatal health services . This encompasses pregnancy, childbirth, early parenthood and early childhood. Staff bring various perspectives to bear in analysing these issues - medical, social and economic. A key principle which lies behind the Unit's work is that routinely collected data should be used to the full . Accordingly, considerable emphasis is placed on literature reviews and compilations and analyses of existing data sets. These activities are supplemented by studies to obtain data not available from other sources, and collaborative research with others, often clinicians.

Of particular interest is the investment made in meta-analysis. This involves

bringing together the results of a number of controlled trials of a particular technology and using the results to provide more comprehensive evidence about the technology than would be available from any single source. Meta-analysis proceeds through a number of stages: first, all relevant trials are identified; second, the quality of each trial is assessed; and third, the results of similar trials are analysed within a pooled analysis. The Unit has undertaken a number of meta-analyses including studies of the effects of continuous electronic foetal heart rate monitoring in labour and the effects of routine screening with ultrasound in pregnancy.

Staff of the Unit have made an impressive contribution to the analysis of perinatal services during the last decade, much of it in the field of technology assessment. This is well illustrated by the annual report for 1985 which lists 256 publications since the Unit was established. However, it is worth noting that the vast majority of these publications have been in medical and scientific journals, or have been in book form. It would seem that the main audience for the Unit's output is the medical and academic community. Few efforts appear to have been made to influence directly health service policy makers.

It is worth emphasising the multidisciplinary approach to technology assessment taken by the Unit. Given the multiple implications of health care technologies, this approach is of particular value. Like the Unit, the Institute is well placed to conduct multidisciplinary assessments, exploring the social and organisational aspects of technology, as well as the clinical and economic dimensions.

MRC/Clinical Trials

The most significant contribution made by the MRC to technology assessment has been in the field of randomised controlled trials, a field in which Britain has taken a leading role. The work of the Council in this area has involved supporting and coordinating trials conducted at a number of centres. Examples include work on cancer trials, carried out in association with cancer charities, and trials involving the treatment of hypertension. Furthermore, in recent years the MRC has taken a close interest in magnetic resonance imaging and has sponsored both economic and clinical evaluations of this technology. The MRC also gives support to specialist units, including the National Perinatal Epidemiology Unit.

Despite the work funded by the MRC, a consistent theme in the literature is the limited part played by the Council in technology assessment. Thus , Stocking and Morrison noted "it would seem appropriate that the MRC should play a much more active role in the clinical evaluation of CT scanning and other medical technology" (1978, p64). This was supported in the analysis conducted by the Council for Science and Society which noted the emphasis placed by the MRC on bio-medical research as opposed to clinical research. The same point has been made by Jennett (1986) who has drawn particular attention to the relative neglect of clinical research by the MRC, particularly research concerned with the evaluation of medical practices and procedures. Instead of encouraging and actively inviting proposals for evaluative research, the MRC has tended to react to proposals initiated by others. As a result, the activities of the Council have been compared unfavourably with those of its sister organisations in Sweden and the United States, discussed more fully in the following chapter.

Drug licensing and surveillance

Procedures for controlling the introduction of new drugs and monitoring their effects in use are more rigorous than those which apply to other health care technologies. These procedures stem from the 1978 Medicines Act. The Act seeks to protect consumers by improving the safety, quality and efficacy of drugs (see Hartley and Maynard, 1982). Specifically, the Act provides for:

- a licensing system operated by the Medicines Division of the DHSS;
- a Medicines Commission to advise Ministers;
- a series of expert committees including the Committee on Safety of Medicines; and
- controls of drug advertising and promotion.

The procedures introduced in 1968 replaced previous voluntary arrangements which were found to be inadequate in the light of the thalidomide tragedy.

The procedures have been criticised by consumer groups for failing to introduce sufficiently tight controls. They have also been criticised by the pharmaceutical industry for excessive bureaucracy and for causing delays in the introduction of new drugs. In 1987 the government announced that it was setting up a review of drug licensing arrangements to establish how the backlog of applications for licensing might be eased. It was suggested that one possibility was for an independent medicines board to be created in place of the Medicines Division in the DHSS (Guardian, 19 March 1987). Whatever the outcome of the review, the basic principles enshrined in the Medicines Act are of considerable interest in the broader context of technology assessment, not least in offering one model of how the introduction of new technologies might

[redacted]

be controlled and their effectiveness kept under review. However, it should be emphasised that the procedures for assessing drugs are solely concerned with safety and do not encompass other key elements normally associated with technology assessment.

Summary

Analysis of UK experience points to the following conclusions:

(1) To date, the DHSS has been concerned mainly with medical equipment and drugs. In the case of equipment, the Scientific and Technical Branch of the Department has focused principally on technical performance, safety and mechanical reliability;

(2) Health authorities do not follow a set procedure in assessing health care technologies. Regional scientific officers play an important part in relation to equipment, and community physicians may be involved in health services evaluation. Although in theory the development of health services is carefully planned and managed, in practice new technologies are often introduced as a result of skilful lobbying by the clinicians concerned, often supported by patients and the public, and not because they are part of agreed policy;

(3) Consensus conferences have developed as one way of reviewing a diverse range of evidence and offering guidelines for clinical practice and policy makers;

(4) The medical equipment industry has an ambivalent relationship with

government. Government support of the industry on economic grounds may run counter to attempts to introduce more rigorous systems of technology assessment;

(5) There has been a significant investment in economic appraisal of health care technology but the results of appraisals have had a limited impact. Often, this is a result of the gap that exists between researchers, policy makers, and users (clinicians and their patients);

(6) The National Perinatal Epidemiology Unit is a good example of a specialist academic unit involved in technology assessment. Of particular interest is the work done by the Unit in analysing existing data;

(7) The MRC has made an important contribution to technology assessment through randomised controlled trials but in general the Council's contribution to technology assessment has been modest;

(8) Procedures for drug licensing and surveillance offer one possible model for technology assessment in other areas, although the focus is narrower than technology assessment as defined in earlier chapters.

REFERENCES

Buxton M et al. (1985) Costs and Benefits of the Heart Transplant Programmes at Harefield and Papworth Hospitals, DHSS Office of the Chief Scientist Research Report No12, HMSO, London

Cabinet Office (1986) Advisory Council for Applied Research and Development. Medical Equipment, HMSO, London

Council for Science and Society (1982) Expensive Medical Techniques, London

Drummond M F and Hutton J (1986) Economic Appraisal of Health Technology in the United Kingdom, Centre for Health Economics Discussion Paper II York

Griffiths Report (1983) NHS Management Inquiry, DHSS, London

Ham C J (1986) Managing Health Services: Health Authority Members in Search of a Role, School for Advanced Urban Studies, University of Bristol, Bristol

Hartley K and Maynard A (1982) The Costs and Benefits of Regulating New Product Development in the UK Pharmaceutical Industry, OHE, London

Jacoby I (1985) 'The Consensus Development Program of the National Institutes of Health', International Journal of Technology Assessment in Health Care, Vol 1, No2, pp420-432

Jennett B (1986) High Technology Medicine, Oxford University Press, Oxford New Edition

Lomas J (1986) 'The consensus process and evidence dissemination', CMAJ, Vol 134, June 15, pp. 1340-1341

Stocking B and Morrison S L (1978) The Image and the Reality, Nuffield Provincial Hospitals Trust/Oxford University Press, Oxford

Stocking B (1985) 'First consensus development conference in United Kingdom: on coronary artery bypass grafting'. Views of audience, panel, and speakers, BMJ, 14 September, 291, 713-716

Stocking B (1987) Book in press on medical technology in the EEC

CHAPTER 4

TECHNOLOGY ASSESSMENT IN EUROPE AND NORTH AMERICA

The aim of this chapter is to illustrate how technology assessment is carried out in Europe and North America. As in previous chapters, the approach is selective, the purpose being to indicate how a small number of countries have involved themselves in technology assessment, rather than to offer a comprehensive review. Following some general introductory comments on the state of technology assessment in the EEC, the chapter analyses in detail the experiences of three countries: the Netherlands, Sweden and the United States. Sweden and the United States are examined because they are widely regarded as being at the forefront of medical technology assessment. The Netherlands is included as an example of an EEC country whose experience provides an interesting contrast to that of the UK. The conclusion highlights differences and similarities between the three countries.

Technology Assessment in the EEC.

A conference held at the King's Fund in April 1986 brought together a wealth of information about medical technology and technology assessment in the EEC. Country reports were prepared for the member states of the EEC and these reports included analyses of how six technologies were deployed in each country. At the conference, particular attention was paid to the way in which decisions on the use of technology were made in practice. This included a specific focus on the regulatory mechanisms in place in each country for controlling medical technology. The country reports and conference papers are

"provides an interesting [illegible]

differences and similarities [illegible]

Technology assessment in [illegible]

A conference held at the [illegible]

of information about medical [illegible]

Country reports were prepared [illegible]

reports included analyses of [illegible]

country. At the conference [illegible]

decisions on the use of technology were made [illegible]

specific focus on the regulatory mechanisms [illegible]

controlling medical technology. The country [illegible]

due to be published in the second half of 1987 (Stocking, 1987) and they have been drawn on extensively in this part of the chapter.

The papers revealed great diversity in the approach adopted to technology assessment in the EEC. In his contribution, Groot (1987a) distinguished between health care systems owned, planned, managed and financed by public authorities, and health care systems planned by public authorities and financed by insurance agencies belonging to the social security system. The former group included Denmark, Ireland, Italy and the UK, the latter comprised Belgium, Germany, Greece, France, Luxembourg and the Netherlands. The publicly run and financed systems tended to control medical technology through budgetary constraints and this often delayed the use of technologies. In contrast, social security systems relied more on specific regulations and guidelines to limit the use and diffusion of expensive technologies and to control the expansionist tendencies built into these systems.

Despite these differences, Stocking (1987) noted that health service agencies in all countries were generally reactive and passive in their approach to medical technology. No country could be said to be involved in technology assessment in a systematic way, although the elements of a comprehensive strategy were evident in each country. Against this background, it is worth noting Groot's view that the UK is at the forefront in the EEC in assessing new technologies (Groot, 1987a). This is reinforced by Drummond's conclusion that the UK is the European country with the strongest tradition of economic evaluation in health care (Drummond, 1987). Accepting this analysis, and making use of Groot's distinction between public health care systems and

health insurance systems, we now examine in more detail the experience of the Netherlands to illustrate the approach to technology assessment adopted in one insurance-based system. As Stocking (1986) has noted, the Netherlands is the insurance-based system which "has shown perhaps the most interest in technology assessment" (p 27) and its experience is therefore of relevance to our analysis.

The Netherlands

As in the UK, technology assessment in the Netherlands is undertaken in a number of ways. For our purposes, three aspects of the approach adopted in the Netherlands are of particular interest. First, the National Organisation for Quality Assurance in Hospitals (CBO) has organised a series of consensus conferences (see Casparie et al 1987). These conferences stemmed from an interest in quality assurance rather than technology assessment. The aim was to provide national guidelines on important medical issues as an aid to assessing the quality of clinical practice. The main emphasis is on achieving a professional consensus on the technology concerned, and by the end of 1987 21 conferences will have been organised. The topics of these conferences are illustrated in the accompanying table.

Health insurance systems

Netherlands

Insurance-based

Insurance-based

Technology

our analysis

The Netherlands

As in the US,

number of

The Parliamentary

for Quality

conference

interest in

to provide

research in

a professional

of conference

Clinical

The main emphasis is on acute

[illegible] reviewed, and by the end of [illegible]

[illegible] The topics of these conferences are

Table 1: Consensus Development Conferences Organised in The Netherlands with Assistance of CBO

Year	Topic
1982	Blood transfusion
1983	Traumatic lesions of the back
	Breast Cancer
1984	Serious brain damage
	Melanoma of the skin
	Platelet transfusion
1985	Solid, solitary thyroid nodule
	Prevention of bedsores
	Osteoporosis
	Diabetic foot
1986	Diagnosis of deep venous thrombosis
	Orchidopexy
	Treatment of bedsores
	Treatment of drug-addiction in prison
1987	Haemophilia
	Follow-up of colon polyps
	Diagnosis of lymph nodes in the neck
	Diagnosis of atopic syndrome
	Prevention of Herpes in the new born
	Hypercholesterolaemia
	Total hip replacement
1988	Screening for breast cancer

The procedure used is as follows. Topics are selected according to the amount of controversy, the relevance to health care, the availability of scientific data, the consequences for medical practice, and the estimated chance of reaching consensus. An expert working committee is then appointed and the committee develops a draft consensus statement with support from CBO staff. This draft is sent with background papers to conference participants. The conference itself involves presentations by working committee members and discussion with the audience, made up of doctors and other health service personnel. A final consensus is prepared taking account of these discussions.

These procedures have evolved over time and have been amended in the light of experience. Casparie et al note the importance of consensus statements offering practical guidelines rather than vaguely formulated compromises. They also stress the value of letting a group of experts work together over a period of months as a way of building cohesion and creating responsibility for consensus development. A key point to note is that Dutch consensus conferences involve less open debate and discussion than in other countries and are more akin to professional state of the art meetings (Vang, 1986).

The second aspect of the Netherlands approach which is of interest is the attempt to plan the provision and diffusion of specialist services throughout the country (Groot, 1987b). Under Article 18 of the Hospital Provisions Act, the following 10 technologies are subject to control:

(1) renal dialysis;
(2) kidney transplantation;
(3) radiotherapy;
(4) neurosurgery;
(5) cardiac surgery;
(6) nuclear medicine
(7) prenatal chromosome examination
(8) heart catheterisation
(9) CT scans
(10) neonatal care for early births

Hospitals cannot provide these services without government approval. National plans and special funding have also been provided for heart transplantation,

liver transplantation, pancreas transplantation, bone marrow transplantation, lithotripters and MRI scanners. Teaching hospitals are not subject to these controls as their affairs are supervised by the Minister of Education rather than the Minister of Health (Groot, 1987b). Partly for this reason, and also because the technologies subject to control represent only a small proportion of total health service activity, national planning has had only limited success in controlling costs.

Third, a distinctive contribution made by the Netherlands has been the study of health care services undertaken by the Steering Committee on Future Health Scenarios in the Ministry of Health. This has involved the creation of alternative pictures of possible and desirable futures in the field of public health care. The Steering Committee selects specific topics for detailed analysis and an independent scenario committee is then established. Attached to each committee is a research team. In 1983, work started on ageing, cardiovascular diseases, lifestyles, cancer, and medical technology. The project on medical technology involved the identification of new medical technologies at an early stage, and a prospective study of the consequences of these technologies. Six areas have been examined in depth: neurosciences, bio-technology, genetic testing, laser technology and coronary artery surgery, imaging techniques, and home care technologies. The first reports from the project were published in 1987 in association with an international conference.

The principal conclusion reached by the Commission on Future Health Care Technology was that a permanent, well-funded system for technology assessment in health care should be developed in the Netherlands (Steering Committee on

Future Health Scenarios, 1987). The system should apply to new and existing technology as well as to future possibilities because of the general lack of knowledge concerning the benefits, risks, costs and social implications of health care technology. The Commission argued that such a system would identify technologies requiring assessment, collect data for making an assessment, synthesise such data, and disseminate the results, especially to policy makers. As we note in the next section, similar proposals have recently been put forward in Sweden.

Sweden (Ministry of Health and Social Affairs 1986)

Like the NHS the Swedish health service is publicly funded and publicly provided. At national level responsibility for health services rests with the Ministry of Health and Social Affairs and the National Board of Health and Welfare, while locally it is the county councils who provide services. Considerable interest has been shown in technology assessment by a variety of agencies. Of particular relevance is the work done by the Medical Research Council which actively supports the assessment of medical technology as an area of priority. Since 1977 the Council has had a separate division for health services research and an expert committee of advisers on medical technology assessment. The committee has initiated research projects on particular technologies and has been involved in the Swedish consensus conference programme. Stocking has commented that of all the research councils in Europe, the Swedish MRC "is one that has shown perhaps the most interest in technology assessment" (1986, p22).

The Swedish Planning and Rationalisation Institute (SPRI) is another body active in this field. SPRI is funded jointly by national government and the

county councils and it is particularly concerned with the economic consequences of medical technologies. SPRI has actively promoted the need for technology assessment within Sweden and it has been involved with the MRC in organising consensus conferences. Seven conferences have been held to date covering a wide range of topics including the treatment of depressive disorders, diagnostic imaging of liver tumour and total hip joint replacement. An evaluation of the conference programme completed in 1986 indicated that there was high awareness of the conferences among doctors, politicans and administrators. The evaluation also demonstrated that politicians and administrators placed a particular value on consensus statements. Although it was found that only a small proportion of doctors reported changing their practice in the light of consensus statements, the evaluators argued that influencing the behaviour of doctors was difficult to achieve and even a small shift in practice was to be welcomed (Calltorp, 1987).

SPRI has also assisted in the establishment of medical care programmes. These programmes are written local agreements containing guidelines governing the content and organisation of care and services to be offered to individuals with a given disease or risk of disease (SPRI, 1985). A further aspect of SPRI's work is its involvement in a joint Nordic association for medical technology known as Nordic Evaluation of Medical Technology. Apart from SPRI, this comprises the Danish Hospital Institute, the Norwegian Institute for Hospital Research, and the Finnish Hospital Association.

Two other bodies involved in technology assessment at national level are the National Board of Health and Welfare and the County Councils' Federation. The Board is responsible for national planning and it offers guidance to the

county councils making use of the work of expert committees. The County Councils' Federation represents the interests of county councils at national level. The Federation is also involved in shaping health policy and it works with the National Board in developing guidelines for the health service.

A recent development of considerable interest is the establishment of the Centre for Medical Technology Assessment at Linkoping University. The Centre is an independent research unit within Linkoping University and it is funded by one of the county councils. The main emphasis is on the economic aspects of medical technology assessment. Projects include an evaluation of extracorporeal, shock-wave lithotripsy and the cost effectiveness of beta-blockers in post-myocardial infarction care. In future, the Centre hopes to carry out assessments within the areas of primary care, home care and self care.

It can therefore be seen that Sweden has made a significant investment in technology assessment. Despite this, a review conducted by a task force appointed by the Minister of Health in 1985/1986 concluded that a need existed for a new independent agency for medical technology assessment in the Ministry of Health and Social Affairs. The functions identified for the agency were:

- to identify new and existing medical technologies in need of assessment;
- to formulate scientifically based syntheses of the value of different technologies based on medical, humanitarian and social-economic perspectives;
- to compile a knowledge base written in easily comprehensible language which provides different levels of detail for decision makers within

government, the county councils, and the health delivery system;

- to formulate strategies for accelerating the introduction of new and effective medical technologies;
- to develop strategies for decommissioning and replacing less effective medical technologies;
- to transfer information concerning assessment results;
- to incorporate international experiences and results from various assessment activities;
- to serve as a national and international contact concerning medical technology assessment.

The task force emphasised that the agency might need to conduct limited studies, but it should not carry out original research nor fund research. A budget of SEK 10 million (approximately £1 million) was proposed. A decision on the task force's report is awaited.

United States

At first sight, the scale of technology assessment activities in the United States is impressive. A large number of agencies are involved in one aspect or another of technology assessment and annual expenditure that could be regarded as to do with technology assessment was estimated to be over $1 billion dollars in 1985 (Institute of Medicine, 1985, p9). Furthermore, organisations like the Office of Technology Assessment are held up as examples for other countries to follow, and methods of technology assessment such as consensus development conferences which originated in the United States have been taken up elsewhere. However, a recurring theme in analyses of US experience is the fragmented nature of these activities, and the associated call for greater coordination.

A number of agencies within the executive branch of the Federal Government are involved in technology assessment. One of the most important is the Food and Drugs Administration (FDA). FDA controls the introduction of new drugs and medical devices. It is a regulatory agency and the approval of the FDA must be obtained before a drug or device is brought into use. FDA concerns itself with safety and efficacy rather than with cost and effectiveness. Despite this, it has been argued that "the premarket approval processes for drugs and medical devices regulated by the FDA is the only coherent, coordinated system for medical technology assessment" (Institute of Medicine, 1985, p 60) in the United States.

Another important body is the National Institutes of Health (NIH). This is the American equivalent of the Medical Research Council. Like MRC, NIH allocates most of its budget to basic biomedical research. However, it does make a substantial investment in clinical trials ($275 million in 1985) and it is also involved in various forms of data sysnthesis, most notably consensus development conferences (CDC). These conferences are coordinated by the Office of Medical Applications of Research (OMAR) within NIH. Since the first conference was staged in 1977 on breast cancer screening over fifty conferences have been organised on a wide range of topics. The following guidelines are used in selecting topics:

(1) the subject under consideration should be medically important
(2) there should be a scientific controversy that would be clarified by the consensus approach or a gap between current knowledge and practice, that a CDC might help to narrow;
(3) the topic must have an adequately defined and available base of

scientific information to answer the previously posed questions;

(4) the topic should be amenable to clarification on technical grounds and the outcome should not depend mainly on the impressions or value judgements of panelists;

(5) the timing of the conference should be such that it is likely to have a meaningful impact ie, it should neither be so early in the developmental course of a new technology that data are insufficient nor so late that the conference merely reiterates a consensus already arrived at by the professions. (Jacoby, 1985).

Conferences usually last two and a half days. For the first one and a half days, invited experts make presentations to the panel in open session on the state of scientific knowledge about an issue. The panel then goes into executive session to prepare a consensus statement. This is presented at the final plenary session of the conference and it is widely disseminated.

Over the years, OMAR has refined the conference process in a number of ways (Jacoby, 1985). For example, panels containing members with opposing views on a topic are no longer used, and instead every effort is made to achieve neutrality in the composition of the panel and the selection of a chairman. Also, particular care is taken to frame the questions for the panel in a way that can be answered using available scientific information rather than opinions or value judgements. Perhaps the most significant change concerns the consensus statement itself. The emphasis now is on concrete and specific proposals rather than discursive, general statements. As these comments suggest, a particular feature of the NIH approach to consensus conferences is the concern to evaluate and improve the process in the light of experience.

In an attempt to achieve better coordination of the work of the FDA, NIH and other agencies involved in technology assessment, the National Centre for Health Care Technology (NCHT) was formed in 1978. In its three year life, NCHCT completed evaluations of 75 technologies, but it was beset by funding difficulties, bureaucratic infighting (including competition with NIH) and opposition from powerful external groups (Foote, 1986). Accordingly, its financing was not renewed, and its work now continues in a more limited way in the National Centre for Health Services Research and Health Care Technology Assessment (NCHSRHCTA) through its Office of Health Technology Assessment (OHTA). Like the NCHCT this is located within the Department of Health and Human Services (DHHS). The principal objective of OHTA is to conduct evaluations of selected technologies to assist the Health Care Financing Administration (HCFA) in deciding what techniques and procedures should be covered by Medicare.

OHTA is quite different from the Office of Technology Assessment (OTA) which works to support Congress in identifying the beneficial and adverse impacts of the application of technology. OTA started work in 1974 and its health programme was initiated in 1975. The programme has been described as "the largest and one of the oldest in OTA" (Herdman and Behney, 1985, p163) and by 1985 it had generated 24 main reports on technology assessment issues, 34 case studies, and other related technical memoranda and background papers (See Table 2)

TABLE 2

Main Reports

Drug Bioequivalence. July 1974.
Development of Medical Technology: Opportunities for Assessment. August 1976.
Cancer Testing Technology and Saccharin. October 1977.
Policy Implications of Medical Information Systems. November 1977.
Policy Implications of the Computed Tomography Scanner. August 1978.
Assessing the Efficacy and Safety of Medical Technologies. September 1978.
Selected Topics in Federal Health Statistics. June 1979.
A Review of Selected Vaccine and Immunization Policies Based on Case Studies of Pneumococcal Vaccine. September 1979.
Forecasts of Physician Supply and Requirements. April 1980.
The Implications of Cost-Effectiveness Analysis of Medical Technology. August 1980.
Assessment of Technologies for Determining Cancer Risks from the Environment. June 1981.
Cost-Effectiveness Analysis of Inactivated Influenza Vaccine. December 1981.
Technology and Handicapped People. May 1982.
Strategies for Medical Technology Assessment. September 1982.
Medical Technology Under Proposals to Increase Competition in Health Care. October 1982.
Postmarketing Surveillance of Prescription Drugs. November 1982.
Medical Technology and Costs of the Medicare Program. July 1984.
Federal Policies and the Medical Devices Industry. October 1984.
Blood Policy and Technology. February 1985.
Medical Devices and the Veterans Administration. February 1985.
Preventing Illness and Injury in the Workplace. April 1985.
Biomedical Research and Related Technology for Tropical Diseases. August 1985.

Technical Memoranda

Compensation for Vaccine-Related Injuries. November 1980.
Technology Transfer at the National Institutes of Health. March 1982.
MEDLARS and Heath Information Policy. September 1982.
Diagnosis Related Groups (DRGs) and the Medicare Program: Implications for Medical Technology. July 1983.
Quality and Relevance of Research and Related Activities at the Gorgas Memorial Laboratory. August 1983.
Scientific Validity of Polygraph Testing: A Research Review and Evaluation. November 1983.
Update of Federal Policies Regarding the Use of Pneumococcal Vaccine. May 1984.
Review of the Public Health Service's Response to AIDS. February 1985.

Background Papers

Computer Technology in Medical Education and Assessment. September 1979.
Methodological Issues and Literature Review. September 1980.
The Management of Health Care Technology in Ten Countries. October 1980.
Policy Implications of Computed Tomography (CT) Scanner: An Update. January 1981.
The Information Context of Premanufacture Notices. April 1983.
The Impact of Randomized Clinical Trials on Health Policy and Medical Practice. August 1983.

Case Study Series

Case Study No.

1. Formal Analysis, Policy Formulation, and End-Stage Renal Disease. April 1981.
2. The Feasibility of Economic Evaluation of Diagnostic Procedures: The Case of CT Scanning. April 1981.
3. Screening for Colon Cancer. April 1981.
4. Cost Effectiveness of Automated Multichannel Chemistry Analyzers. April 1981.
5. Periodontal Disease: Assessing the Effectiveness and Costs of the Keyes Technique. May 1981.
6. The Cost Effectiveness of Bone Marrow Transplant Therapy and Its Policy Implications. May 1981.
7. Allocating Costs and Benefits in Disease Prevention. May 1981.
8. The Cost Effectiveness of Upper Gastrointestinal Endoscopy. May 1981.
9. The Artificial Heart: Cost, Risks, and Benefits. May 1982.
10. The Costs and Effectiveness of Neonatal Intensive Care. August 1981.
11. Benefit and Cost Analysis of Medical Interventions: The Case of Cimetidine and Peptic Ulcer Disease. September 1981.
12. Assessing Selected Respiratory Therapy Modalities: Trends and Relative Costs in the Washington, D.C. Area. July 1981.
13. Cardiac Radionuclide Imaging and Cost Effectiveness. May 1982.
14. Cost Benefit/Cost Effectiveness of Medical Technologies: A Case Study of Orthopedic Joint Implants. September 1981.
15. Elective Hysterectomy: Costs, Risks, and Benefits. October 1981.
16. The Costs and Effectiveness of Nurse Practitioners. July 1981.
17. Surgery for Breast Cancer. October 1981.
18. The Efficacy and Cost Effectiveness of Psychotherapy. October 1980.
19. Assessment of Four Common X-Ray Procedures. April 1982.
20. Mandatory Passive Restraint Systems in Automobiles. September 1982.
21. Selected Telecommunications Devices for Hearing-Impaired Persons. December 1982.
22. The Effectiveness and Costs of Alcoholism Treatment. March 1983.
23. The Safety, Efficacy, and Cost Effectiveness of Therapeutic Apheresis. July 1983.
24. Variations in Hospital Length of Stay: Their Relationships to Health Outcomes. August 1983.
25. Technology and Learning Disabilities. December 1983.
26. Assistive Devices for Severe Speech Impairments. December 1983.
27. Nuclear Magnetic Resonance Imaging Technology: A Clinical Industrial, and Policy Analysis. September 1984.
28. Intensive Care Units: Clinical Outcomes, Costs, and Decisionmaking. November 1984.
29. The Boston Elbow. November 1984.
30. Market for Wheelchairs: Innovation and Federal Policy. November 1984.
31. The Contact Lens Industry: Structure, Competition, and Public Policy. December 1984.
32. The Hemodialysis Equipment and Disposables Industry. December 1984.
33. Technologies for Managing Urinary Incontinence. July 1985.
34. Cost-Effectiveness of Digital Subtraction Angiography in the Diagnosis of Cerebrovascular Disease. May 1985.

Whereas OHSTA in the executive arm of government undertakes assessments which result in decisions on whether or not a technology should be paid for by government, OTA deals more with broad policy implications and "its work leads to information and advice for general policy initiative more than specific decisions on individual technologies" (ibid p164).

The major assessments undertaken by OTA typically extend over 1-2 years and cost over $500,000. OTA concentrates on data synthesis rather than original research. Multidisciplinary teams of staff do much of the analysis involved in assessments but they work closely with a panel of advisers appointed to assist with each assessment. Further advice and expertise is provided by external consultants hired for this purpose. Panels usually comprise 10-20 individuals and meet 3-4 times during the conduct of an assessment. Panels are not required to reach a consensus. Rather, their role is "to represent relevant consistuencies and areas of expertise, to insure that assessment are complete, accurate, and fairly representative of viewpoints, and include reasonable options for Congress" (ibid, p165).

Outside government, a number of the specialist associations have shown interest in technology assessment. Two examples are the Clinical Efficacy Assessment Project (CEAP) established by the American College of Physicians in 1981 , and the Diagnostic and Therapeutic Technology Assessment Project (DATTA) of the American Medical Association. CEAP carries out 10-12 assessments each year (White and Ball, 1985). Assessments begin with an announcement in professional journals and an invitation to interested parties to send comments. Consultants are employed to review and synthesise the literature and they work with CEAP staff to produce a draft statement on the

technology in question. The statement then goes through a process of review, comment and consultation before final approval. CEAP statements vary in length from 1 paragraph to 10 pages. In essence, the statements offer guidelines to clinicians on the indications for a procedure, and statements become the official ACP position on the subject.

DATTA was established by the AMA in 1982. It involves panels comprising at least 40 physicians who rate procedures and therapies. DATTA assessments are based on the extent of agreement of the individual panelists. The panel operates by responding to questionnaires sent out by DATTA staff and assessments take 4-6 months. The resulting statement is not intended to be a standard or guideline but rather a reflection of medical opinion.

Despite the wide range of activities undertaken in the United States, it remains the case that technology assessment is not organised in a comprehensive or systematic manner. This is certainly the view of the OTA which in a series of reports has called for a bigger investment in technology assessment and greater coordination of existing activities.

Thus, Strategies for Medical Technology Assessment noted that existing arrangements for medical technology assessment "do not constitute a coherent system for assessing all classes of medical technologies. The present approach is characterised by multiple participants from the public and private sectors and by uncoordinated activities" (OTA, 1982, p3). OTA argued that there was a need to develop a strategy for assessing medical technologies, and it emphasised that " the most important policy need is to bring forth a rational, systematic approach from the present multiplicity of agencies and

comment and consultation before [illegible] statements vary in
length from 1 paragraph to 10 pages. In addition, the statements offer
guidelines to clinicians on the [illegible] for "procedures" and structures
become the official ICP position [illegible]

DATTA was established by the [illegible] comprising
least 40 physicians who [illegible]
based on the extent of [illegible]
operates by responding to [illegible]
assessment tube and [illegible]
standard or guideline [illegible]

Despite the wide [illegible]
remains the case that [illegible]
comprehensive or [illegible]
which is a series of [illegible]
assessment and greater [illegible]

Thus, Strategies for [illegible]
arrangements for medical [illegible]
system for assessing [illegible]
approach is characterised by [illegible]
sectors and by according [illegible]
there was a need to develop [illegible]
it emphasised that "the [illegible]
national, systematic approach [illegible]

activities to promote and coordinate medical technology assessment " (ibid, p18). A number of options were outlined for achieving greater integration.

An appendix to the report set out a model for an Institute of Health Care Evaluation (Bunker and Fowles, 1982). It was suggested that the Institute would be under the control of a consortium involving the government, the medical profession, representatives of the public, health maintenance organisations and insurance companies. The Institute's role would be to generate new data by supporting clinical trials, retrospective studies and other activities; to act as a communications clearing house, disseminating results of analyses to health professionals and policy makers; to identify newly emerging technologies; and to assist in the development of a uniform data base.

The OTA 's analysis has been echoed by the Institute of Medicine of the National Academy of Sciences. In particular, the Institute has proposed the establishment of a consortium for assessing medical technologies, involving partnership between the public and private sectors (Institute of Medicine, 1983). This idea has been reiterated more recently as part of a comprehensive overview of technology assessment in the United States undertaken by the Institute (Institute of Medicine, 1985). An initial level of funding of $30 million was proposed for the consortium increasing to $300 million over ten years. A first step in this direction has been made with the setting up in 1986 of a Council on Health Care Technology under the aegis of the Institute of Medicine, with support from public and private funds. The statutory purposes of the Council are to promote the development and application of appropriate health care technology assessments and to review existing health

care technologies in order to identify obsolete or inappropriately used technologies. One of the main functions of the Council is to act as a clearing house for information on health care technologies and assessment.

Despite this, Foote has noted that in the US "we do not have one super agency directing assessment activity, nor do we have a simple set of structures. Instead there is a complex web of interlocking institutions situated in both the legislative and executive branches of government, as well as in the private sector" (Foote, 1986). Whether this is interpreted as a manifestation of healthy pluralism or an example of uncoordinated and disjointed planning depends on the position of the commentator. What is certain is that US experience demonstrates the wide range of possible approaches to technology assessment. As such, it offers models and options to other countries seeking to expand their own investment in technology assessment.

Postscript

As a postscript to this discussion, it is relevant to note that a recently published study of technology assessment in Canada suggests the creation of a National Health Technology Assessment Council to coordinate existing technology assessment activities. The functions identified for the Council include identification of new and emerging technologies, selection of important technologies for assessment, funding of evaluation, providing a point of contact with international technology assessment efforts, and acting as a clearing house. The Council would help to counter the fragmented nature of technology assessment in Canada, and as part of its work would sponsor consensus conferences. Stoddart and Feeny (1986) note that in addition to the Council, changes are needed in the Canadian health service to create

incentives and systems for the appropriate use of technology. We return to this theme in Chapter 6.

Conclusion

The experience of the Netherlands, Sweden and the United States points up some important similarities and differences between the three countries in the approach adopted to technology assessment. These can be summarised as follows:

(1) all three countries make use of consensus conferences, although the method differs somewhat in different countries;

(2) the Swedish MRC and the NIH in the United States are actively involved in supporting clinical trials;

(3) the Netherlands and the United States seek to control the use of technologies by identifying particular technologies whose use is either limited or not covered by health insurance plans;

(4) in all three countries a need has been identified for new agencies to provide leadership and coordination of technology assessment activities;

(5) a distinctive feature of the Netherlands is work on future scenarios in health care, designed in part to serve as an early warning system for policy makers;

(6) a distinctive feature of the United States is the use of the OTA to provide support on technology assessment to the legislature;

(7) also in the United States, some of the specialist associations have taken a particular interest in developing guidelines, standards and advice on specific technologies and procedures;

REFERENCES

Bunker J P and Fowles J (1982) 'Model for an Institute for Health Care Evaluation' in OTA (1982) op. cit.

Caspaire A F et al. (1987) 'Health-care providers resovle clinical controversies: the Dutch Consensus Approach' Australian Clinical Review March, pp43-47

Calltorp J (1987) 'Consensus conference in Sweden: effects and impact on health care providers'. Paper prepared at IHF Congress, Helsinki

Drummond M (1987) 'Economic Evaluation and the Rational Diffusion and Use of Health Technology' Health Policy, Vol 7, pp.309-324

Foote S B (1986) 'Assessing Medical Technology Assessment: Past, Present and Future' Paper submitted to Milbank Memorial Fund Quarterly/Health and Society

Groot L (1987a) 'An Overview of the Systems and Regualtions Concerning Medical Technology and the Diffusion of Six Technologies' in Stocking (1987) op. cit

Groot L (1987b) 'Medical Technology in the Netherlands' in Stocking (1987) op. cit

Herdman R C asnd Behney C J (1985) 'Health and Life Sciences at the Office of Technology Assessment US Congress' International Journal of Technology Assessment in Health Care Vol 1(1) pp160-168

Jacoby I (1985) 'The Consensus Development Program of the National Institutes of Health' International Journal of Technology Assessment in Health Care Vol 1 No2 pp420-432

IOM (1983) A Consortium for Assessing Medical Technology National Academy Press, Washington DC

IOM (1985) Assessing Medical Technologies National Academy Press, Washington DC

Ministry of Health and Social Affairs (1986) The Assessment of Medical Technology - A Proposal for National Coordination Task Force report

OTA (1978) Assessing the Efficacy and Safety of Medical Technologies US Government Printing Office, Washington DC

OTA (1982) Strategies for Medical Technology Assessment US Government Printing Office, Washington DC

SPRI (1985) Swedish Medical Care Programmes Stockholm

Steering Committee on Future Health Scenarios (1987) Anticipating and Assessing Health Care Technology, Martinus Nijhoff, Dordrecht

REFERENCES

Bunker J P and Powles J (1982)
Evaluation' in OTA (1982)

Caspaire A P et al
controversies: the
March, pp43-47

Califorp J (1982)
health care providers

Drummond M (1982)
Health Technology

Foote S B (1980)
Future' Paper

Groot L (1982)
Medical Technology
op. cit

Groot L (1982)
op. cit

Herdman R C
of Technology
Assessment

Jacoby
of Health
Nos pp620

ION (1982)
Press, Washington

ION (1980)
Washington DC

Ministry of Health
Technology - A

OTA (1978) Assessment
Government Printing

OTA (1982) Strategies
Office, Washington DC

SBU (1985) Swedish

Steering Committee on
Assessing Health Care

Stocking B (1986) 'Strategies for Technology Assessment and Implementation in Some European Countries' International Journal of Techmnology Assessment in Health Care Vol 2 No1 pp19-26

Stocking B (1987) Book in press on medical technology in the EEC

Stoddart G and Feeny D (1986) 'Policy Options for Health Care Technology' in Feeny D et al eds. Health Care Technology: Effectiveness, Efficiency and Public Policy, Institute for Research on Public Policy, Montreal

Vang J (1986) 'The Consensus Development Conference and the European Experience' International Journal of Technology Assessment in Health Care, Vol 2(1), pp. 65-76

White L J and Ball J R (1985) 'The Clinical Efficacy Assessment Project of the American College of Physicians' International Journal of Technology Assessment in Health Care Vol 1(1) pp169-174

CHAPTER 5

METHODS OF ASSESSMENT

In earlier chapters we referred to the range of different methods of technology assessment that have been used. These include randomised controlled trials, economic appraisal, consensus conferences and meta analysis. In this chapter we examine the strengths and weaknesses of these methods in more detail.

Background

Whilst no-one any longer doubts the need to assess both old and new medical technologies as critically as is practical there are differing views about how this is best done. It has become clear that no one method is suitable for all types of technology. Moreover it is realised that assessing a technology embraces a much more complex set of questions than arise when a drug is being tested. It is widely accepted that for the assessment of a drug the gold standard is a double-blind randomised controlled trial (RCT). However, technologies present problems that do not arise with drugs, in particular the time taken to acquire skills in the use of a new technology. Consequently there are increasing doubts about the practicality of randomised controlled trials in some circumstances and a feature of the literature on technology assessment in the last five years has been a steadily increasing willingness to consider alternatives to the randomised trial. Yet these are proposed with due deference to the near sanctity of the RCT, aware that to some it is almost heretical to admit the possibility of other approaches to evaluation.

officers (see chapter 3). These concern basic safety and technical performance and can be considered as the assessment of feasibility. This says nothing about the clinical value of a technology or about the indications for its use. Many technologies, however, evolve step-wise using existing technologies and staff (eg. surgical procedures and intensive care regimens); there is then little opportunity to impose any requirement for formal assessment even at this first level.

The next step in assessment comprises the establishment of efficacy. This is performance under ideal conditions on selected patients. If a technology will not work under these favourable circumstances, it certainly cannot be expected to do so elsewhere. Trials of efficacy are commonly done in academic departments and teaching hospitals where, however, the unusually high standard of alternative (existing) methods of treatment may make testing the benefits of a new technique more rather than less rigorous. For example, if surgery is being tested against best medical treatment, that medical treatment may be of a higher standard than could be expected in the country in general.

If a technology does pass this test of efficacy under ideal conditions, guided by the innovator or product champion, it may still prove insufficiently robust when routinely used in a less selected population - the test of effectiveness. This will depend upon how readily learnt are not only the skills associated with the technical procedure but also those of selecting suitable patients. The frequency of such appropriate patients in the population will also be a factor that determines whether or not there is a case for widespread adoption of a technique. Such patients may be so infrequent as to require a policy of centralising the technology on a regional basis, both for economy of scale and

in order to ensure a degree of expertise that can only be achieved by a reasonable throughput of patients.

When formal trials of therapies are carried out it is not uncommon to discover that the patients treated by conventional means (or perhaps not treated at all) have a more favourable outcome than had hitherto been realised. This may be because the true natural history emerges only when the rigorous data collection and follow-up required by a formal technology assessment are imposed. A striking example was the trial of EC/IC bypass surgery, undertaken to reduce the risk of stroke in patients with evidence of atheroma of cerebral supply vessels. Surgery was offered on the assumption that there was a high risk of a stroke and deaths precipitated by surgery exceeded the risk of natural strokes and deaths over the next few years (EC/IC bypass study group, 1985). Sometimes the explanation for the "better than expected" outcome of the control population in a trial is that such patients are in fact being better looked after in various ways other than the modality of therapy under trial. There is no doubt that inclusion of patients in the control arm of a trial does result in their becoming the focus of additional attention. Whether this confers benefit by way of a placebo effect or as a result of extra care varies according to the condition and kind of treatment involved.

It is a common misperception to expect that the outcome of technology assessment, even at this technical level, will be to discover whether a technology is either effective or is not. It does occasionally turn out that a technology is of no value, but no technology can ever be declared to be always efficacious, let alone effective. For no technology is of value when used on the "wrong" patients. An essential component of assessment therefore

In order to ensure a degree of

reasonable thoroughput or

When formal trials of

that the patients

all have a more

be because the

collection and

imposed. A

to reduce the

supply vessels

risk of a

natural

(CBS)

the current

better

trial.

trial does

Whether

extra

It is a common

assessment,

technology

a technology

always

used on the "wrong"

is to discover the sliding scale of benefits and burdens for its use for different types of patient, whose conditions are of varying severity and pattern, and who are of different ages (Jennett, 1986).

Economic appraisal becomes important once effectiveness has been shown - for effectiveness in average conditions is a necessary but not sufficient condition for the adoption of a certain technology. If a technology can be shown to alter significantly the pattern of disease or disability in a community, the next step is to discover how it can be delivered most efficiently. The case has then to be made that society would be well served by diverting resources to this particular technology, because of its benefit:cost ratio; finally it is necessary to make an estimate of the affordable level of provision that appears to be appropriate.

Cost-benefit analysis provides the information necessary to achieve or increase efficiency. The assumption of cost-benefit analysis as applied to health care is that resources should be transferred to those activities with high marginal benefit:cost ratios and away from those with low ratios. In short, it is a means of allocating scarce resources in the most efficient manner.

Economic appraisal should not be deferred too long because only this approach will challenge clinicians to define clearly their objectives and options, and will require them to identify costs and benefits. These should include those that are indirect and perhaps not met by the most obvious agencies. For example, costs may be borne by the family and by the community as well as by the hospital sector that is delivering the technology. Benefits are much

more difficult to assess than costs but again economists will commonly insist on a broader view of benefits than is likely to occur to a clinician. A well conducted economic appraisal will also make allowances for the uncertainty of benefits and it will put these in an appropriate time frame. It will recognise that immediate risks and benefits are perceived differently than those that will be deferred, sometimes by several years; such deferred benefits will be discounted at an agreed rate.

Comparatively few technologies are subjected to thorough economic appraisal. Such a valuation is itself a consumer of resources and the exercise may not be justified unless large quantities of resources are expected to be involved, either because a technology is an expensive one, or else because it is likely to be adopted on a very wide scale should it prove to be effective. Economic appraisal will also be important when there are complex aspects of cost, such as a shift in care from the hospital to the community, or when a new diagnostic test is likely to have consequential indirect costs - for example, by leading to a demand for further follow-up tests or even a new constituency of patients requiring therapeutic intervention. Another consideration is the increased future cost in hospital or in the community of prolonging the survival of disabled patients. On the other hand the benefits of an immediately successful outcome may have to be discounted because of the expected rate of recurrence that will lead to a further episode of costly treatment.

It should be noted that there are a number of problems in applying cost-benefit analysis to health care. One is the difficulty of isolating the costs associated with the use of a technology, given the rudimentary nature

of NHS information systems. Even more problematic is how to measure benefits. It is usual for cost-benefit analysis to enumerate and value benefits in monetary terms. In the case of health care, although there have been valuations of human life in monetary terms, the principal measure of benefit used in the UK is the quality adjusted life year (QALY).

The QALY seeks to summarise in one measure the benefits of medical interventions in terms of the number of years of life saved and the quality of these years. Quality assessments are based on the judgements of a sample of individuals about the relative severity of different states of illness. These are then combined with life years saved to obtain a QALY. Cost data can be added and the costs per QALY of different treatment compared. Economists refer to this as cost utility analysis (Drummond, 1987).

One application of the QALY approach is to construct a league table of the costs per QALY of different treatments. When this is done it can be shown that hospital dialysis is much less cost beneficial than renal or even heart transplantation, while hip replacements or pacemaker implants are even 'better bargains' (Jennett, 1986). Another application is to compare the costs per QALY gained by patients with differing degrees of severity of the same disease. In the case of coronary artery bypass grafting for patients with angina, Williams (1985) showed that there is much greater benefit per unit of cost operating on more severe cases, whose outlook without surgery would be much poorer. The opposite is the case with neonatal intensive care, when the more severe cases (lower birth weight babies) have a less good chance of survival and a greater chance of handicap if they do survive. As a result the overall cost of securing one QALY is some seven times greater for the more

severe than the less severely affected infants (Bennett et al, 1985). This is because the cost of unsuccessfully treating the non-survivors and the badly disabled survivors has to be counted in the price of success for the good results.

QALYs can thus be used both as an aid to the decision-making of doctors dealing with patients with the same condition but of varying severity, and as a tool for policy makers in making investment decisions. However, by no means everyone welcomes this approach to assessment in the value of health care (Smith, 1987). Even the proponents of QALYs agree that more work is needed in order to refine the methods (see for example Maynard, 1987), in particular more studies of patients with specific conditions in order to gain better measurements of quality of life and of health status. But some question the ethical aspect of any person assessing the quality of another person's life, and particularly putting some value on it. Another concern is that because expected duration of life is a factor such measures may count against older patients. Yet technological interventions can bring substantial benefits to well selected elderly patients (Jennett, 1987).

For an assessment to be considered comprehensive it should take account of the acceptability of a technology to patients, and also of any societal and ethical repercussions that widespread adoption might entail. The acceptability of a technology to patients has two components, most clearly illustrated by reference to surgery. There are always the immediate discomfort and risks associated with the intervention, and there may be lasting effects such as disfigurement or dysfunction. Giving due weight to

quality of life when assessing and valuing outcome should allow for the latter.

Societal and ethical repercussions are never easy to anticipate. They are perhaps most obvious in the possibilities that technologies present for dealing with problems at the two ends of life. There are various aspects of the treatment of infertility and also the possibility of detecting a wide range of deformities and diseases in the foetus. This immediately raises the question of what action is justified when such handicaps can be reliably anticipated. Is abortion then to be welcomed into the armamentarium of preventive medicine? Then there is intensive care and surgery for abnormal or very small birth weight babies. Issues involved in using technologies to extend the lives of hopelessly ill adults include the inhumanity of prolonging such lives (or the process of dying), as well as the waste of resources involve in futile treatment. It is often a false antithesis to regard economics and ethics as conflicting; the problem is to persuade professionals and the public to resist the imperatives to inappropriate action when faced with hopeless situations.

However, such issues do not bear immediately on the assessment of the technical capability of technologies. Indeed it is only when they have been shown to be effective that ethical or societal problems may be encountered. There are also examples of technologies that have been created initially for one specific use, but that have later been adapted to a much broader spectrum of uses than had originally been envisaged. It is difficult to see what can be done to anticipate such developments, but when some really innovative technology emerges there may be a place for initiating a "futures" type of

discussion. This might focus on a number of alternative scenarios, based not so much on the results of formal assessment but rather on what might be the implications in terms both of financial cost and practical consequences if such a technology were to prove successful and therefore widely demanded.

Assessing Outcome

Many of these issues apply most obviously to therapeutic technologies, because for them there is no doubt that the measure of effectiveness is an improvement in patient outcome. Diagnostic technologies, which represent a considerable proportion of the innovations of the last decade, are more difficult to evaluate and different answers may emerge according to whether they are used on patients presenting for medical care or in a screening mode; and in the latter case whether used antenatally or later. Assessment of such technologies is initially concerned to demonstrate efficacy at the level of producing a reliable diagnostic result, taking account of false positives and false negatives, specificity and sensitivity.

There are those who consider that assessment of such technologies can legitimately go no further than that, considering it unrealistic to expect an impact on patient outcome. However, diagnosis is not a self-evident good in itself. Technologies that identify conditions that cannot be modified by any known treatment, or that may indeed by asymptomatic, may bring no benefit. Indeed the main consequence of knowledge of the presence of such abnormalities can be to provide the patient and doctor with something to worry about. In that event diagnosis will have done more harm than good. When diagnostic technologies are able to detect treatable disease, their cost-effectiveness will depend on the frequency of the condition in the population subjected to

will depend on the [illegible]

[illegible] are [illegible]

that event [illegible]

can be to provide [illegible]

Indeed the main [illegible]

Beyond [illegible]

[illegible]. Technology [illegible]

Apart on [illegible]

Legislatively or [illegible]

There are [illegible]

[illegible] negotiations [illegible]

testing, and on how cost-effective available therapies are. If repeated tests are needed before a screening method will detect a large proportion of true positives the cost may escalate considerably, even though the unit cost of the test is small (eg. the sixth stool occult blood test).

The remainder of this review of methods of assessment will deal with therapeutic technologies. As patient outcome is the variable endpoint, considerable care and effort are needed to ensure that this is adequately assessed. The method of assessment will vary according to the kinds of outcome that are under discussion. If death in the short term is the endpoint, as in certain interventions for immediately life-threatening situations, there is unlikely to be ambiguity about it. For that reason neither a placebo nor blinding of observers will likely be appropriate because it is difficult to see how anyone could be influenced in counting the dead. When, however, the endpoint is either delayed death due to a new episode of illness, or counting the occurrence of non-fatal episodes, then it may be important to have observers other than those who were involved in the treatment to assess whether or not death or an episode that was survived should in fact be attributed to the disease process under consideration. When the outcome is improvement in symptoms or in quality of life it is also wise whenever possible to have assessments made by observers not involved in the technological treatment. Whenever feasible they should not know whether or not a patient has had one form of treatment or another but this may be difficult when the intervention has been as obvious as major surgery.

Another important variable is <u>when</u> outcome is best assessed. This should not be too soon after the intervention because most technological interventions

are associated with considerable placebo effect, resulting in temporary improvement which is not sustained. On the other hand it may take time to get over the immediate effects of intervention and acute illness, and only then is it possible to judge the beneficial effect of that intervention. In the case of chronically recurring conditions sufficiently long follow-up may be needed to estimate that the frequency of episodes of illness since the intervention has been reduced.

Randomised Controlled Trials

This elaborate and expensive method has come to be highly regarded because of several striking examples of treatments considered to be effective but then shown by RCT to be valueless. That, however, tells us more about the inadequacy of the alternative methods then used than about the unique validity of RCT for evaluating a technology. What so frequently proved deceptive were either historical controls or an uncontrolled contemporary series of cases. There is good evidence that with the passage of time other factors may change as well as the variable under examination - other methods of treatment, types of patients presenting, features of the treating team, inter alia. When contemporary controls are used the compounding factors are different.

The problem then is usually that control and treated groups are not adequately matched, either because all the variables that influence the outcome are not known (they rarely are), or because doctors involved in choosing patients for a new treatment may be unconsciously biased in selecting those likely to do better. The object of randomisation is to minimise the possibility of such bias when choosing patients to be treated with the technology under assessment

or for the control group. What is needed is two groups of patients whose prognosis and outcome would be expected to be similar if they were treated similarly.

That a trial is randomised is no guarantee that it has been well designed or well conducted. The protocol may have been inadequate; but even if it was properly constructed it may not have been strictly adhered to. Many trials are flawed by the exclusion of large numbers of cases before randomisation, or by cross-overs to the alternative method of treatment after randomisation. In other instances small numbers frustrate valid statistics. Whatever the reasons, the outcome of many RCTs is equivocal. Most effective treatments have been adopted without the benefit of controlled trials and many useless ones have been abandoned without this method having been used. Indeed a review of 30 years of papers in JAMA, New England Journal and Lancet found that only 55% were longitudinal trials; only a third of these were truly trials, of which half had no control, and only 5% were RCT (Fletcher and Fletcher, 1979).

Increasing reservations about the appropriateness of RCTs for assessing technologies (as distinct from drugs) derive from several features. Randomisation means that many cases must be accumulated before results are statistically significant, particularly if the events being observed are not very frequent ones. This is turn means that studies must often be multi-centre with all the difficulties and expense that that involves. They are time-consuming, in that many years may pass before the numbers are sufficient, during which period a technology itself may change as experience grows. The cost of such trials is large and there is usually no equivalent

of a pharmaceutical company to share that cost with academic or research funding agencies. Ethical issues sometimes concern clinicians when dealing with life-threatening conditions, especially when decisions about interventions have to be made in emergency situations. However, an argument can be followed that if there is genuine uncertainty about the relative value of two methods of treatment, then there should be no hesitation about randomising. Moreover there is plenty of evidence that many patients are willing to enter randomised trials after adequate explanation (although some critics have doubted whether for some trials that explanation has been as full and honest as it should have been).

More significant criticisms are that an RCT tests efficacy rather than effectiveness, and that its results are seldom widely generalisable. Not only may conclusions be regarded as applying only to the subset of patients studied, but as the study has usually been carried out in an academic centre the circumstances of the trial may be regarded as atypical. A cynic has observed that if the benefit claimed for a therapy is so marginal that it needs a randomised trial to show it then it cannot be a very important effect. In fact trials are much more often of significance when they show that a treatment is not of value - but when a negative trial is believed to have led to the abandoning of a therapy there is often evidence that it was already falling into disuse before the declaration of the trial results.

The trial of EC/IC surgery to prevent stroke provides a striking example of the difficulties associated with an RCT that claims to demonstrate that a widely used procedure brings no significant benefit. This trial randomised 1377 patients from 71 centres in several countries. It cost $9m but was

estimated to have saved some $11m by reason of randomising several hundred patients to medical rather than surgical treatment during the eight years of the study. The trial was hailed as a model of its kind and the surgeons praised for their courage in co-operating (Plum, 1985). Savings in the US of some $30m a year were forecast if the procedure were to be abandoned - a sequel expected if reimbursement for the operation were to be discontinued.

Eighteen months later, however, the same journal published two post-mortem reports on this trial - one from a single surgeon (Sundt, 1987) and one from a committee set up by the American Association of Neurological Surgeons (Goldring et al, 1987). Each revealed that some surgical collaborators in the study had withheld large numbers of medically eligible patients in order to operate on them, and that the trial organisers had not been informed of this and that no outcome data were available for these patients. Whether this is regarded as evidence that the surgeons cheated or that the trial organisers were insufficiently vigilant in detecting them matters less than whether this revelation means that the trial is flawed, as the surgeons claim. The trial organisers have defended their position, maintaining that the results still stand (Barnett et al, 1987). Meanwhile editorialists on both sides of the Atlantic have taken the opportunity to reflect yet again on the difficulties that arise when attempts are made to assess surgical procedures by randomised trials (Relman, 1987; Dudley, 1987). This episode seems set to become a landmark in the history of the randomised trial, whatever the final resolution.

Strategies have been proposed to improve the efficiency of RCTs, by reducing the numbers required - and therefore both the cost and the time taken. The

first is to limit entry to cases at high risk of an untoward event (death or complication), or whose type of illness is likely to be responsive to the intervention under scrutiny (Sackett, 1980). This information may be available from previous studies in the form of prognostic factors and pre-randomisation stratification, which could be regarded as triage for trials. Whatever the reasons, it is of interest that the Institute of Medicine review of methods of assessing technology considered that the main thrust now should be in finding valid alternatives to the randomised trial (Institute of Medicine, 1985). A subsequent issue of the International Journal of Technology Assessment in Health Care that was devoted to methods of assessment had only four pages out of 134 on the randomised trial, and those dealt with the shortcomings of this technique (Danielsson et al, 1986). Clearly there will continue to be some aspects of assessment that can only be resolved by such a trial. But its place should probably be as a court of last resort for settling matters that other methods have identified as requiring this particular technique.

Meta Analysis

This type of assessment attempts to analyse and combine the results of previously reported randomised controlled trials. The aim is to increase the statistical power and to resolve uncertainty when numbers have been small or reports have disagreed; it may also be possible to answer questions not posed at the start of individual trials. A report of 86 trials in the English language literature in the 20 years up to 1986 found most meta analyses to be very incomplete (Sacks et al, 1987). It concluded that there was an urgent need for improved methods of literature searching, quality of evaluation of trials and synthesising of results. Nonetheless they believed that

be obvious that such comparisons were poorly founded, whether the two series had been consecutive or contemporaneous.

The problem with all studies of this kind is that they depend on a retrospective review of data that were available in case records, or in the day books of operating theatres, radiological departments or laboratories. Data that have been routinely collected, without a view to a particular study, are always incomplete and inconsistent in one way or another and this greatly limits the value of many such comparisons. However, some units collect more systematic data than others and this is becoming more common as formal audit is introduced. But the main lesson learnt by the researcher in this kind of study is the inadequacy of routine medical records and the limited conclusions that can be reached from analysis of them.

However, it would be foolish to discount such reports as worthless. Not only may they give indications of changing trends in the utilisation of technologies and of various other aspects of process, they may sometimes suggest that there is a phenomenon worth investigating more carefully. In this regard they are similar to the observations of a scientist, whose curiosity is thereby fired to formulate a hypothesis. If outcome for a series of patients with a certain condition is apparently improved, then one hypothesis might be that some change in the method of treatment is responsible. Alternative explanations are that the patient population has changed, so that fewer are severely affected or that more of them are younger. But the possibility that there are other differences that are unsuspected, either in the patient mix or in the management, may justify a randomised

important to do this.

The use of audit to explore variations in

revealed differences related to aspects of

wide-held difference in complications rate

study of over 80 British surgeons operated

differences in patterns or their

control of the surgeon (leading

to be the most important variable

by the way in which he used

disposal. By contrast the

teaching and district

variables, there being more

hospital, where also more

1971).

effective in prolonging the lives of patients with breast cancer, for example, first emerged from scrutiny of large scale registers.

Routinely collected national statistics, such as the hospital in-patient enquiry and hospital activity analysis, are even less trustworthy and informative. Simple questions about process can be answered - about numbers admitted, length of stay, age, sex and main diagnosis and hospital mortality. One type of data that has emerged from review of routine statistics is the variation in utilisation of various procedures, in particular of surgical operations. As we noted in chapter two, if it can be shown that in communities with widely varying utilisation rates the mortality and morbidity from the conditions for which this procedure are done are similar, there is at least a prime facie case for considering that the procedure may be less effective in influencing outcome than is believed by those who use it more frequently. Such data are, however, crude and do no more than pose a question that more detailed enquiry is needed to resolve.

Prospectively Accumulated Data Bases

The coming of computer technology has transformed the scene of data collection, storage and analysis. It is now practical to accumulate prospectively large numbers of cases according to strict protocols designed to answer specific questions and that can include many details. These can relate to severity of illness, age and other factors likely to determine prognosis; also about therapies used, and about outcomes - with strenuous efforts to ensure follow-up. Rigorous data collection of this kind is commonly claimed as characteristic of RCTs. Although such data have emerged as a by-product of RCTs the technique per se has nothing to do with randomisation or with

quantitative synthesis of this kind could be valuable. A more optimistic note was struck by a commentary based on the use of this overview method to uncover the benefits of adjuvent radiotherapy and chemotherapy for breast cancer (Lancet, 1987). In this case the value of a small benefit for a common disease became evident by synthesis when it had not been apparent in small individual trials. And as we noted in chapter 3, the National Perinatal Epidemiology Unit has given a central place to meta analysis in its work.

Reports of Clinical Experience

These make up the bulk of the papers that are the reading of those clinicians who do watch the journals. As they are usually more informally and more attractively written than are some statistically complex reports their influence should not be underestimated. Students of technology assessment may sub-classify them as consecutive series, sample surveys, case studies and historical control studies, according to how they have been conducted. At their simplest a series of patients treated with X is described, the only comparison being with previous reports in the literature. The focus of such a report may be the use of a new technology, or of an old one in a novel way; or it may be to report its use in one type of patient such as the elderly; or in a particular setting, such as one geographical location or type of hospital.

More ambitious reports may compare one hospital with another during the same time period, or in two different time periods within the same institution. In the latter case, as also when a comparison is made with published literature, the study is said to depend on historical controls. It is such comparisons that have gained a bad reputation as the basis for false claims for various technologies, which RCTs have frequently refuted. Even without an RCT it may

[redacted]

controlled trial - if the matter under consideration is of sufficient important to do this.

The use of audit to explore variations in process and outcome in surgery has revealed differences related to aspects of technological input. Thus the six-fold difference in complication rate after anastomosis of the bowel in a study of over 80 British surgeons appeared to be related not so much to differences in patients or their disease as in aspects of management under the control of the surgeon (Fielding et al, 1980). Although the surgeon was found to be the most important variable, it was thought that he influenced matters by the way in which he used various components of the technologies at his disposal. By contrast the difference in death rates from prostatectomy in teaching and district hospitals proved to be largely due to patient variables, there being more older and complicated cases in a district hospital, where also more operations were done as emergencies (Ashley et al, 1971).

Registers and Routine Statistics

For many years registers have been kept of patients with certain diseases or those receiving certain treatments. The most common are those for cancer and for radiotherapy. Such limited information has been available from these sources that little has been learnt other than trends in incidence and death rates for different cancers and different modes of treatment. The most that is likely to emerge for technology assessment is the occasional observation that there are changes in death rates that might be related to various treatment regimens. The suspicion that radical mastectomy might not be

trials. Whatever their source, data bases of this kind provide biostatisticians with the possibility of producing prognostic indices that estimate the probability of various outcomes, given certain clinical features. The main source of bias when assessing alternative therapies is imbalance in the prognostic factors in the groups of patients being compared.

Randomisation is only one way to minimise such bias - its simplicity and wide acceptance making it the most obvious evidence that a trial has been based on a valid comparison. However, randomisation may result in an uneven distribution between groups of patients of factors that influence outcome unless the numbers are very large, and it can be prohibitively expensive to achieve such a balance by chance.

Two large international studies have shown the power of the data base approach to the assessment of complex technologies such as those involved in intensive care. The Glasgow based study of severe head injuries began in 1968 and now has some 3,000 cases in file from Glasgow, two centres in the Netherland and two in California. Large variations in the frequency of use of certain therapies for patients with similar prognosis was associated with no significant difference in outcome (Jennett et al, 1979; Jennett, 1984). The same has been shown for patients in general intensive care units in several parts of the United States (Knaus et al, 1982a), and in a comparison between one United States centre and a number of French units (Knaus et al, 1982b). That the outcome should be similar in spite of wide differences in treatment suggests that the individual components of the treatment packages were not crucial influences on outcome.

A more detailed study can be made of an individual technology, predicting what the death rate would be in matched cases treated with that technology and comparing this with those not so treated (Jennett et al, 1980). Whilst the objection can be raised that the allocation of patients to treatment methods was not random, the fact is that large numbers of cases of similar severity were found in this study to have been treated in some places with one technology and in other places without - yet another example of variations in clinical practice. The criticism that unknown prognostic factors were operating is countered by the accuracy of the predictive model, which suggests that no important influences on outcome had been ignored (Murray, 1986). Such a data base with predictive modelling can also be used to stratify cases before randomisation if it is later felt that a randomised trial of some particular subgroups of patients or of treatments is justified.

Another value of data bases of this rigorous kind, prospectively collected and maintained over several years, is that the stability of the patient population and of the outcome of treatment can be observed. Were a new mode of treatment to emerge it should then be possible to test it on such a stable population, without the fear that other matters have been changing over time. A much smaller contemporary control than usual would then be acceptable to verify that there had not been some very recent secular change independent of the new therapeutic modality.

In his essay "Evaluating the physician and his technology", McDermott (1977) claimed that constructing usable data bases for large parts of medicine could have as much impact as a major therapeutic breakthrough. He was in no doubt about the difficulty of doing so and thought it might take 10 years to do so

for one major disease of one system. The IOM study concluded that data bases offered rich possibilities for detailed study but that their cost was very considerable, perhaps as great as for a randomised trial. However, the view of Moses (1986) was that the practical difficulties and discipline required might make the realisation of these expectations much more limited than was hoped. The international head injury study has, however, demonstrated beyond doubt the feasibility of this method (Murray et al, 1986).

Expert Systems and Decision Analysis

Judgements about the diagnosis, the prognosis and the optimum course of action for an individual patient depend on the informal and intuitive weighing of evidence by the clinician, who compares the features of the patient in his mind with his previous experience and with knowledge acquired by various means from other practitioners. The relationships between these many variables are complex and the computer age has seen the evolution of formal systems to provide calculated probabilities for alternative diagnoses, given certain constellations of clinical features and diagnostic tests. Similarly for prognosis, as already referred to in relation to data bases on head injuries and intensive care. For making decisions about management such models can take account of the probability of certain specified outcomes, given particular interventions; moreover these outcomes can be ascribed explicit values or utilities so that the net benefits expected from alternative lines of action can be compared (O'Brien, 1986).

This emphasises that the benefit gained from the use of a technology depends on its being employed in a rational way, taking account of the calculated trade-offs between its use under different circumstances. This may be

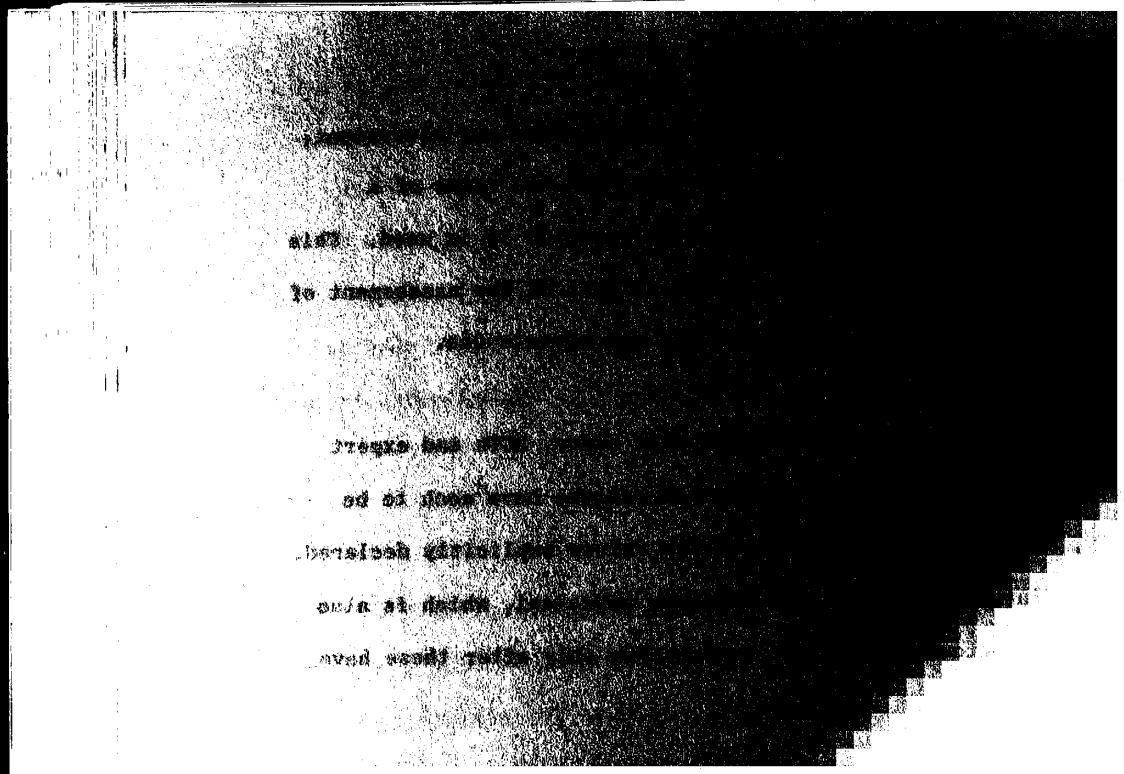
regarded as more to do with using than with producing technology assessment; but it is part of the evaluation process to discover how the value of a technology may vary according to the circumstances in which it is used. This emphasises a point already made - that an essential part of the assessment of a technology is to define the indications for its appropriate use.

An advantage of all these approaches, detailed data bases, RCTs and expert systems, is that disease state, health status and outcome have each to be defined precisely, and assumptions about relative values explicitly declared. There is clearly an interaction here with economic appraisal, which is also concerned with valuing outcomes and which is possible only after these have been defined.

Group Judgements

Tempting as it is to employ artificial intelligence, as described in the above section, to solve some dilemmas in medicine, it is worth considering how to tap the accumulated human intelligence of experienced clinicians. Given the evidence from the variations' studies that doctors' practice is so variable, however, the question is how best to gain access to their experience and views by engaging them in a dialogue that might lead to a consensus statement, or at least to a clearer explanation of why their opinion differ so much (Fink et al, 1984).

A Delphi technique has been used to predict future developments in medical technology. Experts are invited to return by mail questions about the probability of certain events or developments. Their own ranking together with the median of the total group and the range, are then fed back to



individuals who are invited to modify their original estimates. The nominal group technique (NGT) similarly involves a list of ideas or statements which have to be ranked or commented on, but the participants meet for limited discussion. They first generate their responses silently, these are collected and distributed without identification of the author. The responses are then clarified without confrontation but in verbal exchanges, and further round robins circulated. Some studies have been carried out in America of both NGT and Delphi for a variety of organisational medical matters such as strategies against drug abuse and handling emergency medical cases. When surveyed six months later the NGT participants were found to have changed their opinions to a significantly greater extent than had the Delphi group.

The best established method of securing a group judgement in medicine is the NIH Consensus Development Conferences, which have been running in the United State since 1977. These have already been described in chapter 4 and in chapter 3 where the more recent adoption of such conferences in the UK, Sweden and Netherlands was discussed. A considerable literature of comment is emerging about these conference in the United States, including formal attempts to assess how widely disseminated the data have been, and what impact they may have had on practice. It is widely agreed, however, that the key to success lies in a topical subject and a good panel. The possibility of using decision analysis as a synthetic tool for achieving a consensus has been explored on five completed conferences (Pauker, 1986). In two conferences the panel's activities were passively observed and then decision models shown to selected speakers. In the next two conferences the models were presented to the panel after their first formal meeting. The suggestion is that a formal model and a consensus panel might address the same questions, perhaps

focusing on different aspects of the problem, and the two processes could thereby complement each other.

Developments of the consensus movement in Europe have taken various forms (Vang, 1986). That which seems to depart most from the American model is in the United Kingdom, where a deliberate attempt is made to embrace economic and ethical issues. It is also the practice in the UK to include on the panel a minimum of experts with at least half the members being non-medical, including on occasions the chairman. One of the first three conferences dealt with a public policy issue (The Place of Asylum in Society). This produced a less clearcut statement than the other conferences and the British Medical Journal chose to publish it only in shortened form. In the Netherlands the opposite trend has been observed, as the conferences are closed and made up mainly of doctors. It seems that the consensus statement there omits any reference to matters on which consensus could not be reached. By contrast it has been a feature of the American and British conferences to point out where consensus cannot be reached and to draw attention to whether this is due to lack of data - as a stimulus to the production of better data.

The Rand Corporation has developed a different method for synthesising expert opinion about the appropriateness of medical and surgical procedures (Park et al, 1986). Twelve topics were divided into cardiac, gastrointestinaL and neurological and three panels each of nine persons were assembled. They were asked to rank appropriateness of a large number of indications (more than 200 for most technologies and more than 1,000 for two of them). These were ranked on a scale of 9 and the scores divided into three groups - 1-3 negative, 4-6 equivocal, and 7-9 positive. Experts were asked to score for a favourable

The Rand Corporation has developed a [illegible]
opinion about the appropriateness of [illegible]
al, 1968). Twelve topics were [illegible]
meteorological and three panels each of [illegible]
asked to rank appropriateness of a [illegible]
for most technologies and more than [illegible]
on a scale of 8 and the scores divided [illegible]
equivocal, and 7-8 positive. Experts were [illegible]

balance between positive features (increased life expectancy, relief of pain and anxiety, and improved function) as compared with negative outcomes (mortality, morbidity, anticipatory anxiety, pain of the procedure and time off work). They were asked to exclude cost considerations. The panels met over two days when they often changed the categories of the indications originally given to them, either splitting or merging categories. The endpoint was regarded as equivocal either if agreement was in the 4-6 range, or if there was marked disagreement among the experts. Indications which raised disagreement among the expert panel were highlighted as areas where further knowledge was required. Where agreement about inappropriate use has been reached it is then possible to explore variations in practice in different places in order to discover whether these reflect inappropriate usage as determined by the panel. It should be possible to combine the work of such an expert group with a consensus audience. Having reached a consensus within an expert group, this matter might then be opened to broader discussion, and experts invited to consolidate or modify their original statements in the light of points raised.

Another approach to discovering how expert clinicians react to specific clinical problems is to distribute to them a group of "paper patients" or clinical vignettes. Each member of the circulated group is asked to answer a number of specific questions - which may be ranking the cases in order of severity for prognosis, or to indicate whether they would recommend certain investigations or lines of treatment. It is then possible by analysis of the results to see what degree of consensus there is. But it is also possible to analyse what features or clusters of features appear to correlate with certain decisions - another approach to artificial intelligence, if you will. What

accumulates. The McMaster team have suggested a "technology assessment iterative loop" (TAIL) to describe this process (Feeny et al, 1986). That is why initial assessment methods that take 5 to 10 years to complete, and cannot be modified until then, are so cumbersome. Indeed the technology may have moved on by the time the results are published. It is then easy to discount the findings, even if some part of the study would still be relevant. The dilemmas involved in the timing of assessments have been summarised by Buxton in the following way: "It's always too early to evaluate a technology, until, suddenly, it's too late".

Summary

Analysis of different methods of assessment points to the following conclusions:

1) Technology assessment is concerned with the feasibility, efficacy, effectiveness, efficiency, cost-effectiveness and social acceptability of clinical interventions;
2) Diagnostic technologies are more difficult to evaluate than therapeutic technologies because their relationship with improved patient outcome is less direct;
3) The RCT is the most elaborate and expensive method of assessment but there are difficulties involved in conducting trials as the controversy surrounding the EC/IC surgical trial demonstrates. Many of those involved in the technology assessment field make use of the other methods of evaluation described in this chapter;
4) These other methods include meta analysis, published reports of

clinical experience, clinical audit (including studies of variations in practice), the use of registers and routine statistics, prospectively accumulated data bases, expert systems and decision analysis, and group judgements, including consensus conferences;

5) A key issue in technology assessment is the timing of assessment. This should be neither too early (otherwise the technology may have developed insufficiently) nor too late (otherwise it may be difficult to convince clinicians of the need for assessment).

REFERENCES

Ashley J S A et al (1971) 'Case fatality of hyperlasia of the prostate in two teaching and three regional-based hospital', Lancet, 2: 1308-11

Barnett H J M et al (1987) 'Are the results of the extracranial-intracranial bypass trial generalizable?' New England Journal of Medicine, 316: 820-24

Bennett K J et al (1985) 'Guidelines for health technology assessment: the efficacy, effectiveness and efficiency of neonatal intensive care', International Journal of Technology Assessment in Health Care, 1: 873-92

Danielsson H et al (eds) (1986) 'Technology assessment: policy, clinical and methodological issues' Interational Journal of Technology Assessment in Health Care, 2: 5-131

Drummond M (1987) 'Cost Benefit Analysis in health Care: Future Directions', in Teeling Smith G (ed.) Health Economics: Prospects for the Future, Croom Helm, London.

Dudley H A F (1987) 'Extracranial-intracranial bypass, one; clinical trials nil', British Medical Journal, 294: 1501-02

EC/IC bypass study group (1985) 'Failure of extracranial/intracranial arterial bypass to reduce the risk of ischemic stroke: results of an international randomised trial', N Engl Journal of Medicine, 313: 1316-18

Feeny D et al (eds.) (1986) Health care technology; effectiveness, efficiency and public policy, Halifax: Institute of Research in Public Health Policy

Fielding L P et al (1980) 'Anastomotic integrity after operations for large-bowel cancer: a multicentre study' British Medical Journal, 2: 411-14

Fink A et al (1984) 'Consensus methods: characteristics and guidelines for use', American Journal of Public Health, 74: 979-83

Fletcher R H and Fletcher S W (1979) 'Clinical research in general medical journals: a 30 year perspective', New England Journal of Medicine, 301: 108-183

Goldring S et al (1987) 'The extracranial-intracranial bypass study', New England Journal of Medicine, 316: 817-20

Healy M J R (1978) 'Truth and consequences in medical research', Lancet, 2: 1300-01

Institute of Medicine (1985) Assessing Medical Technologies, National Academy Press, Washington DC.

Jennett B (1979) 'Prognosis of patients with severe head injury' Neurosurg, 4: 283-89

Jennett B et al (1980) 'Treatment for severe head injury' Journal Neurol Neurosurg Psychiat, 43: 289-95

Jennett B (1984) 'Outcome of intensive therapy for severe head injury: an inter-center comparison' in Major issues in critical care medicine, Parillo J E, Ayres S M (eds), Baltimore: Williams & Wilkins

Jennett B (1986) High Technology Medicine - benefits and burdens, (2nd ed), Oxford University Press, Oxford: new York

Jennett B (1987) 'High technology medicine and the elderly' in Health Problems of an Ageing Population, Wells N, Free C (Eds.), Macmillan Press Ltd., London and Basingstoke

Knaus W A et al (1982a) 'Evaluation of outcome from intensive care: a preliminary multihospital comparison' Crit Care Med, 10: 491-96

Knaus W A et al (1982b) 'A comparison of intensive care in the USA and France' Lancet, 2: 642-46

Lancet (1987) 'Whither meta-analysis?' 1: 887-88

Maynard A (1987) 'Health Economics Research in the UK', in Teeling Smith G (ed.) Health Economics: Prospects for the Future, Croom Helm, London.

McDermott W (1977) 'Evaluating the physician and his technology' Daedalus, 106: 135-57

Moses L E (1986) 'Data acquisition for assessment of medical technology: methods other than technical trials' International Journal of Technology Assessment in Health Care, 2: 27-38

Murray G D et al (1986) 'Assessing the performance and clinical impact of a computerized prognostic system in severe head injury' Stats in Med, 5: 403-10

Murray G D (1986) 'Use of an international data bank to compare outcome following severe head injury in different centres', Stats in Med, 5: 103-112

O'Brien B (1986) 'What are my chances doctor': a review of clinical risks, Office of Health Economics, London

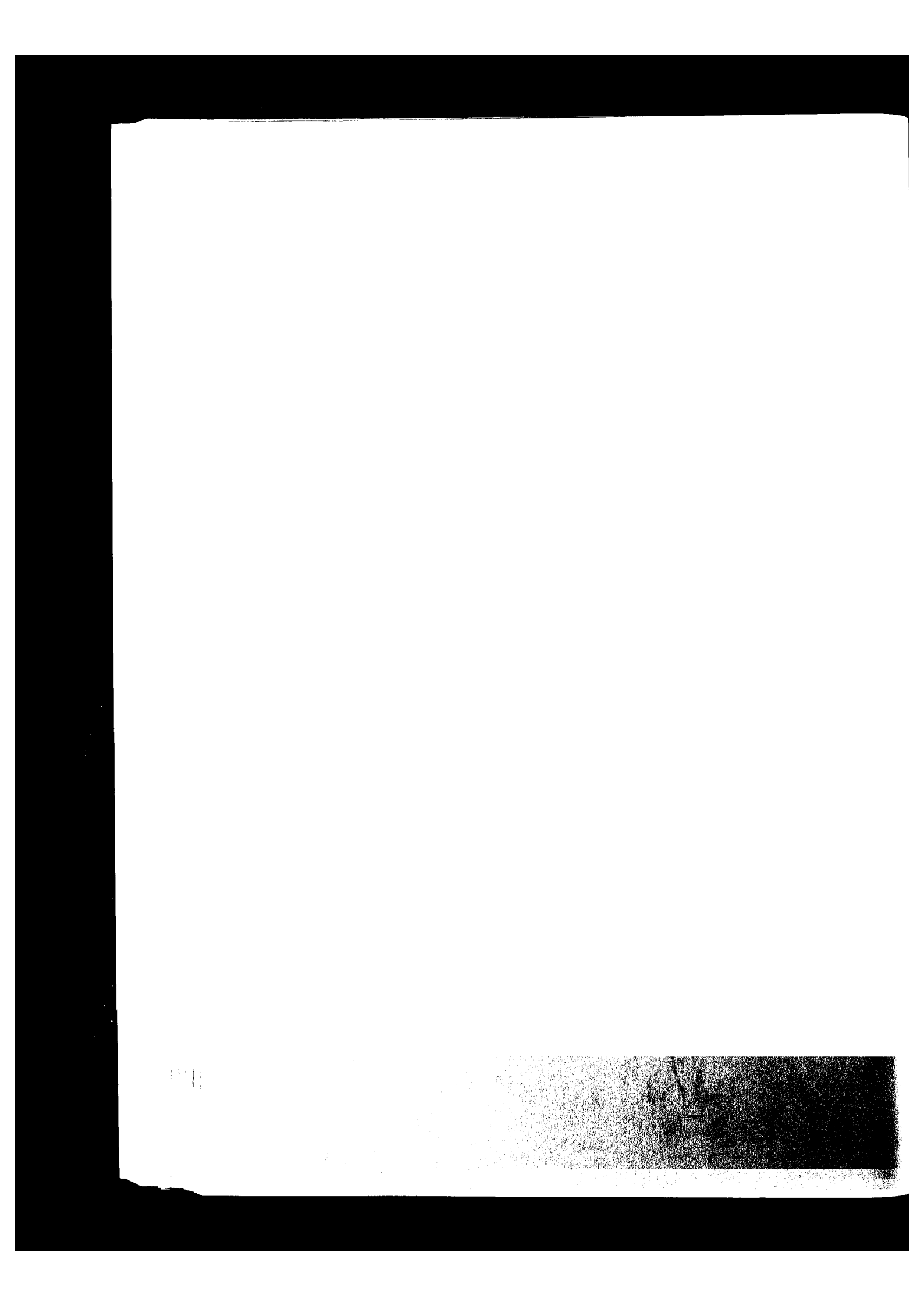
Park R E et al (1986) 'Physician ratings of appropriateness indications for six medical and surgical procedures', American Journal of Public Health, 76: 766-72

Pauker S G (1986) 'Decision analysis as a synthetic tool for achieving consensus in technology assessment', International Journal of Technology Assessment in Health Care, 2: 83-98

Plum F (1985) 'Extracranial-intracranial arterial bypass and cerebral vascular disease' New England Journal of Medicine, 313, 1221-23

Relman A S (1987) 'The extracranial-intracranial arterial bypass study. What have we learned', New England Journal of Medicine, 316: 809-10

Sackett D L (1980) 'The competing objectives of randomised trials' New England Journal of Medicine, 303: 1059-60



Sacks H S et al (1987) 'Meta-analyses of randomized controlled trials', New England Journal of Medicine, 316: 450-55

Smith A (1987) 'Qualms about QALYs', Lancet, 1: 1134-36

Sundt T M (1987) 'Was the international randomized trial of extracranial-intracranial arterial bypass representative of the population at risk?', New England Journal of Medicine, 316: 814-16

Vang J (1986) 'The consensus development conference and the European experience', International Journal of Technology Assessment in Health Care, 2: 65-76

Williams A (1985) 'Economics of coronary artery bypass grafting', British Medical Journal, 291: 326-29

CHAPTER 6

THE INFLUENCE OF TECHNOLOGY ASSESSMENT ON CLINICAL PRACTICE

This chapter examines the influence of technology assessment on clinical practice. One approach to this issue concentrates on how best to ensure that clinicians are informed about data on technology assessment. The other questions whether awareness of information about assessment is the main problem, because of the evidence that few clinicians act consistently in accordance with technology assessment even when they are aware of its conclusions. This second approach focuses less on the provision of data on technology assessment to clinicians than on the use of audit, education and incentives to change clinical practice.

The Customers of Technology Assessment There are a variety of motivations behind technology assessment. If in some cases the prime intention is to maximise benefits by promoting appropriate use and minimising unjustifiable exposure of patients to technologies that could do harm, in others the main thrust is the containment of costs. The latter is particularly important in countries where staff are reimbursed on a fee-for-service basis for providing technology. Where comprehensive health systems are provided an important purpose of technology assessment is to help in planning the acquisition of technologies and the provision of the facilities necessary for them to function optimally: the equivalent use of assessment at the macro level in other systems is in decisions about which technology should be reimbursable.

There are therefore a wide range of customers for technology assessment. These include policy makers, planners and the allocators of resources at national and regional levels. Lower down there are the managers of health care facilities in districts or units. Clinicians should certainly be concerned, for it is the aggregate of their individual decisions that determines the use of technologies. Patients are becoming increasingly interested in technology assessment, in part through concern about the over use of inappropriate technologies, and in part through an interest in making effective technologies more widely available.

One approach would suggest that the first impact of technology assessment should be on policy makers, and that only later will it affect clinicians. However, all the evidence in Britain is that few policies about acute medical emerge without there having been consultation with the leading clinicians in the field. Indeed there is abundant evidence that use of technology depends primarily on the decisions of clinicians. It has to be remembered that policy makers cover such a wide canvas of activities that they can spend little time focusing on any one technology, whereas this may form a lifetime concern for clinicians in a specialty. Moreover the doctrine of clinical freedom, however threatened and threadbare it is, gives clinicians a wide range of discretion in the use of technologies.

Influences on Clinical Practice

It is important to emphasise that technology assessment is only one of several factors that bear both on policy and practice. It may be the most rational of those influences but others may be equally or more powerful. When dealing with seriously ill patients, clinicians are susceptible to imperatives to act

positively rather than with restraint, particularly in the presence of distressing symptoms or the diagnosis of a dread disease such as cancer or heart disease. There is also the natural desire to be innovative, or to make use of equipment and facilities that are available. Then there is the wish to resist criticism from the patient or her family as well as from other doctors and nurses, whose expectations are that the doctor will act. In the UK legal harassment cannot seriously be regarded as a reason for misusing technology although it certainly is a factor in the United States where defensive action distorts the practice of medicine in some specialties.

Quite apart from these imperatives, the commonest reason for inappropriate use of technology (whether overuse or underuse) is ignorance of whether or not it is indicated, or is likely to be useful in these circumstances for this kind of patient. This is the phenomenon of professional uncertainty discussed in Chapter 2. The uncertainties surrounding clinical practice have been identified by Reiser (1986) as related to the biological response of different patients to the same disease; to the definition of diseases; to the performance characteristics of technologies and of the experts who use them; to the unpredictability of the reactions of patients and lawyers to adverse outcomes; to perplexity about societal and ethical dilemmas; and to inadequate information about the uncertainties and expectations of individual patients. Various strategies are used to cope with these uncertainties. One is coherence - doing only what seems to make sense in terms of accepted theories of disease mechanisms, and rejecting theoretically inexplicable outcomes even when they appear favourable. Related to this is orthodoxy, where doctors adopt the teaching of seniors with its seeming certainty, and dismiss data that do not conform with this as being exceptional or unusual.

Technology itself may become a shield against uncertainty, as when laboratory or radiological findings are given precedence over the history and examination because these tests seem more certain and concrete.

In some circumstances, professional uncertainty stems from ignorance in the medical profession as a whole concerning indications for treatment. In other circumstances, evidence is available but is either not known to the individual clinician, or it is known but is ignored. One reason for ignoring known data is yielding to the imperatives already mentioned; another is disbelief on an intellectual level, or a belief that the data do not apply to this patient in these particular circumstances. Whereas we may debate how likely it is that the fully informed clinician will always act rationally in accordance with knowledge about assessment, it is certain that he cannot do so if he is unaware of what data are available or if there are no data. It is therefore logical first to consider how data about assessment are disseminated when a study has been completed.

Dissemination of data

Initial publication will normally be in a refereed medical journal of high repute, often a weekly publication so that no undue delay is entailed in publication. Publication at greater length in specialist journals can entail several months delay, even a year or more. This initial publication, with the necessary technical details concerning methods of data collection and statistical analysis, does not make easy reading for the average clinician. Added to this there may be scepticism that even a multi-centre trial would apply generally; when a series is reported from a single centre, it is even more likely to be suspected of having only local validity.

to by the peers or specialist group of a given clinician. Alternative clinical strategies that involve close relationships with other specialties are apt to be threatening and are therefore less likely to be adopted. Industry may influence adoption by organising meetings where a technology is presented in a favourable light; published procedeedings of this kind of meeting are unlikely to be refereed. For embodied technologies, those that involve a single instrument, industry may also subsidise courses that offer to teach the use of the new technique, but in effect these are designed to promote its adoption.

The influence of the public on the adoption of new technologies or on the increased use of existing ones can be important in highly visible areas where the benefits are seemingly obvious. Examples are screening for cervical or breast cancer, or the use of diagnostic imaging - often, however, based on exaggerated expectations of benefits. Both television programmes and the columns of medical correspondents of daily papers and weekly magazines now form an important avenue of dissemination to the public. With increasing involvement of patients in decisions about their own treatment, this heightened awareness of the public may in some instances become a significant incentive to doctors to pay heed to technology assessment. This has been particularly evident in the UK in the area of obstetrics and ante natal care, where midwives and nurses have become a potent force to reinforce messages about appropriate use. Publicity in Switzerland about hysterectomy rates and in the United States about surgical mortality rates in different hospitals led to questioning of doctors by the public. Awareness of the possibility of less radical surgery for breast cancer on the part of patients and their relatives may prove to be a more powerful influence on surgical practice than all the

academic articles about trials. For this reason, the organisers of the UK conference on breast cancer made a deliberate effort to make sure that the consensus statement was made available to women.

Resistance to Technology Assessment

It is always difficult to relate change or lack of change in medical practice to specific influences. Whilst negative findings may be ignored for years, a positive finding in favour of technology may result in much more wide spread adoption than the assessment in fact justifies. On the other hand a technology might already be in the process of being abandoned and this is only accelerated by a formal technology assessment that reinforces the trend. Findings of technology assessment that go against widespread expectations or trends or traditions, or that threatened vested interests are likely to be resisted. The status of the recommending body that publishes the assessment can also influence how much notice is taken of it.

Gutzwiller and Charzanowski (1986) note that there is a range of evidence that doctors do not behave in accordance with the results of RCTs. They also note that negative assessments may result in the abandonment of a technology or a reduction in its use, but in some cases there may be little impact on clinical practice. An example of the latter is the routine use of certain x-ray examinations, which continue to be performed despite a professional consensus that they are not necessary.

In her work, Greer (in press) has highlighted the scepticism of hospital doctors on both sides of the Atlantic about scientific literature. Reservations were expressed about the motives and methods of research workers

and criticisms made about the impractical nature of the reports and recommendations. These clinicians were doubtful about how generalisable were the findings of academic assessment. Greer also found that many doctors who took part in trials did not feel any responsibility to translate their findings into practice. Non-teaching doctors found meetings more useful than papers, mainly because the product champions were likely there to be challenged about complications, difficulties and poor results. Not only could the personalities of the protagonists be assessed, they frequently answered questions reasonably honestly, even though they would not include such difficulties in their published reports.

In practice, as we noted in Chapter 2 in discussing clinical variations, subjective factors and practice style may be just as important as scientific evidence in explaining clinical behaviour. As Wennberg (1984) has commented, the clinical hypotheses underlying surgical decision making are often weak, implicit and untested, while clinical decision rules are mostly derived from traditional teaching of one or other surgical schools of thought, or are personal and reflect the idiosyncracy of the individual surgeon. In view of this, methods other than the simple provision of information to clinicians are likely to be needed if the use of technology is to be effective.

There is increasing awareness of the importance of taking positive steps to ensure that more notice is taken of technology assessment in the planning of health services and in medical practice. Thus, Blanpain (1986) has argued that there is a need to bridge the communication gap between the scientific community and policy makers, and he suggests the policy-oriented journals and health policy forums have a key role to play. As far as clinicians are

[redacted]

concerned, medical education, audit, guidelines and financial incentives are important mechanisms for translating the findings of medical assessment into medical practice. We now discuss these mechanisms in more detail.

Medical Education

In theory medical education, both undergraduate and post graduate, is a significant influence on the use of technology, and teachers have a key role in emphasising the need for careful scrutiny of technologies and for seeking clear indications for their appropriate use. Certainly the opportunity is there in the teaching hospitals where university clinical departments often include both innovators of technology and those involved in clinical trials. Moreover such teaching units commonly have regular meetings to review their practice, the forerunners of today's audit. Too often, however, the students are excluded from exposure to these activities because many medical school teachers are still overly concerned with diagnosis rather than management and this is reflected both in teaching and in examinations. Moreover far from academic clinicians setting a good example in the parsimonious use of technology, they are often suspected of being largely responsible for the frequency with which technologies are used in circumstances that do not justify it. Partly this may arise from a confusion between the use of technologies as part of a research endeavour and their use in routine practice. It would be helpful if more explicit acknowledgment of this were made. Too often therefore any teaching about evaluation is left to departments of community medicine and it is then all to easy for students to regard this as an activity divorced from everyday clinical medicine. Until students see clinicians setting an example by themselves questioning what they

do by way both of tests and treatment and criticise their juniors for overuse rather than underuse the situation is unlikely to improve.

Earlier this year the International Journal of Technology Assessment in Health Care devoted a whole issue to educating practitioners in the appropriate use of technology and this included references to costs and ethics and rationing. The pressures to include medical ethics teaching has perhaps had greater impact than that in regard either to cost or assessment of technologies and yet the three are inextricably interwoven.

Clinical Audit

Audit is receiving increasing attention in the light of public demands that the profession be more rigorous in its self-policing and the demand of managers that clinicians be better able to account for their activities (Dawson, 1985). Audit relates to structure, process and outcome, hopefully all three. Clinicians are much more willing to audit structure because that relates to their demands or requirements for certain facilities in the way of equipment and staff and does not seriously question how these are used. Audit of process has a tendency to focus on features such as length of stay and other efficiency measures which do not examine whether what is being done is effective or not. Manifestly it is of no benefit to deliver more efficiently measures that are not effective. Clearly the ultimate measure of effectiveness is outcome, yet remarkably little audit is to do with outcome. However, this is now beginning to emerge and just as there is a set standard to indicate what procedures are appropriate in certain circumstances so there is need for some standard over what kinds of outcomes are acceptable - by way of survival, improvement and quality of life.

[illegible] their [illegible]

[illegible] to improve.

[illegible] Department of Technology Assess[illegible]

[illegible] affecting practitioners in the [illegible]

[illegible] conferences to create and [illegible]

[illegible] within teaching [illegible]

[illegible] to cost or [illegible]

[illegible] observed.

[illegible] attention in the [illegible]

[illegible] in its self-[illegible]

[illegible] better able to [illegible]

[illegible] to structure, [illegible]

[illegible] are much more willing to [illegible]

[illegible] requirements for [illegible]

[illegible] not seriously [illegible]

[illegible] focus on [illegible]

[illegible] do not examine [illegible]

[illegible] is of no benefit [illegible]

[illegible] Clearly the [illegible]

[illegible] little [illegible]

[illegible] energy and just as [illegible]

[illegible] appropriate in certain [illegible]

[illegible] some of outcomes are [illegible]

[illegible] of life.

The Lothian Surgical Audit is an example of what is being done in one district. This involves all 31 consultant general surgeons and urologists in the Lothian Health Board. Routinely collected statistics were found to provide an unsatisfactory basis for audit, and instead the surgeons collected their own information making use of computers for the analysis. The results were fed back to the clinicians involved, and specific difficulties arising were discussed at regular meetings arranged for this purpose. Gruer et al (1981) report that as the surgeons themselves are responsible for providing the data they regarded them as reliable. Furthermore, the data demonstrate improvements in the outcome of surgery with significant falls in re-operations and post-operative mortality. Gruer et al conclude that it is likely that audit has contributed to these improvements.

The Lothian example demonstrates that surgical audit on a district basis can be organised given commitment and enthusiasm on the part of the clinicians concerned. Gruer et al note that simplicity was an important factor in the success of the system, and it is also worth recording that a feature of this initiative was collaboration between specialists in community medicine and their surgical colleagues. As others have noted (Mitchell and Fowkes, 1985), passive feedback of information rarely has an impact on clinical behaviour. However, if information feedback of reliable data is combined with educational and review mechanisms, supported by leadership from key clinicians, it may well be effective (Fowkes, 1986). This is particularly the case where there is a prior agreement on the standards of service to be expected.

The interest of the Royal Colleges in audit is becoming increasingly positive. The Royal Colleges of Surgeons of England and of Edinburgh have each held

meetings in 1987 about audit. It seems likely that some of the Colleges may soon require formal audit in hospital units that wish to have posts approved for higher training, as has been the case for some years in the Royal Colleges of Physicians in Australia and Canada. Colleges could further reinforce the importance of audit by making reference to its role in their examinations.

Guidelines

As we noted above, audit is likely to be strengthened if it occurs within the context of agreed standards or guidelines. These guidelines may be drawn from the results of technology assessment, or from the experience of experienced senior clinicians nationally or locally. Audit can be seen as the means of discovering whether or not guidelines are being adhered to, and quality assurance is the process of acting on the results of audit to produce improvements in clinical practice.

There are those who are dubious about guidelines, concerned that they may restrict innovation and may limit contrasting styles of clinical management. In particular there is concern that they may not be sufficiently readily adapted to the individual patient. It is however too easy to find merit in contrasting styles which are regarded as healthy diversification, but it is a thin line between this and eccentric variation. Another concern is that junior staff may be left to decide on their own, left to guess what their seniors might wish but not really knowing about this. Furthermore, there is fear that legal bodies might give undue weight to guidelines and that professionals not conforming to them might find themselves in difficulties. On the other hand guidelines may be helpful in giving some corporate support for difficult decisions. However, it should be clear that anyone may depart

meetings in 1981 about audit. It seems likely that some of the Colleges may soon require formal audit in hospital units that wish to have posts approved for higher training, as has been the case for some years in the Royal [illegible] of Physicians in Australia and Canada. Colleges could further reinforce the importance of audit by making reference to its role in [illegible] examinations [illegible]

__Guidelines__

As we noted above, audit is likely to be strengthened if [illegible] context of agreed standards or guidelines. These guidelines [illegible] the results of technology assessment, or from the experience [illegible] senior clinicians nationally or locally. Audit can be [illegible] discovering whether or not guidelines are being adhered to, [illegible] assurance in the process of [illegible] results of audit to improvements in clinical practice.

There are those who are dubious about guidelines, concerned [illegible] restrict innovation and may limit contrasting styles of c[illegible] In particular there is concern that they may not be sufficiently adapted to the individual [illegible] contrasting styles which [illegible] thin line between this and [illegible] junior staff may be left to [illegible] seniors might wish but not really [illegible]. Further, there is fear that legal bodies might give undue [illegible] guidelines and that professionals not conforming to them might [illegible] in difficulty. On the other hand guidelines may be helpful in giving some corporate [illegible] for difficult decisions. However, it should be clear that anyone [illegible]

(5) Medical education, audit, guidelines and financial incentives all have a part to play alongside the dissemination of information in changing clinical practice.

al education, audit, guidelines
part to play alongside the diagnostic
clinical practice.

- 13 -

REFERENCES

Blanpain J E (1986) 'Transfer of technology assessment to health policy makers' International Journal of Technology Assessment in Health Care 2: 117-126

Dawson A (1985) 'The Accountability of Doctor to Doctor' The Lancet, 10 August, pp 323-4

Drummond M (1986) 'Financial Incentives to Change Behaviour towards health Technology' Paper prepared for an EC Workshop, April 22-25, 1986

Fowkes F G R (1986) Strategies for changing the use of diagnostic radiology King's Fund Project Paper No 57, London

Greer A L (in press) 'The state of the art vs the state of the science: the diffusion of new medical technologies into practice'

Gruer R et al (1986) 'Audit of Surgical Audit' The Lancet, 4 January, pp23-26

Gutzwiller F, Chrzanowski R (1986) 'Technology assessment - impact on medical practice' International Journal of Technology Assessment in Health Care 2: 99-106

Maynard A (1987) 'Incentives for Cost-Effective Physician Behaviour' Health Policy, Vol 7, No 2, pp 109-124

Mitchell M W and Fowkes F G R (1985) 'Audit Reviewed: Does Feedback on Performance Change Clinical Behaviour' Journal of the Royal College of Physicians of London, Vol 19, No 4, pp 251-254

Reiser S J (1986) 'Assessment and the technological present' International Journal of Technology Assessment in Health Care 2: 7-12

Wennberg J E 'Dealing with medical practice variations: a proposal for action' Health Affairs 1984, 3: 6-32

REFERENCES

Diamond I A (1986) 'Transfer of [illegible] to [illegible] policy'
makers' International Journal of [illegible]
117 126

Dawson A (1985) 'The Accountability of [illegible] to Doctors' The
August, pp 223-4

Drummond M (1986) [illegible] Charge Delaying [illegible]
Technology' Paper prepared for an EC Workshop. Aug 11 25-35. [illegible]

Fowkes F G R (1986) Strategies for [illegible] the use of diagnostic
King's Fund Project Paper No 47. [illegible]

Greer A L (in press) 'The state of the art vs the state of [illegible]
diffusion of new medical technologies' [illegible]

Greer R et al (1986) [illegible] 'The [illegible]
pp23-29

Gutcmiller F, Chrzanowski [illegible]
practice' International [illegible]
99-106

Maynard A (1987) [illegible]
Policy. Vol 7 No 3 [illegible]

Mitchell M W and [illegible]
Performance Charge [illegible]
Physicians of London [illegible]

Reiser S J (1986) 'Assessment [illegible] procedures'
Journal of Technology [illegible]

Wennberg J E 'Dealing with [illegible] variation
action' Health Affairs [illegible]

CHAPTER 7

FUTURE DIRECTIONS

We began this report by noting that technology assessment has been applied patchily in practice, and even where it has been undertaken results of assessment have not always been acted upon. In subsequent chapters we developed this theme. Chapter 2 focused on variations in the use of health services. The analysis in that chapter demonstrated that there are significant variations in the use of services both between and within countries. Many possible explanations have been put forward for these variations. Although there is a continuing debate about the relative importance of different factors in accounting for variations, much research points to the importance of supply variables and the practice style of individual clinicians. The existence of professional uncertainty concerning indications for treatment and the outcomes associated with different interventions creates a situation in which doctors have a large measure of discretion in deciding whom to treat and how. In this situation, the subjective judgement of clinicians is at least as important as scientific evidence in the decision making process.

In discussing variations, we emphasised that the relevant literature raises more questions than it answers. Areas of work for the future identified in the chapter were:

* The need for systematic monitoring of variations in service inputs, use rates and outcomes;

* The importance of feeding back information on variations to clinicians, policy makers and the public;
* The need to investigate the outcomes associated with different treatments;
* The need for greater efforts on the part of the medical profession to engage in clinical audit;
* The value of the more widespread use of consensus techniques of varying kinds in order to develop guidelines and appropriateness indications;
* The importance of further analysing the reasons for variations, including the relationship between epidemiology and use rates, and the relative importance of demand factors, supply factors and practice style;
* As part of this, the need to complement statistical analyses of large data sets with research into the component parts of practice style to determine how treatment decisions are made by clinicians;
* The possibility of developing standards for use by policy makers and managers, for example concerning GP referral rates, use of outpatient services, and the number of operations to be performed for a given population in particular specialties.

In Chapter 3 we reviewed UK experience of technology assessment. We began by noting that independent commentators have been consistent in arguing for a greater investment in technology assessment and a more coordinated approach. Analysis of UK experience points to the following conclusions:

* To date, the DHSS has been concerned mainly with medical equipment and drugs. In the case of equipment, the Scientific and Technical Branch

of the Department has focused principally on technical performance, safety and mechanical reliability;

* Health authorities do not follow a set procedure in assessing health care technologies. Regional scientific officers play an important part in relation to equipment, and community physicians may be involved in health services evaluation. Although in theory the development of health services is carefully planned and managed, in practice new technologies are often introduced as a result of skilful lobbying by the clinicians concerned, often supported by patients and the public, and not because they are part of agreed policy;
* Consensus conferences have developed as one way of reviewing a diverse range of evidence and offering guidelines for clinical practice and policy makers;
* The medical equipment industry has an ambivalent relationship with government. Government support of the industry on economic grounds may run counter to attempts to introduce more rigorous systems of technology assessment;
* There has been a significant investment in economic appraisal of health care technology but the results of appraisals have had a limited impact. Often, this is a result of the gap that exists between researchers, policy makers and users (clinicians and their patients);
* The National Perinatal Epidemiology Unit is a good example of a specialist academic unit involved in technology assessment. Of particular interest is the work done by the Unit in analysing existing data;
* The MRC has made an important contribution to technology assessment through randomised controlled trials but in general the Council's

contribution to technology assessment has been modest;

* Procedures for drug licensing and surveillance offer one possible model for technology assessment in other areas, although the focus is narrower than technology assessment as defined in earlier chapters.

In an attempt to identify how a more coordinated approach might be introduced into the UK, Chapter 4 reviewed the experience of three countries in detail: The Netherlands, Sweden and the USA. The analysis in that chapter indicated that:

* All three countries make use of consensus conferences, although the method differs somewhat in different countries;
* The Swedish MRC and the NIH in the United States are actively involved in supporting clinical trials;
* The Netherlands and the United States seek to control the use of technologies by identifying particular technologies whose use is either limited or not covered by health insurance plans;
* In all three countries a need has been identified for new agencies to provide leadership and coordination of technology assessment activities;
* A distinctive feature of the Netherlands is work on future scenarios in health care, designed in part to serve as an early warning system for policy makers;
* A distinctive feature of the United States is the use of the OTA to provide support on technology assessment to the legislature;
* Also in the United States, some of the specialist associations have taken a particular interest in developing guidelines, standards and

- Procedures for

for technol

personnel

In an attempt to identify

into the IA, [illegible] a

the Netherlands, Sweden and

that:

- all three countries made

- national [illegible]

- The Swedish Act and the

- In supporting [illegible]

- The Netherlands and the

technologies by [illegible]

limited or not [illegible]

- In all three countries

provide leadership and

activities:

- A distinctive feature of

health care, designed in

policy makers;

- A distinctive feature of

provide support on technology

- Also in the United States,

taken a particular interest in

advice on specific technologies and procedures;

* Finally, the demise of the NCHCT in the United States highlights the political aspects of technology assessment and the threat some organisations perceive from greater coordination of technology assessment.

In Chapter 5 we examined the strengths and weaknesses of different methods of assessment. The following conclusions emerged from our analysis:

* Technology assessment is concerned with the feasibility, efficacy, effectiveness, efficiency, cost-effectiveness and social acceptability of clinical interventions;
* Diagnostic technologies are more difficult to evaluate than therapeutic technologies because their relationship with improved patient outcome is less direct;
* The RCT is the most elaborate and expensive method of assessment but there are difficulties involved in conducting trials as the controversy surrounding the EC/IC surgical trial demonstrates. Many of those involved in the technology assessment field make use of other methods of evaluation;
* These other methods include meta analysis, published reports of clinical experience, clinical audit (including studies of variations in practice), the use of registers and routine statistics, prospectively accumulated data bases, expert systems and decision analysis, and group judgements, including consensus conferences;
* A key issue in technology assessment is the timing of assessment. This should be neither too early (otherwise technology may have developed

advice on specific

* Finally, the desire

political sensitivities

organizations arising from greater consideration of technology

assessment.

In Chapter 5 we examined the strengths and weaknesses of

assessment. The following conclusions emerged

* Technology assessment is concerned with the

effectiveness, utilization, cost-effectiveness

of clinical interventions.

* Diagnostic technologies are more clinically

technical because their relationship with

is less direct.

* The RCT is the most elaborate and expensive

there are difficulties involved in conducting

surrounding [illegible] surgical trial design

involved in

of evaluation.

* These other methods may require

clinical aspects, clinical audit (including routine

practice), the

accumulated data

judgements, including

* A key issue is

should be

insufficiently) nor too late (otherwise it may be difficult to convince clinicians of the need for assessment).

Finally, in Chapter 6 we examined the influence of technology assessment on practice. The main conclusions here were:

* The customers for technology assessment include policy makers, managers, clinicians and patients. Technology assessment seeks to influence both policy and practice;
* There are a variety of influences on clinical practice. Technology assessment is only one of these influences and it may not be the most powerful;
* It is vital to communicate the results of technology assessment to clinicians through the medical press and other channels;
* The evidence indicates that clinicians do not always act in response to scientific evidence. In view of this, methods other than the simple provision of information to clinicians are likely to be needed if the use of technology is to be affected;
* Medical education, audit, guidelines and financial incentives all have a part to play alongside the dissemination of information in changing clinical practice.

One other point emphasised in Chapter 6 was the need to bridge the gap between the scientific community, clinicians and policy makers. This leads us into a discussion of the implications of our analysis for the work of the Institute.

much smaller budget, and with technology assessment as only one of its interests, can only hope to make a limited contribution in the UK context. Given the Institute's mission and resources, the following goals seem appropriate as a starting point for discussion:

* To develop into a coordinating body, building on the work of other institutions and individuals;
* To concentrate on reviewing and synthesising existing data, supplementing this with focused exercises in data collection where appropriate;
* To focus in particular on the evaluation of established procedures, while not ignoring opportunities for doing work on emerging technologies;
* To present results and reports in a form which can be used by policy makers (ministers, civil servants, health authority chairmen, members and managers);
* To work in conjunction with Barbara Stocking and other colleagues in the Fund who also have interests in technology assessment. This may well mean working on similar technologies but the nature of our contributions will be different (eg. we may produce a literature review in association with a consensus conference run by Barbara. We may also prepare a fuller account of consensus conference proceedings and papers than is currently done. We may look too at the impact of consensus results);
* To ensure that the social acceptability of medical technology receives adequate attention and that ethical issues are given due consideration;
* To disseminate results to the medical profession, targeting reports at key individuals and associations.

In view of the size of the field and our limited resources, we will have to be highly selective. For the most part, this will involve selecting specific technologies for detailed analysis. It is suggested that the following criteria are used to identify technologies for assessment of one kind or another:

* There should be disagreement and debate about the technology;
* There should be data available to analyse;
* The technology should be of some significance in resource terms;
* There should be possibilities of building on what others have done, either here or abroad;
* There should be possibilities of linking to other Fund work;
* Analysis of the technology should not have been overdone.

On this basis, one technology we have already committed ourselves to work on is intensive care units. This is to be the subject of a consensus conference in late 1988 and the Institute will collaborate with Barbara Stocking and colleagues in preparing for the conference. This will involve convening a consensus panel in January 1988 and working with the panel on an analysis of the literature on intensive care units. The result will be a preliminary consensus statement and the statement will be discussed, debated and refined at the conference. The resulting publications will include not only a definitive statement but a more comprehensive analysis of the literature on intensive care units prepared by the Institute.

Looking further ahead, there are several other possible topics which suggest themselves for analysis. These include:

a) Work on some aspect of variations in the use of common surgical procedures, developing further the analysis undertaken in Chapter 2. As an example, this might involve taking a high variation procedure, convening an expert group, and establishing guidelines for clinical practice;

b) Undertaking a local study integrating data on variations in inputs, provision, use rates and outcomes. As an example, this might involve taking two districts in a similar RAWP position and identifying variations in use rates and outcomes. The case for carrying out a study of this kind is set out more fully in Chapter 2;

c) Analysis of the use of technology in the obstectric services. Possible topics are the use of electronic foetal monitoring, ultrasound and caeserian sections. However, it may be that these issues have already adequately been analysed by others including the National Perinatal Epidemiology Unit;

d) Work on some aspect of the use of technology in primary health care. As an example, this might involve assessing the opportunities being created by the availability of radiology and pathology testing in general practice. Another possibility is to investigate further the reasons for variations in GP referral rates, responding to the current policy concern with this issue;

e) A study of the comparative costs and benefits of different treatments for heart disease, bringing together data on CABG, transplants, pacemakers, angioplasty etc. There has already been some work carried out in the Fund on these issues, and it might be possible to capitalise on this with appropriate expert assistance.

f) Exploring the wider use of clinical guidelines as a means of reducing

variations in the use of common services.

further the analysis undertaken

might involve taking a high-level

areas, and establishing guidelines

some study integrating data on

practices and outcomes. As an example,

patients in a similar RAMP practice

practice rates and outcomes. The

kind is set out more fully

the use of technology in the

the use of electronic foetal

practice. However, it may be

areas analysed by others including

Health

some aspect of the use of technology

case, this might involve moving to

the availability of radiology and

practice. Another possibility is

implications in GP referral rates

with this issue:

preventive costs and benefits

managing together data on

practice. There has

practice teams, and it might be

practice management

practice guidelines as a

variations in practice, based on examples where guidelines are already acepted.

As part of its work on specific technologies, we suggest that the Institute responds opportunistically to the emergence of new technologies and in selected cases prepares a policy briefing note on these technologies. Inevitably, given the wide range of technologies emerging, the Institute will not possess the expertise required to produce a briefing note, but it is in a good position to convene a group of experts and to act as a secretariat to such a group in the development of a briefing.

More generally, a case can be made for a piece of work on technological futures. This would build on the investment already made in the Netherlands (see Chapter 4) and would seek to identify some of the likely implications of current developments in medical technology for both service users and service providers. The sorts of issues a futures exercise is likely to raise are:

- the impact of advances in laboratory testing (biosensors, diagnostic kits) on hospital laboratories;
- the social and ethical implications of the greater availability of self diagnosis kits;
- the impact of developments in laser surgery and endoscopy on bed numbers, bed use and hospital design.

The Netherlands future health scenarios programme suggests that there will be much greater potential for treatment and care based in the home and in GPs' surgeries in future, and that this will have important implications for the

already

suggests that the [illegible]

of new technologies and [illegible]

uses on these technologies

technologies emerging, the [illegible]

produce a building more [illegible]

needs and to act as a [illegible]

setting.

to a piece of work on [illegible]

assessment already made in the [illegible]

briefly some of the likely [illegible]

technology for both service and [illegible]

process exercise is likely to [illegible]

testing (discussed in [illegible]

of the greater availability [illegible]

energy and embodiment in [illegible]

suggests that there [illegible]

to the home and in [illegible]

implications for the

planning and management of hospital services. There is little evidence that these issues have received serious consideration by policy makers in the UK, although a Delphi Study has been conducted from Guy's Hospital.

Another over-arching area of work concerns the development of health outcome measures. A recurring theme of our analysis has been the difficulty of determining the appropriate use of technologies because of the limited evidence on outcomes. A major priority for the future is to rectify this, in particular by developing more sophisticated quality of life measures. The interest shown in QALYs illustrates the importance and topicality of this issue, yet the basis on which QALYs are constructed is crude as we noted in chapter 5. Williams (1987) has argued that future work should involve asking ordinary people to rank those aspects of quality important to them personally, taking account of variables such as physical mobility, pain and distress, and ability to perform normal social roles. In some cases, measures will be required for a specific condition, in others a broadly based measurement will be appropriate to establish priorities across disparate illnesses and conditions. We do not believe that the Institute has the resources to conduct original research in this area, but we would strongly endorse the arguments advanced by Williams for more work on outcomes.

A further and more ambitious possibility is that the Institute should publish a current awareness bulletin on technology assessment. The raison d'etre for this is that the results of technology assessment do not at the moment find their way onto the desks of those responsible for making decisions on the purchase and use of technology. The technology assessment bulletin proposed here would summarise the results of RCTs and other research and would present

up to date information in a readable and readily accessible form for policy makers. The bulletin would rely on a small number of experts in relevant fields who would monitor developments and submit articles and news items for publication. The bulletin would be available on a subscription basis and might be published initially six times a year. The Institute would be responsible for producing the bulletin and preparing some of the copy, but its principal role would be to establish and coordinate the network of experts who would supply most of the material to be published.

Another possible area of analysis is a review of DHSS policy on acute services. In comparison with services for the priority groups, relatively little guidance has been published by the Department on acute services. Specific targets are set in relation to technologies such as hip replacements and CABG, and special funds are allocated to support supra-regional services. But a policy vacuum appears to exist as far as other aspects of acute services provision are concerned. There has been no follow-up to the acute services review published by the Department in 1981 (see chapter 2) and it would appear timely to undertake an analysis in this area. In this context, it is worth noting that acute services is to be the theme of the annual conference of NAHA in 1988.

Finally, we suggest that the Institute publishes a short and accessible guide to technology assessment for the main consumers of our work. It is undoubtedly the case that technology assessment is a term which is interpreted in a variety of ways by policy makers, managers, clinicians and others. At its narrowest, it may be seen as involving the evaluation of medical equipment, at its broadest an Americanism which has little meaning in the UK

context. To overcome this, we believe the Institute should do a marketing job for technology assessment, giving examples both of promising innovations which did not turn out to be successful as well as those that were undervalued prior to assessment, and suggesting some questions which should be asked when technological developments are proposed. A guide of this kind has been prepared in Denmark, and we would like to produce something similar in the UK.

To summarise, the main thrust of the Institute's work in this area would be to use existing data on medical technology to help policy makers in their business of setting priorities and allocating resources and to influence clinicians to review their practice. The principal output of the Institute's work would be a series of publications. As currently envisaged, these would include:

- A short guide to technology assessment;
- Detailed studies of particular technologies;
- Briefing notes on emerging technologies;
- A bulletin on technology assessment.

These publications will be supported by conferences and workshops as appropriate, such as the one held in June 1987 on variations in the use of services. In the longer term, it may be possible to offer a consultancy service to health authorities along similar lines to the Inter Authority Comparisons Project run by John Yates at Birmingham University. In all of this work, it is vitally important that the Institute draws on appropriate medical expertise and guidance, and collaborates with relevant colleagues in other parts of the Fund. In this way, the policy anaysis contribution of

the Institute can be coordinated with the service development role of the Centre and the management development role of the College.