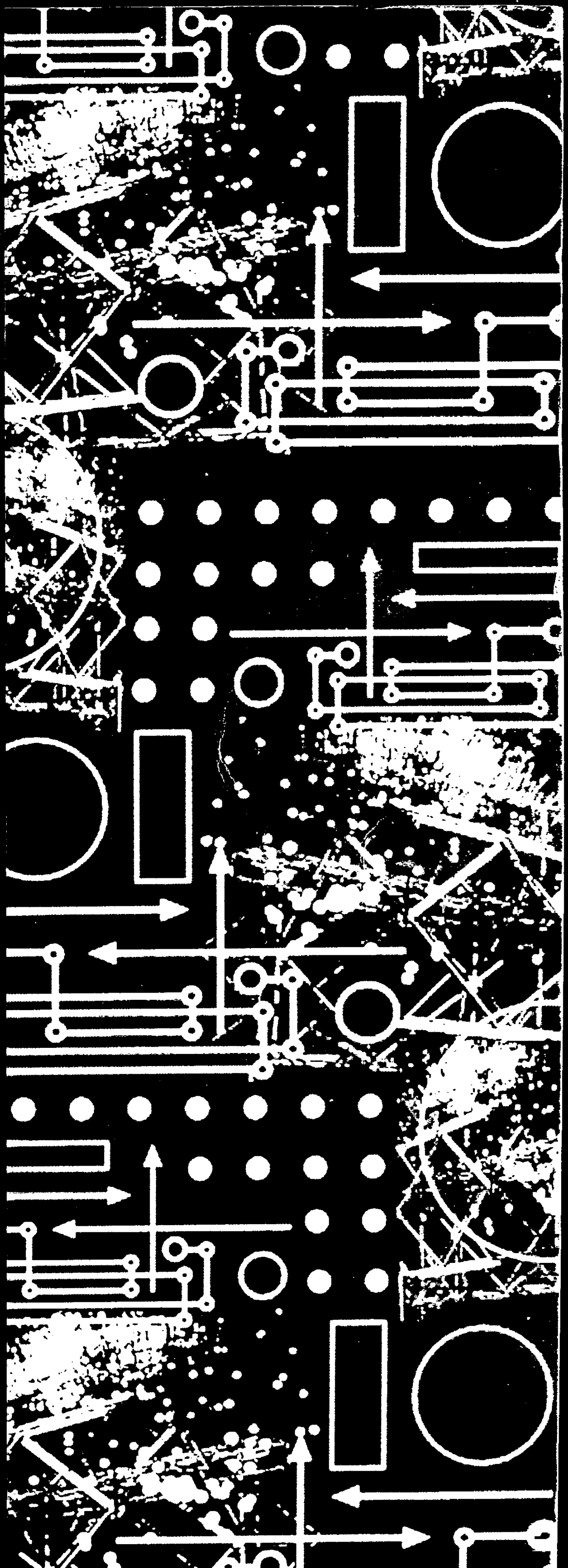
King's Fund

Access to elective care

What should really be done about waiting lists

Anthony Harrison

and Bill New



King's Fund
Publishing
11–13 Cavendish Square
London W1M 0AN

Access to elective care

What should really be done about waiting lists

Anthony Harrison and Bill New

King's Fund

Published by
King's Fund Publishing
11–13 Cavendish Square
London W1M 0AN

First published 2000

ISBN 1 85717 299 X

A CIP catalogue record for this book is available from the British Library

Available from:

King's Fund Bookshop
11–13 Cavendish Square
London
W1M 0AN

Tel: 020 7307 2591
Fax: 020 7307 2801

Printed and bound in Great Britain

Contents

List of Figures, Tables and Boxes

Figures

Tables

Boxes

Introduction

Almost since the foundation of the NHS, the existence of waiting lists for access to elective hospital treatment has posed a challenge to policy-makers. Initiative has followed initiative, as successive governments have attempted to reduce or even eliminate waiting lists. Overseas observers, while expressing admiration for the NHS as a cost-effective means of providing universal health care, have seen waiting lists as its Achilles heel (Light, 1998).

Since the Labour Government came to power in 1997, however, waiting lists have attracted unprecedented attention. Before the General Election, Tony Blair made a reduction in waiting lists of 100,000 one of his 'promises to Britain'. Since its election, the Government has made unprecedented efforts to fulfil that promise and, for a time, appeared to be succeeding in reducing the number of patients recorded as waiting for hospital treatment. Between April 1998 and April 1999, the numbers fell by some 220,000. However, the numbers recorded as waiting more than 13 weeks for an out-patient consultation rose by some 100,000 in that period – the numbers waiting for more than 13 weeks for an out-patient appointment were 247,510 in March 1997 and 456,039 two years later – suggesting that any success the Government may have had with the in-patient waiting list reflected, at least in part, a redistribution of waiting from a later to an earlier stage of the process by which patients achieved access to care. Moreover, from April 1999 onwards, the numbers waiting for in-patient care showed little change.

Such an outcome had been anticipated in an earlier King's Fund paper (Hamblin *et al.* 1998), which argued that not only was the Government pursuing the wrong objective by focusing on numbers waiting for treatment rather than access times, but also that the policies it adopted to achieve that objective were misconceived. In particular, it left out of consideration the forces at work bringing new patients on to the list and the response of patients and professionals working within the NHS to any changes it did succeed in bringing about. The paper argued that recent experience suggested that it was possible for the average time spent waiting to fall and numbers on the list to rise.

However, if times did fall, more people would seek and be referred for treatment. As a result, short of a massive and sustained injection of finance, total numbers waiting would rise even if the Government succeeded in reducing them for a limited period.

The earlier paper based its arguments on a description of the whole of the elective care system, including both the NHS and the private sector, and on analysis of the factors that determine how that system works. This report builds on the earlier framework but it extends it in four main ways:

1. it considers all the stages through which NHS patients must pass to gain access to elective care, including both primary and secondary care
2. it expands substantially the empirical description of how the elective care system works
3. it considers the role of waiting in the elective care system, specifically how it functions as a rationing device
4. it sets out in more detail two broad policy options for 'solving' the waiting list problem.

The report is structured as follows. In Chapter 1, we describe how policy towards access to elective care has developed over the years. In Chapter 2 we set the scene, describing the numbers waiting, the times they wait and recent changes in both. In Chapter 3 we consider a number of issues raised by this account:

- why the numbers of patients waiting rise even when activity increases
- why waits vary between different patients
- why waits matter
  - to the system as whole
  - to the individual
- why waits vary from one area to another.

In Chapter 4 we consider how the broad objectives of any area of public policy – efficiency, equity, and accountability – apply to elective care, and in Chapter 5 we consider a range of policy options.

This report, as well as building on the earlier work by Richard Hamblin, Seán Boyle and Anthony Harrison, draws on a large amount of work by other researchers, references to which can be found in the bibliography.

The elective care system

Although the term *'system'* is now in common use, there is no agreed definition in the health care field. In this report, we use the expression *'elective care system'* to refer to:

a) all the stages through which patients must pass to obtain treatment – in other words, their care pathways between their decision to seek medical help and the ultimate treatment, if any, they receive, and
b) the way in which the different stages may interact with each other.

The care pathway is essentially an ordered sequence of events of the kind shown, in very simplified form, in Figure 2.1 below. It can be described as all the points or stages by which care may be accessed and decisions made as to whether a patient progresses or 'leaves the system'. The pathway implicitly assumes that the various stages can be regarded as a series of linked processes through which patients move more or less rapidly. In such a sequence, Stage 1 influences Stage 2 and Stage 2 influences Stage 3, and so on as patients are referred forward.

The second meaning of 'system' cannot be represented so simply. We argue here that what happens at the initial stages of the pathway may be influenced by the patients' or their adviser's knowledge or belief about what is happening to other patients further down the pathway. In this way, Stage 3 may influence Stages 1 and 2. It is therefore a dynamic or behavioural concept involving interdependence and feedback between its various stages.

Although it is clear what the main components of the elective care system are – patients, GPs and hospital consultants – there is no completely satisfactory way of defining precisely what other elements should be included within it. Some of the factors we consider, such as medical technology, could be regarded as 'in' because many innovations are the result of work by hospital consultants and 'out' because others stem from technical innovation outside the NHS. Other factors, such as the impact of emergency care, could be regarded as forming, along with elective care, the wider hospital system and that in turn interacts with the rest of the NHS and many other areas of interest.

In what follows, we have defined the elective care system fairly narrowly on pragmatic grounds, while acknowledging at various points in the analysis the way it interacts with the wider world. In particular, we have omitted from this report any consideration of the private sector, which, as the earlier King's Fund paper acknowledged, interacts with the public sector in a number of ways. The existence of these interactions does not, we believe, undermine the line of argument developed here.

1. Policy development

The NHS inherited a waiting list for elective care and in nearly every year since its establishment, the numbers waiting for hospital treatment have continued to rise. This has not occurred for want of trying to reduce them: as we shall see, every decade has seen the launch of a new initiative designed to do so. But the problem has always proved intractable. Although the new Government has succeeded in reducing the list to below the record level reached soon after it took office, the rate of decline slowed sharply during 1999 and elective activity devoted to waiting list cases fell from a March 1998 peak by over 17 per cent in less than a year. Furthermore, even if the Government succeeded in reaching its pre-election target, that would still leave more than 1 million on the list.

In the first part of this chapter, we briefly describe the measures taken over years and put forward our interpretation of why successive initiatives have failed. We then describe the policies of the new Government, which appear to address directly some of the weaknesses of those of previous administrations, but which, we argue, are still unlikely to succeed.

Box 1.1: Definitions of a waiting list

Patients are defined as being on a waiting list when they are awaiting elective, i.e. not immediate, admission to hospital. This definition includes booked admissions where patients are given a date for admission at the time of the decision to admit as well as those who had to wait to receive a date for their treatment.

Elective care covers a very wide range of conditions. At one end of the scale are conditions verging on emergencies, which must be treated promptly; at the other, are conditions that may be troublesome but that are not life-threatening. Over the years, official statistics have sometimes acknowledged this fact by distinguishing between urgent and waiting list cases but they no longer do so. Currently, official figures – the Department of Health's quarterly (now monthly) reports on waiting lists – distinguish between those 'with admission date' and the total, and between day case and 'ordinary' in-patient admissions.

Elective activity also includes those with planned admissions, i.e. courses of treatment carried out over a period of time. These admissions are not included in the waiting list figures. In 1997/98, they amounted to about 700,000 cases.

1.1 Policies in retrospect

The 1950s saw what was termed at the time a 'concerted effort' to reduce the length of the list (DHSS, 1980). A few years later, in 1963, the Ministry of Health issued a circular (Ministry of Health, 1963) designed to guide GPs and others on how they should go about gaining admission to hospital on behalf of their patients – primarily elderly people and the chronic sick. It also recognised that there was what it termed a 'backlog of cases' in other specialties. Where patients had to be put on waiting lists, it set out some basic requirements for administering such lists and promised some more detailed advice.

That advice (Ministry of Health, 1964) was issued in the following year. It comprised further practical measures, but this time largely concerned with the way hospitals worked – the use of theatre and surgical time, beds and the use of out-patient care. Optimistically, it stated that:

> *A continuous effort will be needed to prevent a backlog of cases arising again once it has been eliminated.* (para.11)

In fact, the practical measures proved insufficient. The backlog was not eliminated: numbers waiting continued to rise.

The next decade produced another initiative. HSC(IS)181, *Reduction of Waiting Times for In-patients Admission: management arrangements* (DHSS, 1975) was issued in 1975 and set out more practical advice, acknowledging that earlier efforts had not been successful:

> *The seeming intractability of the problem of long waiting lists and times for in-patient admission has been a matter of concern to the NHS authorities, to the staff concerned and to the community, for many years. The problem is naturally greater in some specialties than in others, and varies from place to place. It may be aggravated or abated by major changes in medical techniques or treatment. Other factors are the impact of emergency admissions and the effect of special situations such as epidemics or industrial disputes.* (p.1)

Despite its clear identification of the wide range of factors affecting the length of waiting lists, the circular went on to suggest that shortage of facilities was probably the main reason for delays in admitting patients. But even this was not a straightforward matter since 'remedying one deficiency may do no more than accentuate another in the chain' (p.1), i.e. relief of a blockage at one part of the care pathway might simply put more pressure on another part of it.

The circular did not attempt to tackle head-on the complexities it identified, but focused again on practical matters, particularly the management of lists themselves, drawing on the results of earlier work by the 1972 Working Party on Waiting Lists. Subsequently, £5 million was made available to get rid of bottlenecks, but total numbers still rose.

The Royal Commission (1979), reviewing the NHS in the late 1970s, made a number of pertinent observations about the significance of waiting lists. In particular, it emphasised that it was length of wait rather than numbers waiting that was important, and that the length of wait should include time spent waiting for an out-patient appointment. It also recognised the function of waiting within a service free at the point of delivery – to control access and hence perform a rationing function. But it had nothing substantial to say about how effectively or fairly waiting lists discharged that role and what, if anything, might be done to improve matters. Indeed the Commission acknowledged that it 'had no instant solutions to offer' (para.10.13) to waiting lists.

In 1986, the Conservative Government launched a Waiting List (later, Waiting Time) Initiative, backed by a dedicated fund – initially £25 million for 1987/88 – and consultancy support to help hospitals improve their operational efficiency. Anticipating a pledge made some ten years later, that extra was intended to treat 100,000 more patients from waiting lists by the end of March 1988. Between 1987 and 1990/91, £119 million was allocated through the Fund, primarily to tackle the longest lists. Despite this cash injection, the House of Commons Health Committee (House of Commons, 1991a), reporting in 1991, discovered that lists had continued to rise and that long waits – those over two years – had not been eliminated. The

Conservative Government nevertheless renewed its commitment to reduce waiting times. The *Patient's Charter* published in 1991 embodied that target.

The Waiting Time Initiative was formally ended in 1994/95 (in England), by which time £252 million had been allocated from central funds to support it. At the time the Initiative began, there were some 100,000 people who had been waiting over two years. When it came to an end, that number had been substantially reduced. However, total numbers waiting continued to rise and by the time the Conservatives left office in 1997, they were higher than they had ever been.

Within the totals, there continued to be marked differences between the length of lists in different areas. Reflecting the development of consumerism in the NHS, the College of Health began publishing in the 1980s league tables of waiting times in different parts of the country, with a view to helping patients decide where to go for treatment. The College's annual reports consistently found very large variations between different parts of the country, which the enterprising consumer, or the GP on their behalf, could exploit by 'shopping around' for hospitals offering shorter waiting times. The same perception had underpinned the suggestion made by the Office of Health Economics (Office of Health Economics, 1982) in the early 1980s that waiting lists should be genuinely national, i.e. a managed version of shopping around.

The existence of differences in numbers waiting in different parts of the country was further highlighted during the 1990s by the league tables published as part of the *Patient's Charter* initiative. In effect, these took over the role of the College of Health, by explicitly encouraging users, or GPs on their behalf, to 'shop around', in line with the market principles on which the 1990s reforms were based. In principle, such behaviour could have equalised the length of lists (whether measured in numbers or times) in different parts of the country.

The Government, in its response to the Health Committee's 1990/91 questions relating to public expenditure on health (Department of Health, 1991), expressed the hope that the competition between providers introduced by the NHS and Community

Care Act 1990 would provide incentives to hospitals to reduce waiting lists lest they lost contracts within the internal market, which the Act was designed to create:

> *13. The NHS reforms will provide the management framework to help reduce waiting times Money will follow the patient, and provider units will have the incentive to ensure their waiting time is as short as possible so as to attract contracts from purchasing health authorities and GP fundholders. All NHS contracts will specify waiting time targets among the quality measures which health authorities require to be met.* (p.52)

Some GP fundholders did 'shop around', as did some health authorities. For example, one northern authority arranged for a number of patients to be operated on within a military hospital in the south (Howell *et al.* 1990). But the overall impact of such measures was limited. Indeed, because the new arrangements limited the freedom of non-fundholders to send patients outside their district, they may have had the opposite effect to what was expected. In any case, there were other barriers to 'shopping around', such as lack of knowledge of the quality of care in other hospitals and patients' reluctance to travel.

The 1964 and 1975 circulars, the Waiting Time Initiative and the Conservative Government's hopes for the 1991 reforms implicitly assumed that the answers lay in the way that hospitals in different parts of the country worked. The measures it suggested or, in the case of the Initiative tried to put into place, focused on the way that lists were managed and the way that the resources of the hospital were deployed.

But by the 1980s there was plenty of evidence to indicate that the recorded variations in waiting lists reflected many other factors, more than the 1975 circular had acknowledged. A number of studies had by then appeared, identifying wide variations in the level of treatment received by apparently similar populations. The significance of these variations was assessed in the 1980s in two King's Fund publications. In the first (Ham, 1988), Chris Ham's overview of the evidence found that:

> *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 focused on*

> *demand-side factors such as population characteristics and morbidity, supply-side factors such as the provision of doctors and beds, and on professional decision-making ... Despite the growth in interest in variations studies and the increasing sophistication of the methodologies used, there is a continuing debate about the relative importance of different variables.* (p.9)

He concluded that:

> *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.* (p.14)

A similar conclusion was reached in the second publication, *Variations in Hospital Admission Rates: a review of the literature* (Sanders *et al.* 1989), as the following indicates:

> *Of major importance is the fact that the existence of variations highlights the considerable uncertainty as to what constitutes a 'correct' (in terms of therapeutic value) level of admission for many common medical and surgical procedures. The question remains as to whether high rate areas are admitting too many patients, or whether low rate areas are not admitting people who would benefit from hospitalisation. In the absence of data on the outcomes associated with different levels of admission, it is difficult to determine correct rates of hospitalisation.* (p.38)

In the light of that conclusion it is not surprising that the report acknowledged that it was not possible to establish norms for treatment levels. In his concluding remarks, Chris Ham was similarly cautious, but recommended that norms of service utilisation ought to be developed to provide some means of judging whether levels of treatment in particular areas were too high or too low.

In the mid-1980s, such norms had been set for a small number of what were then new procedures. A 1990 circular (Department of Health, 1990) set targets for coronary bypass operations (a rate of 300 per million); hip operations (1050 per million); bone marrow transplants (12 per million); end-stage renal failure (40 new patients per million). However, the Department's 1990 report (Department of Health, 1990) simply noted that the number of procedures in all four categories had risen more or less in line with these targets. It confined its comments to the national totals rather than identifying those areas that were below the target rates.

During the 1990s, the evidence on variations across a wide range of services continued to accumulate. A series of reports from the Clinical Standards Advisory Group (CSAG, 1993; 1996a; 1996b) identified variations in access to routine elective care and specialised services. The Group concluded that, given the variations' scale, they could not be attributed to differences in needs. In the case of coronary artery bypass grafts (CABG), CSAG found that the rate varied nine-fold between districts, and the rate of percutaneous transluminal coronary angioplasty (PTCA) 62-fold.

CSAG had been established to assuage the concerns of the medical profession that the internal market would have an undesirable impact on clinical standards. Over a series of reports, CSAG did not identify such an impact; instead, it became, for the first time, an external, clinically-based auditor of the way that a large number of services were provided. In the course of this work, it identified not only major variations in service level, but also the lack of a proper basis to ensure that services were provided in an effective manner. In relation to specialised services, it found (CSAG, 1996b) that the contracts introduced in the wake of the NHS and Community Care Act 1990 were an inadequate means of determining what should be provided. More generally, it concluded that:

> *The information available in the NHS remains inadequate for effective operation and local monitoring of an internal market for specialised services.* (p.51)

In other reports (e.g. CSAG, 1998), it found that the basis for determining what a 'satisfactory' level or form of provision should be did not exist. Against that background, it was fanciful to expect that people in different parts of the country would receive more or less the same level of service. That could only be achieved if there was agreement at national level as to what services should be available and how they should be provided.

In its first report on specialised services, CSAG found that purchasers and providers were looking for national guidelines. In its response (Department of Health, 1993), the Government welcomed:

> *... the suggestion of developing, in discussion with managers and the profession, national guidelines concerned with service specifications and quality standards. The NHSME is actively pursuing this with the profession and NHS management, and will seek to include the services studied by CSAG in any programme agreed to develop such guidelines.* (p.7)

The Health Committee had already made this call. Its 1991 report (House of Commons, 1991a) had recommended that there should be joint work between the Department of Health and the Royal College of Surgeons on clinical protocols, including 'standards or agreements about the criteria for deciding ... which procedures are appropriate for which people and when' (para.10).

The Government response to this recommendation (Department of Health, 1991) had also been positive:

> *The development of clinical protocols is indeed now an important policy initiative. Protocols will help tackle the variation in thresholds for treatment around the country so as to improve the assessment of clinical priority, services to patients and the application of resources. The whole object of achieving greater professional consensus on treatment thresholds is to maximise the availability of NHS services in terms of the clinical benefits they yield to patients.* (p.4)

The issue was also raised by the Welsh Affairs Committee in its examination of elective surgery (House of Commons, 1991):

> *Unless the Government grasps the nettle of clinical freedom inappropriate care will continue on a considerable scale; whilst doctors actions should not be directed by non-clinicians, it is not unacceptable that they should follow clinical guidelines drawn up by leading practitioners in the subject.* (p.xxv)

The Committee went on to set out a series of recommendations:

> *The Welsh Office, in co-ordination with the Department of Health, should establish a permanent collaborative unit (in conjunction with the health authorities and the universities) within the NHS. It should:*
>
> *i. identify best practice and draw up clinical guidelines (including threshold criteria) for appropriate interventions in consultation with and approved by professional bodies such as the Royal College of Surgeons (and building on the protocols drawn up by the Welsh Health Planning Forum);*
>
> *ii. ensure that best practice is disseminated through surgical audit;*
>
> *iii. establish literature review units – such as the Perinatal Epidemiology Unit at Oxford – to cover all areas of health care; and*
>
> *iv. attempt to determine the level of need [as well as of demand] for selected conditions. [It could do this by examining on random population samples the effect of applying the treatment criteria laid down in the guidelines; for example it could, using the threshold criteria to be laid down for hip replacements, discover how many hip replacements are needed in a given population].* (pp.xxv–xxvi)

The Government response to these recommendations (Secretary of State for Wales, 1992) was also positive in tone but fell short of endorsing the search for thresholds and their implications for the population at large.

By the time the Conservatives' period of office came to an end, there was little evidence in official publications that a start had been made in responding to the recommendations of the two select committees. A start had, however, been made, through the expert group – always known as Calman-Hine after its co-chairs – towards developing a national approach to the development of cancer services (Expert Advisory Group, 1995). By the time the Conservatives left office, its recommendations were in the process of implementation and, through the new emphasis on clinical effectiveness, a start had been made on weeding out ineffective procedures. But there was no national programme of protocol development or threshold development across the broad spectrum of elective care.

Moreover, numbers waiting for hospital treatment were higher than ever. The Waiting Time Initiative had had only a temporary impact on the total – only the very longest waits had (virtually) disappeared. Over 40 years of attempts to reduce the numbers waiting had resulted in failure. Surprisingly, the priorities set out in the Conservative's final White Paper, *A Service with Ambitions* (Department of Health, 1996), published almost at the end of their period in office, did not include any mention of waiting lists, and the final statement of their public spending priorities for health (NHS Executive, 1996) simply reported that:

> *The NHS has been working to end the longest waits for treatment while ensuring that urgent cases continue to receive priority.* (para.4.83)

The Conservative Government was, apparently, content to let the total numbers waiting rise provided that very long waits were eliminated in line with *Patient's Charter* commitments. The issue of variations in access was ignored.

1.2 The impact of policy

It is impossible to say whether waiting lists would have been even longer in the absence of the policies introduced in the 1950s, 1960s and 1970s. But there is evidence on the impact of Conservative policies from the 1990s onwards. These comprised:

- the introduction of the *Patient's Charter*
- the Waiting List Initiative
- the provisions of the NHS and Community Care Act 1990.

As far as the *Patient's Charter* is concerned, it does appear to have been successful in eliminating the longest waits – as we shall see in the following chapter. But while that clearly benefited the patients concerned, its overall effect is less clear. According to the Clinical Standards Advisory Group report on district elective surgery (CSAG, 1996), surgeons in the trusts visited considered that clinical priorities had been distorted:

> *Examples were given ... to illustrate the fact that clinical priority was not always being preserved. These included patients that were waiting for admission for carotid endarterectomy (for whom an operation could reduce the risk of a stroke) while patients needing less urgent varicose vein surgery, but who were approaching the 18-month limit, were being admitted. In another Trust, patients were being seen in out-patients for vasectomy counselling ahead of patients with possible aortic aneurysms (who may require surgery to prevent a fatal bleed) because the latter had not waited as long.* (p.41)

This report also found that the way that funds had been targeted on waiting list reduction led to inefficiencies: because the extra money came on an *ad hoc* basis, it could not be used to expand capacity on a permanent basis.

> *Local waiting list initiatives (WLIs) were being widely used by purchasers. The amounts involved in each Trust varied from £60,000 to £3 million per year. WLIs were disliked by Trusts and purchasers, particularly the former.*

> *Money allocated in this way was difficult to use efficiently as it was non-recurring, and, therefore, could not be used to make substantive appointments. It was normally offered late in the financial year (in the winter), to be spent before the following April.* (p.41)

The third strand of Conservative policy was the internal market. As noted above, the Conservative Government expressed the hope that the introduction of a separate purchasing function, with the explicit contracting process that implied, would improve matters. But the impact of the 1991 reforms appears to have been small. The King's Fund assessment of the impact of the reforms (Le Grand *et al.* 1998) concluded that:

> ... *it is impossible to tell whether the internal market has significantly contributed to or hindered [the reduction of waiting lists].* (p.127)

CSAG found that (CSAG, 1996) waiting times varied according to purchaser, as they did between health authorities and between health authorities and GP fundholders. Competition had failed to even-out these differences.

Thus, the policies adopted since the 1980s towards elective care were not successful even though, by eliminating the longest waits, they benefited some patients. There were two main reasons for this failure.

First, none of the policy initiatives have been based on an adequate account of the nature of the problem to be tackled. The simple view, explicit in the 1960s, that the waiting list is a backlog, which extra resources and better management could clear, was never fully dispelled despite the perceptive nature of, for example, the Royal Commission's comments (1979) as well as those of many outside commentators (Frankel and West, 1993).

Thus the Conservatives' Waiting List/Time Initiative embodied the view, inherent in the advice issued in the 1950s and subsequently, that numbers waiting could be reduced by better list administration and better management of existing hospital capacity, supported by modest amounts of targeted funding. Although the two select committee reports we have cited contained the elements of a more fundamental approach, they had no significant impact on policy, which continued to be 'more of the same' until the Initiative was wound up.

Second, the question of equity of access was never systematically addressed. From the 1970s onwards, systematic attempts were made to ensure that the resources available to each part of the country were more or less the same, after taking account of their different levels of need, but the further step of ensuring that the resources allocated to treatments of specific kinds – such as joint replacement – was not taken. This can largely be explained by the general reluctance on the part of the Department of Health/NHS Executive, throughout the whole of the period considered here, to determine clinical practice. That gap was not filled by the Royal Colleges, which, despite their pervasive influence on clinical practice through training and accreditation, were not in a position to enforce common standards.

The persistence of variations between different parts of the country might seem at odds with the fundamental objective of the NHS, with its foundation of offering care according to need. But although the NHS had always had an implicit commitment to equity as between different parts of the country, in practice central policy had never attempted to define in explicit terms what this meant. The 1962 Hospital Plan for England and Wales, the reallocation of financial resources following from the work of the Resource Allocation Working Party in the 1970s and the role of the Medical Practices Committee in respect of the distribution of GPs, were designed to ensure that more or less the same volume and type of services were available in all parts of the country. This broad and continuing commitment, however, was not converted into policies designed to ensure that, within practical limits, people in different parts of the country did enjoy similar levels of access to specific forms of treatment.

1.3 A new Government

The Labour Government came to power in May 1997 with an explicit commitment to cutting the numbers waiting for hospital treatment and to eliminating long (over 18 months) waits. *The National Priorities Guidance* (NHS Executive, 1998) issued in September 1998 describes the first NHS lead priority as follows:

> *Reducing NHS waiting lists and waiting times are key objectives for the NHS and are central to restoring public confidence in the service. The Government is committed to reducing in-patient waiting lists, over the life of this Parliament, by 100,000 from the position it inherited.*

The *Guidance* goes beyond total numbers waiting to set out objectives bearing on numbers waiting, times spent waiting and booked admissions: see Box 1.2.

Box 1.2: Objectives for waiting lists and times – *National Priorities Guidance*

- Contribute to the achievement of the Government's commitment to reduce NHS waiting lists by 100,000 from the position it inherited; and deliver a consequential reduction in average waiting times.
- Improve the dependability of care and reduce the uncertainty of waiting, by increasing the proposition of day case patients offered a booked date for treatment.
- Ensure the delivery of the 18-month maximum waiting time guarantee for admission to hospital for all patients.
- Ensure appropriate investment in out-patient and other ambulatory care services to work towards reducing the time patients wait for first out-patient appointments.
- Pilot the most effective ways of reducing waiting times in Accident and Emergency.

The commitment to reducing numbers waiting has been backed by a series of financial allocations targeted at elective care or at measures designed to make the flow of elective patients through the hospital system easier to manage. These allocations have been made from the NHS Modernisation Fund announced when the results of the 1998 Comprehensive Spending Review were made public. The Department of Health told the Health Committee (House of Commons Health Committee, 1999) that in 1998/99:

> - *£320 million was invested directly on elective activity – to deliver record increases in the number of operations and surgical and medical sessions;*
> - *£65 million was targeted for investments in primary care, community, mental health and social services so that people who didn't need to go to hospital could get better care in their communities;*
> - *£5 million was made available to establish 24 pilot sites as part of the National Booked Admissions Programme. The new system helps to take away patients' uncertainty and stress about waiting for a hospital appointment and enables patients to agree dates which suit them, and take into account their family and work responsibilities.* (p.212)

It also indicated that for 1999/2000, a further £320 million would be spent from the Modernisation Fund. More significant for those working in or managing hospitals, the Government threatened to impose sanctions on those who fail to cut their lists, as well as rewarding those who do.

In taking these measures, the Government could be seen simply as pursuing, if a little more vigorously, the type of measures adopted by its predecessors. In two significant ways, however, it has gone further. First, when allocating some of the targeted funds, it has supported measures to improve the way the NHS as a whole works – the so-called 'whole systems' approach advocated in the *National Priorities Guidance* (NHS Executive, 1998) and described as follows:

> *Bringing down waiting lists and keeping them down in line with clinical priorities, whilst meeting the demand for emergency care, requires a 'whole-systems' approach involving primary care, community, mental health and social services.* (p.18)

We shall argue below that the Department's definition of a 'whole systems' approach falls short of what is required – in particular, it fails properly to take into account the role of primary care in identifying needs for treatment. However, the *Guidance* and subsequent advice (Department of Health, 1997; 1998) make the important point that the elective care system has important connections with other parts of the NHS. The 'whole elective care system' is embedded in the wider system of health care delivery, particularly hospital-based emergency care, and hence there are trade-offs between the two – which we consider in more detail below.

Second, it announced in *The New NHS* (Department of Health, 1997) a series of policies designed to ensure that in practice people in different parts of the country receive similar standards of service. Even before the White Paper was published, the new Government had ordered hospitals to keep common waiting lists, i.e. to cease giving priority to the patients of fundholders. *The New NHS* itself set out a broader-based approach designed:

> *to renew the NHS as a genuinely national service. Patients will get fair access to consistently high quality, prompt and accessible services right across the country.* (p.11)

The subsequent consultation paper, *A First Class Service* (Department of Health, 1998), sets out a number of weaknesses in the current situation which give rise to differences between different parts of the country:

> *Today's NHS does not fulfil the highest expectations for everyone. For a national public service like the NHS there are unacceptable variations in performance and practice. The inequalities go beyond the provision of medicines and other treatments. There are inequalities in the way that some proven treatments get introduced to the NHS too slowly while other unproven treatments can be introduced too quickly. There are inequalities in waiting times for operations; in the time it takes for patients to receive test results; in the number of people given screening tests. There are inequalities in clinical practice – and in clinical outcomes.* (p.5)

In this way the Government signalled its intention to take equity of access seriously. *The New NHS* also reflected the Government's intention to hold the NHS explicitly to account for its performance, in respect of equity, as one of six sets of criteria against which it would be judged. That intention was expressed in the consultation paper on the NHS performance assessment framework which set out a range of proposals for measuring NHS performance including and indicator of equity of access to elective care. This indicator is defined as follows:

> *(i) Surgery rates, composite consisting of age standardised elective rates for:*
> - *CABG and PTCA;*
> - *hip replacement (aged 65 or over);*
> - *knee replacement (aged 65 or over);*
> - *cataract replacement.*

Although it focuses on only a narrow range of procedures, it reflects a greater commitment to equity of access than any previous Government has made.

Against this background it would seem that access to elective care is to be more systematically addressed than ever before. The central argument of this report, however, is while the new initiatives, if successful, will move the NHS in the 'right' direction as far as equity is concerned, they are insufficient to 'solve' the waiting list problem.

To reach that conclusion, we set out in Chapter 3 our view of how the elective care system works. That view rests in part on the way that the list has developed over the years and it is to this we turn in Chapter 2.

2. The current situation

In this chapter, we set out some basic facts about access to elective care. We begin by setting out the main stages through which people must pass to gain access to elective hospital care: see Figure 2.1. The pathway, as we have termed it, from perception of 'a problem' to final treatment, consists of a series of processes, typically carried out by different professionals, in different organisations.

Figure 2.1: The pathway to care

Stage 1: Perceive illness and decide to seek medical advice
↓
Visit GP
↓
Stage 2: GP consultation. Decision on referral
↓
Refer to out-patient consultation
↓
Stage 3: Consultant diagnosis. Decision to place on waiting list
↓
Place on waiting list
↓
Stage 4: Consultant decision on admission for treatment
↓
Treatment: Health outcome

Each stage involves a decision, initially by the individual patient and then subsequently by a professional – albeit in some degree of partnership with the patient, as to whether a further stage is required. Stages 2, 3 and 4 can be seen as a process of filtering the important/serious from the less important and less serious as decisions are made as to whether the patient should proceed to the next stage.

At all three stages, more information may need to be obtained to determine what the decision should be. For example, the nature of a patient's illness may not be clear from the symptoms – Stage 2 may comprise a series of visits before the decision to refer is made and Stage 3 may involve a series of tests as well as consultations. A slightly different sort of delay may occur before Stage 1, depending on how long the patient waits before deciding to visit the GP. That, as we see below, may be influenced by a wide range of personal factors. This 'wait' is outside the system, however, and not directly influenced by health care professionals, unless specific attempts are made to do so, e.g. if GPs discourage 'frivolous visits'.

We can also see clearly from this 'access pathway' that there are three main points where waiting may take place: before seeing the GP, before seeing the consultant for an out-patient consultation, and before being admitted for treatment. Of course, waiting may also take place within the hospital whilst arrangements are being made for the intervention itself, even after admission. And these three waiting categories will have sub-categories: such as the wait for notification of an out-patient appointment, the wait for the day of the consultation, and then the wait in the out-patient waiting room. But from an analytic point of view, these main categories are the most important.

The figure omits many of the key features of this process, but it serves to provide a framework for the descriptive material that follows: the aggregate waiting list figures, variations between specialties, and the time spent waiting; how the aggregate data have changed over the years; and the evidence on variations between different parts of the country.

2.1 The aggregate picture

Table 2.1: Access to elective care

	1991–92	1996–97	1998–98	1998–99
(a) GP consultations[1] (millions)	n/a	270.97	n/a	n/a
(b) GP written referrals to out-patient clinics[2] (thousands)	6,185	7,940	8,165	8,296
(c) Out-patient first attendances[2] (thousands)	8,036	10,415	10,643	10,803
(d) Additions to the in-patient waiting list[3] (thousands)	3,258	4,112	4,192	4,189
(e) Elective episodes of care[4] (thousands)	n/a	4,967	5,064	5,475

1. Source: General Household Survey. Data is a three-year moving average except 1996–97 where only two years' data were available. There was one survey in 1997–98 and results from the 1998–99 survey will be published in November/December 1999.
2. NHS trust based data. Figure for 1998–99 is estimated from quarterly monitoring data.
3. Number of additions to the in-patient waiting list split between waiting list and booked admissions are not available. The count of additions to the waiting list do not include planned admissions (which are also excluded from waiting list figures).
4. Health authority based data. Figures are finished consultant episodes (FCEs) in the general and acute sector. They include activity not related to the waiting list.

Source: House of Commons Health Committee, 1999: 214.

Table 2.1, prepared by the Department of Health in response to a question from the Health Committee, represents the first official (published) attempt to describe the flow of patients through the elective care system. As the notes indicate, the figures come from a variety of sources, reflecting the fact that the NHS itself has never attempted to monitor the elective care system as a whole. Instead, official figures provide only an incomplete snapshot of the system.

Waiting lists have until recently been measured only from the moment when the patient is selected for treatment and is waiting for it to take place. In 1994 the then Government accepted, as the Royal Commission had argued nearly 20 years earlier, that this gave an incomplete picture of the waiting that patients experienced. A further measure of numbers and of waits was subsequently introduced, comprising those who had been referred to a specialist and who were waiting to be seen in an out-patient clinic, i.e. those waiting for a Stage 3 decision.

Measurements of waiting times have only been made since 1988. As Table 2.2 shows, mean and median waits have fallen steadily until 1997. The former has fallen faster than the latter because of the virtual elimination of very long waits as a result of the introduction of a requirement in the *Patient's Charter* to cut out waits of more than 24 months – and later those over 18 months. Elimination of these very long waits has a greater impact on the average, which has more than halved, than the median wait, which reflects only the numbers of people. But both had declined considerably until the last three years.

Table 2.2: Mean and median waiting times for admission to NHS hospitals in England (provider-based), June each year

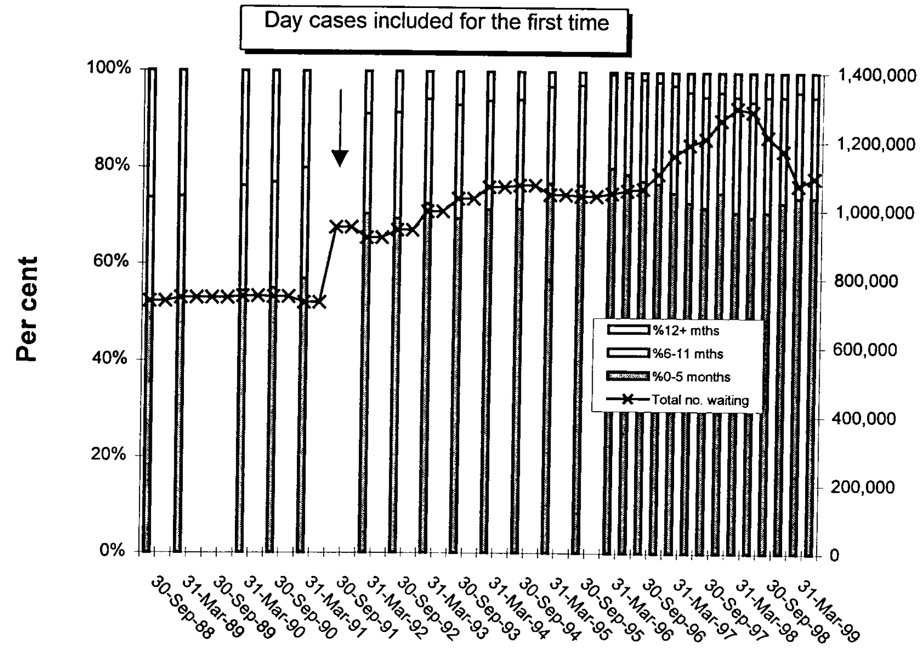
	Mean	**Median**
1988	9.36	5.13
1989	9.15	4.99
1990	8.55	4.77
1991	7.53	4.37
1992	5.02	3.45
1993	4.72	3.47
1994	4.64	3.35
1995	4.08	2.95
1996	3.81	2.84
1997	4.34	3.09
1998	4.67	3.50
1999 (March)	4.30	2.96

Note to table: these figures are based on those who are finally treated not the total of those who waited for some period of time.

Source: House of Commons Health Committee, 1999.

These reductions suggest that a substantial improvement has taken place over these ten years. But in fact the picture is rather more complicated than these summary data suggest. Figure 2.2 presents the most detailed representation currently available of changes in the *proportion* waiting in various categories and the *total* numbers waiting, over the past ten years. For most people, waits are short. About half gain access within three months of being given an appointment and another 25 per cent within six months. But these proportions have been changing: until mid-1996 the proportion waiting under six months had been steadily rising; since that date it has fallen. Conversely, the proportion waiting over 12 months (i.e. all the 'long' waiters) had been falling until that date, since when it has started to rise again. During this period the total number on the list has risen steadily with only two 'dips': one between 1994 and 1995, and the other starting in mid-1998 and lasting until the early part of 1999.

Figure 2.2: Proportion waiting in various time categories, and total waiting, in-patient, 30 Sep 1988–31June 1999, England



Source: Department of Health.

Even this detailed picture hides what has happened to the very long waiters who, because of the way data collection changed over this period, cannot be represented in the same figure. In short, at 31 March 1992 there were still 1490 people waiting over 24 months, but by September of that year there were virtually no patients waiting that long. Numbers waiting over 18 months were again virtually eliminated by 30 June 1996, made a brief reappearance (reaching a peak of 972 at 31 December 1997), but were completely eliminated by 30 June 1998. As noted above, these developments were directly related to the introduction of a *Patient's Charter* 'right', introduced in 1991, that no-one should wait more than two years for elective treatment, subsequently reduced to 18 months in April 1995. From this analysis it should be clear that falling average waiting times *can* co-exist with increasing overall numbers on the list.

The data also suggest that the average may fall while some patients have to wait longer. Very few people have waited over 18 months since mid-1996 – the removal of these long waiters will have contributed to the decline in the average wait. But in order to achieve this reduction, some of these long waiters would have been treated at the expense of shorter waiters (nought to five months), the proportion of which in the total duly increased.

Because those who were relatively urgent cases would still have priority, along with those coming up to the 18-month cut-off, it is likely that something of a 'bunching' phenomenon occurred, with those in between these two extremes waiting longer than previously – the commonly cited problem of 'distorted' priorities. The result has been the increase in the proportion waiting between 12 and 18 months, shown in Figure 2.2.

The existence of this 'bunch' of long waiters (12–18 months) is already placing considerable pressure on the ability of the NHS to meet the 18-month maximum, and requires further 'distortion' of normal prioritisation based on existing clinical criteria.

Table 2.2 refers only to the time patients spend between being accepted for hospital treatment and the time they received it. Since 1995 time spent waiting between GP referral and first out-patient appointment has also been recorded. As Table 2.3 shows, the proportion waiting for less than four weeks for an appointment and the proportion waiting under 13 weeks have both declined, and the proportions waiting over 26 weeks has doubled from 3 to 6 per cent.

Table 2.3: Waiting times for out-patient appointments, England, 1995–99

	Total number of first attendances seen	Total number of GP referrals seen	Per cent GP referrals seen within % <4	% <13 wks	% <26	Numbers waiting 13 to <26 wks	Numbers waiting 26+ wks
30 Jun 95	2,415,900	1,804,000	40	85	97	-	-
30 Jun 96	2,577,036	1,869,782	39	85	97	-	-
30 Jun 97	2,689,803	1,904,564	39	85	97	210,000	85,000
30 Jun 98	2,660,499	1,827,140	38	82	96	269,000	109,000
30 Jun 99	2,714,179	1,837,726	35	78	94	339,000	146,000

Source: Department of Health.

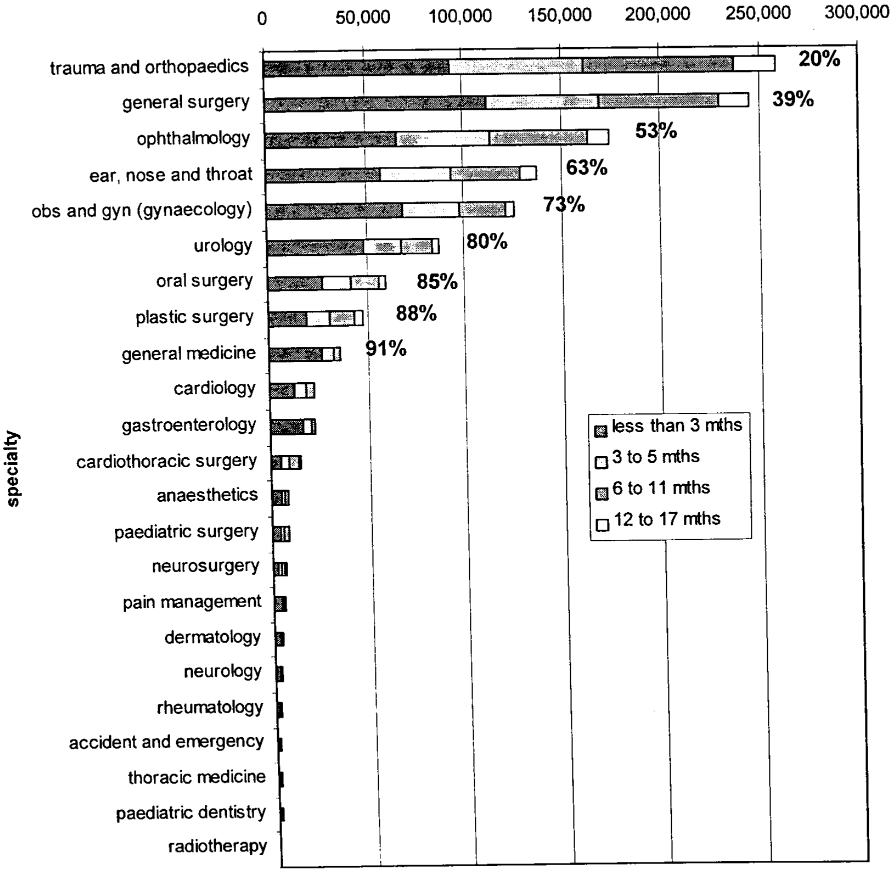
Unfortunately, by virtue of the way that the data on time spent waiting for an out-patient appointment are collected, it is not possible to link the two sources, still less is it possible to track from these sources overall delays in gaining access to hospital care. However, the total recorded as waiting at the end of each quarter has been rising even while the numbers recorded as waiting at Stage 4 – see Figure 2.1 – have been falling. It appears, therefore, that the present Government's successes at the one stage have been offset by losses at the other.

There are no national data bearing on waits at Stages 1 and 2 in Figure 2.1. Delay in taking action over a symptom can only be recorded retrospectively and through special surveys. As we note below, some surveys of this type have been done typically for specific conditions. Waiting times for an appointment with a GP are typically short and are also not regularly recorded on a national basis.

More substantial delays at Stage 2 may arise through delays in reaching diagnoses or referral decisions. Information on these is also not available from any routine data series. However, as we shall show below, for some conditions delays at these stages are important relative to measured delays: the only data available on them, however, also comes from small-scale studies of particular conditions.

The data presented in Tables 2.1–2.3 aggregate all kinds of activity. There are no published data covering the whole of England that describe waits for specific procedures. Figure 2.3 sets out the pattern of waits at specialty level: it shows that six specialties account for 80 per cent of waiters. The proportion of long waiters is even more concentrated, with almost 80 per cent of those waiting over six months in the first five specialties. These are the specialties identified in the 1960s as having the longest waits.

Figure 2.3: Number waiting in various specialties, England, June 1998, and cumulative percentage of total



Source: Department of Health.

Variations in access

We noted in Chapter 1 that variations in the length of waiting lists between different parts of the country have been a long-standing feature of the NHS. In principle, the introduction of the *Patient's Charter* nation-wide in 1991, as well as the scope for 'shopping around' that the NHS and Community Care Act 1990 promised to create, should have led to their elimination. In fact, there continue to be substantial differences between parts of the country. As Table 2.4 indicates, the proportion of the total numbers recorded as waiting for over 12 months in different health authorities varies from 15 per cent of those on the list to zero. Those seen within six months vary from just over half to nearly 100 per cent: see Table 2.5.

Table 2.4: Variation in proportion of people waiting 12–17 months in selected health authorities, 1998

Top and bottom ten health authorities	% total waiting 12–17 months
West Sussex HA	15.4
Lambeth, Southwark & Lewisham HA	13.7
South Humber HA	11.8
Hillingdon HA	11.3
West Hertfordshire HA	11.0
North Essex HA	10.6
North Derbyshire HA	10.1
Ealing, Hammersmith & Hounslow HA	10.1
Enfield & Haringey HA	10.1
South Essex HA	10.1
Doncaster HA	0.5
Birmingham HA	0.5
Gloucestershire HA	0.5
Solihull HA	0.4
Walsall HA	0.4
Sandwell HA	0.3
Northumberland HA	0.2
Sunderland HA	0.1
Tees HA	0.1
Dorset HA	0.01

Source: Department of Health.

Table 2.5: Variation in people waiting 0–5 months in selected health authorities, 1998

Top and bottom ten health authorities	% total waiting 0–5 months	total
Dorset HA	98.1	13,443
Dudley HA	90.2	6038
Walsall HA	88.9	4507
Sandwell HA	86.8	6174
Portsmouth & SE Hampshire HA	85.1	11,494
Gloucestershire HA	84.2	10,040
Calderdale & Kirklees HA	83.4	12,362
Wolverhampton HA	83.3	5397
Doncaster HA	83.3	5287
Wirral HA	82.9	6434
Lincolnshire HA	61.8	20,585
West Hertfordshire HA	61.7	14,726
South Essex HA	61.7	22,457
Bedfordshire HA	61.6	13,862
North Essex HA	61.3	29,009
E&N Hertfordshire HA	60.6	16,753
North Derbyshire HA	59.0	11,369
South Humber HA	56.4	10,912
West Sussex HA	55.1	25,200
Lambeth, Southwark & Lewisham HA	54.8	21,671

Source: Department of Health.

Large variations are also found on the provider side. The Clinical Standards Advisory Group report on district elective surgery (CSAG, 1996) examined waits in 11 trusts in five specialties. As Table 2.6, covering one specialty indicates, there is considerable variation between trusts. Most strikingly, Trust J has virtually no long waits while others have substantial numbers waiting up to a year.

Table 2.6: Time waited for ophthalmology treatment in nine trusts

	Months already waited (per cent)				
Trust	0–2	3–5	6–8	9–11	12+
A	40.1	24.8	18.5	14.7	2.0
B	37.3	27.4	19.6	10.6	5.1
C	40.2	30.5	19.6	9.3	0.3
D	30.9	25.4	19.7	13.1	10.8
F	42.4	25.2	16.4	11.2	4.9
G	49.0	26.2	15.7	8.9	0.1
H	30.8	23.6	17.6	13.3	14.7
J	91.5	7.3	0.8	0.4	0
K	45.7	24.6	15.1	9.2	5.4

Source: CSAG, 1996.

Variations in utilisation

Despite the availability of hospital episode statistics for all parts of the UK, there has been no recent and comprehensive study of the scale of variations in utilisation of elective care for individual procedures. However, there is substantial evidence from other sources that rates of utilisation vary considerably in ways that cannot satisfactorily be explained by differences in the characteristics of local populations. As noted above, the evidence summarised in the earlier King's Fund papers covering the period up to the end of the 1980s has since been amplified by the Clinical Standards Advisory Group in a series of reports in the 1990s, covering both common and specialised procedures.

In respect of district elective surgery, the Clinical Standards Advisory Group found that, after standardising for age and sex, rates in particular specialties and particular procedures varied widely.

Table 2.7: Annual elective procedure rates per 10,000 for inguinal hernia in a sample of ten district health authorities and boards in the UK, directly standardised for age and sex

District or board	1992/93	1993/94	1994/95*
A	17	16	16
B	26	15	16
C	n/a	11	12
D	15	12	10
E	10	10	9
F	20	19	20
G	17	20	27
H	n/a	n/a	15
I	17	16	17
K	16	18	18

Notes: *Rates for 1994/95 are based on data for the first six months only and are provisional. Note absence of J, for which no data was obtained.

Source: CSAG, 1996.

Table 2.7 reproduces its findings for inguinal hernia repair, one of the most common procedures and one for which the indications are straightforward. The highest rate identified – 27 per 10,000 – is three times the lowest. Other procedures show similar or even greater variations.

CSAG findings for common procedures are mirrored in the less common. In its 1996 review of specialised services (CSAG, 1996b) it reached the general conclusion that:

> *Inequitable access to specialised services remains a problem despite current attempts to purchase care on the basis of need.* (p.45)

It found that:

> *Revascularisation rates were higher in districts with least need (as measured by standardised mortality rates) in 1991/92 and this has persisted over the following three years. While rates were higher in more socially deprived districts in 1991/92 (possibly due to the location of tertiary centres in inner city areas), this relationship had disappeared by 1994/95, thus providing further evidence that rates of revascularisation are not linked to the apparent level of need.* (p.40)

Finally, analysis of hospital episode statistics reveals that there are marked variations in treatment rates for specific age groups. Table 2.8 shows the extent of variation in elective care admission rates per age group across England. The spread is limited for the two youngest age groups but is particularly marked for the older age groups.

Table 2.8: Rates per 1000 of elective care in English health authorities by age group, 1998

	0–4	**5–14**	**15–44**	**45–64**	**65–74**	**75+**
0–50	64	66	25	9	4	4
50–100	19	23	46	30	6	8
100–150	8	3	11	23	26	17
150–200	1	4	7	15	15	20
200–250	3	0	4	9	12	12
250–300	0	0	1	3	11	10
300–350	2	0	0	3	7	4
350–400	0	0	0	3	5	9
400+	0	0	0	3	9	14

Note: the data are ‘unchecked’, which may explain some of the more extreme figures

Source: Barrett, 1999.

2.2 Overview

This brief review of the data on access to elective care reveals the following:

- the numbers passing each year through the elective care system continue to rise at all stages of the process
- waiting time for in-patient admission has come down in recent years even when more people have been recorded as waiting, although the average has started to rise
- national figures fail to reveal the full extent of waiting since they cover only the last two stages by which people gain access to care. The earlier delays are not recorded in official data
- rates of treatment vary widely in different parts of the country overall, and both for particular age groups and procedures
- recorded waiting continues to vary widely between different providers and between different parts of the country, as well as between different specialties. Rates of intervention vary widely from one part of the country to another. These data do not in themselves prove that the waits patients experience vary widely, since the way that lists are recorded and managed also vary – as do the underlying epidemiology. But these and other similar findings combine to make a strong *prima facie* case that access is far from equal from one part of the country to another.

3. How the elective care system works

In this chapter, we switch our attention from the numerical data describing the current pattern of waits to a consideration of how the elective care system works. As we noted at the end of Chapter 1, our central critique of waiting list policy since the early days of the NHS is that it has been based on an incorrect view of the nature of the problem to be solved. If the waiting list were simply a backlog, it would have disappeared by now. In this chapter we propose an alternative approach.

We begin by considering the reasons for waiting occurring and the consequences of waiting at various points in the system. Then we examine the workings of the system as a whole, and suggest that the more obvious 'negative' aspects of waiting and delay must be considered against the less obvious, but equally important, rationing function that the waiting list system performs. Finally, we discuss the origin and causes of variations in waiting.

3.1 Why people have to wait

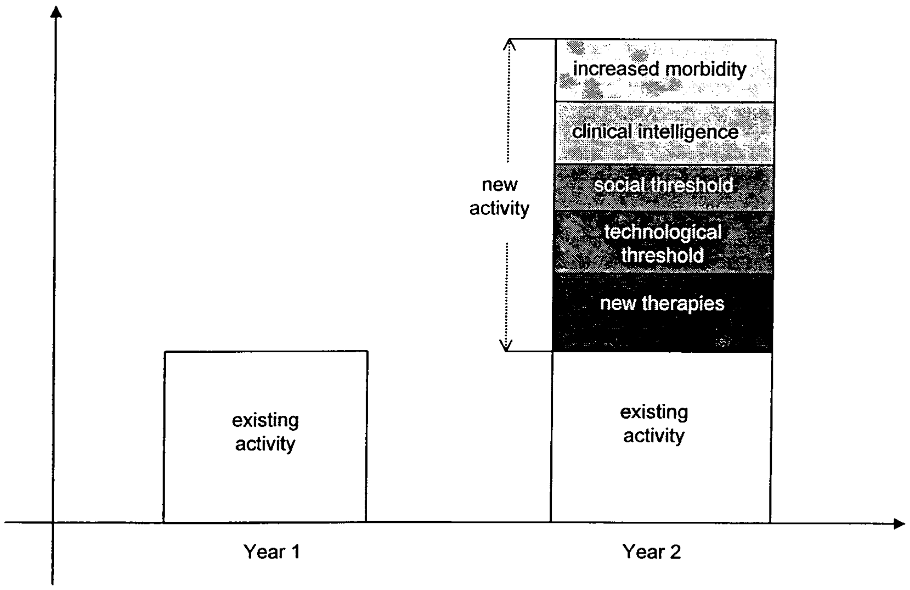
The most obvious explanation of waiting for elective care is that the hospital system does not have sufficient capacity to cope with the patients seeking access to it. As we noted in Chapter 1, for most of the life of the NHS the assumption has been that if hospitals could do better with the resources they have, or if they were given more resources, waiting would reduce or even disappear. That assumption also underlines the policy of the present Government. Variations between different parts of the country have also been assumed to reflect, at least in part, different levels of efficiency in carrying out elective procedures.

The earlier King's Fund paper argued that the assumption that an increase in the volume of elective care would in itself remove waiting lists was based on a misunderstanding of the reasons why lists grew. Changes of definition make long-term comparisons hazardous, but there is no doubt that the number of procedures has increased substantially since concern about waiting lists emerged during the 1950s.

Indeed the increase in total activity since the start of the Waiting Time Initiative in 1986 – well over a million procedures per annum – exceeds the number recorded as waiting at that time. In fact, as the level of activity has gone up, numbers waiting have risen more or less in the same proportion, leaving the ratio of numbers waiting to total treated at a fairly constant level equivalent to about three months' activity (in England).

Why has extra spending not succeeded in eliminating waiting? The answer in brief is that while resources and activity have increased steadily over the life of the NHS, changes in medical technology allied with economic and social factors have extended the range of potentially valuable procedures faster than activity has risen. Figure 3.1 provides a graphical representation of the factors at work.

Figure 3.1: Potential explanations of new elective activity



Note: see Hamblin *et al.* (1997) for a fuller discussion of this figure.

Increased morbidity may reflect the effects of an ageing population, or 'resistant' bacteria immune from previous therapies, or of deteriorating environmental conditions. *Clinical intelligence* refers to the increased knowledge, awareness and expertise of clinicians, particularly GPs, who become more adept at diagnosing a wide range of conditions. *Social thresholds* reflect increasing expectations and perhaps social changes in people's willingness to 'put up' with ill health. This category may also include a widening of the scope of health services to include conditions, such as mild depression, which were once considered non-medical. *Technological thresholds* change when existing therapies become appropriate for a higher proportion of patients within particular disease categories – for example, end-stage renal therapies being used on increasingly elderly people, or on those with co-morbidities. Finally, *new therapies* relates to the constant development of new interventions that ameliorate or cure previously incurable conditions.

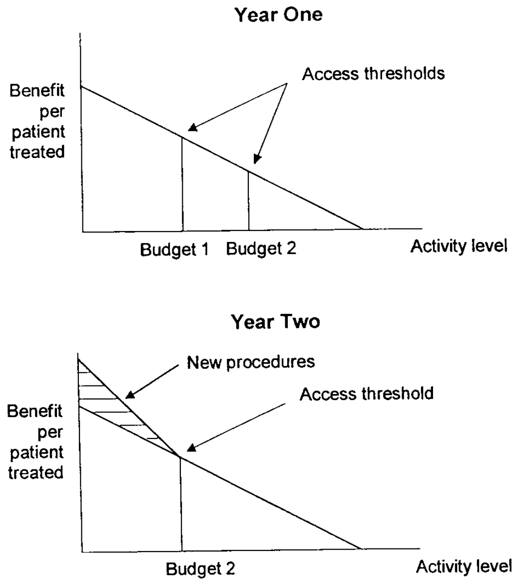
We have already noted that the rise in activity over the past ten years has been more than sufficient to eliminate the numbers waiting at the beginning of the period. The list, in other words, is not a backlog but a pool of potential patients, which is continually being topped-up. This topping-up reflects the factors set out in Figure 3.1. Against this background, the persistence of waiting to gain access to hospital care is a reflection of the fact that increases in the resources available, and the efficiency with which they are used, have been unable to keep up with the forces set out in the figure.

Some of the factors set out in Figure 3.1, particularly morbidity and demography, can be regarded as largely external to the elective care system particularly in the short run (in the long run some part of the increase in life expectancy may be attributable to more effective elective care). These work to top-up the pool, whatever is happening within the elective care system as outlined in Figure 2.1. Although there is no evidence that overall levels of morbidity are increasing across the population, demography alone can partly explain why more patients come forward (Harrison, 1997). Rates of treatment are higher than average for the very young and the old and very old. As the proportion of the population in the latter group has been rising, maintaining existing rates of utilisation implies an overall increase in activity.

Treatment rates within each age group have been steadily rising (Hamblin *et al.* 1998), reflecting the other factors shown in the figure. Of these, advances in medical technology have probably been the most important. Some of the most common and effective procedures such as joint replacement were unknown 20 years ago. As a result, the growth in clinical activity has not resulted in diminishing returns, but has been absorbed by new high-value procedures.

Figure 3.2 illustrates this in simple terms. The upper part of the diagram represents the scope for expanding activity offered by a larger budget; that expansion, however, yields lower benefit per pound spent – on the assumption that the most beneficial procedures have already been funded. The lower part represents the impact of new technology, which is assumed to offer scope for highly beneficial procedures. Funding these means that the access threshold remains the same and hence that a queue of people waiting for treatment remains despite the increased level of activity. The figure is of course highly simplified – not all new procedures are highly beneficial and there is no precise relationship between extra funding and the cost of medical innovation – but it is sufficient to bring out the nature of the process involved.

Figure 3.2: The impact of new technology



Other factors, particularly social and clinical thresholds, lie within the elective care system as we have defined it, and may therefore be influenced by the way it works. Thus, the generally successful attempts to reduce waiting times noted in the previous chapter have probably encouraged more people to enter the pool since the price they have to pay, in time rather than money terms, has been reduced. The reductions have also encouraged clinicians to refer people at earlier stages in the development of their condition or at lower levels of perceived or expected benefit from treatment. We consider in more detail how this works later in this chapter.

As the example of other countries shows, sufficient extra spending can overcome both the external and the internal forces that top-up the pool of potential patients. Other countries, with higher levels of spending than England, also have waiting lists but they are typically shorter. In Scotland, which spends about 23 per cent more per capita on the NHS than England (Dixon *et al.* 1999), waiting lists represent about two to three months' activity as opposed to three to four months' activity in England. Nevertheless Scotland's total numbers waiting have risen throughout the 1990s despite a substantial increase in activity (see Box 3.1). In Northern Ireland – also a high spender relative to England – numbers waiting represents about four months' elective activity (see Box 3.2).

Box 3.1: Scotland – trends in waiting times and waiting lists

	1991	1994	1998
In-patients & day case discharges during year			
Number of IP & DC discharges during year	886,862	1,056,464	1,193,724
of which number admitted from WL	363,711	448,805	487,567
Median wait (days)	30	31	33
% admitted within 3 months	81.7	80.7	80.5
% admitted within 12 months	98.2	97.8	99.2
% admitted 12 months or later	1.8	2.2	0.8

	1991	1994	1998	
			31 March	30 Sept
Patients on waiting list on 31 March				
Number waiting	79,866	82,065	89,525	84,931
% waiting under 3 months	55.1	61.7	62.6	62.9
% waiting 6 months and over	27.6	18.9	15.4	15.5
% waiting 12 months and over	14.0	5.1	1.2	1.3
Number on deferred waiting list	9,417	15,397	21,366	20,641

	1991	1994	1998
New out-patients referred during year			
Number of referrals			
of which referrals from GMPs/GDPs	–	1,241,699	1,371,443
Median wait (days)	–	1,007,845	1,110,314
% waiting within 9 weeks	–	38	41
% waiting 9–18 weeks	–	70.9	66.9
% waiting 18 weeks and over	–	21.9	23.3

Source: Scottish Health Statistics, 1998, Table M6.1.

Box 3.2: Northern Ireland – trends in waiting times and waiting lists (year ending 31 March)

	1995	1996	1997	1998	1999
In-patient & day case discharges during year					
No. of IP & DC discharges during year	309,079	323,765	320,322	327,187	340,447
of which no. admitted from WL*	118,911	126,922	124,037	128,728	131,404
Median wait (days)	32	31	28	29	29
admitted within 3 mths	88,316	95,434	94,813	95,292	95,199
admitted within 12 mths	111,719	122,131	119,530	120,888	122,765
admitted 12 mths or later	7192	4791	4507	7840	8639
*Patients on waiting list 31 March**					
Number waiting	35,015	36,377	45,769	47,005	44,868
% waiting under 3 mths	57.2	54.5	41.9	40.1	43.3
% waiting 6 mths and over	22.1	24.6	37.8	40.4	36.7
% waiting 12 mths and over	6.1	6.5	17.1	22.0	17.9
*Out-patients referred by a GP***					
who attended 1st appt	275,096	304,110	299,067	305,677	318,585
of which % seen within 3 mths	87.4	86.0	86.7	83.7	80.2
% seen within 3–6 mths	8.1	10.0	10.4	12.1	15.3
% seen after 6 mths or over	4.5	4.0	2.9	4.3	4.4

* includes waiting list and booked admissions (excludes planned admissions)
** does not include walk-ins or self-referrals

Source: Personal communication.

We have not been able to find data for countries outside the UK covering all elective procedures. Data for other countries usually refer to only a small range of procedures, such as CABGs or cataract. In Canada for example – which is also a high spender relative to England but which does not have comprehensive waiting list data at either national or provincial level – delays for CAGBs have attracted a great deal of public criticism (Naylor and Slaughter, 1994). In Sweden, also a high spender by comparison to the UK, there have been targeted programmes on a narrow range of conditions such as cataract where waiting times were considered to be too long (Hanning, 1998). The overall pattern of waiting is, however, unknown.

Although higher levels of spending might reduce the time spent waiting as well as the numbers waiting, the Scottish evidence suggests that the level of expenditure needed and the rate at which it would have to increase in order to keep up with the pressures outlined in Figure 3.1 are likely to exceed previously experienced rates of growth and spending. The Prime Minister's announcement of January 2000 suggests that spending increases will exceed historical levels, but even so, it will be years before Scottish levels are reached.

3.2 Why delay occurs

But even if waiting lists could be eliminated or reduced to negligible levels, that would not mean the elective care system was providing patients with immediate access to care or that the NHS would be providing an equitable service. The concept of waiting implies that a need has been recognised, but as we noted above, patients might put off approaching their GP, and the GP may not immediately recognise the need for treatment. Furthermore, considerable time may elapse within the hospital, i.e. at Stage 3 before the need for treatment is recognised or the right form of treatment identified. The time which elapses at these stages we term 'delay'.

In the absence of national data sources bearing on those aspects of delay that occur before patients are referred to hospital or which arise within Stage 3, we have to turn to studies of particular conditions that have attempted to track patients through all four stages of the care pathway. Here we take two examples: cataract and cancer.

Cataract is a progressive, non-life threatening condition but one that is nevertheless very significant in terms of the quality of life people can lead. A study of ophthalmology units carried out in the Northern Region in the early 1990s (Mordue *et al.* 1994) found that there was wide variation in patients' level of visual acuity when they were accepted for treatment – some were much worse than the level at which the consultants advised treatment but some were much better. Some who were offered treatment for their second eye had better visual acuity in it than some yet to receive treatment of their first eye. The study concluded that a standard method of describing and assessing visual handicap was required. The authors pointed out that these variations reflected differences in clinical practice within the units, but also referral patterns from general practice.

The importance of the latter was confirmed in a subsequent study of cataract surgery in Scotland by the National Audit Office (National Audit Office, 1997), which found that there were significant differences between GPs as to the stages of the disease which they thought appropriate to refer patients to hospital. In a significant proportion of cases, people were gaining access to hospital treatment at a very advanced stage of the disease, when they were nearly blind, even though surgery was considered clinically appropriate well before then.

The report found cataract treatment rates varied widely between the patients of different GPs and that the rate depended on:

> - *whether GPs actively search for cataract or respond to patient demand;*
> - *how active local opticians are in identifying cataracts;*
> - *the threshold for visual acuity at which GPs normally refer.*
>
> (pp.30–31)

The report goes on to point out that some GPs look for cataract during the annual health check offered to all patients over 75 years old. But not all do so and there is no standard format for the check.

A recent overview of access to cataract surgery (Gray *et al.* 1999) concluded:

> *Current means of assessing patients for cataract surgery do not provide enough information objectively to assess need and therefore priority. Much more emphasis must be placed on visual symptoms and how these influence social functioning and independence.* (p.1368)

The second example, cancer, is life-threatening, and the speed at which people begin to receive treatment may be important to their ultimate chances of survival. In the case of breast and cervical cancer, screening programmes exist in the hope that they will enable cancers to be detected at an early stage of their development. The present Government has introduced a two-week target – which came into effect in April 1999 – between the time a GP identifies a possible cancer and an out-patient appointment.

Only very limited evidence is available on the total care pathway and the factors explaining the time taken by patients to pass along it. As we noted in the previous chapter, national recording of waiting or delay takes places only at Stages 3 and 4. For information of the scale of delays at Stages 1 and 2, we have to turn to case studies. According to one systematic review of studies of patient and provider delays (Ramirez *et al.* 1999), most existing work is of poor quality and had done little to establish why delays occurred. However, as the following examples indicate, case results do suggest that delays at the early stages of access can be important.

A study of cancer patients in London (David, 1997) found that patient delay, i.e. at Stage 1, constituted the main source of delay, more than twice as much as system delay, i.e. delays within the NHS itself. Another study, in this case of oesophagastric cancer (Martin *et al.* 1997), found that total delay broke down as follows: 29 per cent at Stage 1, 23 per cent at Stage 2 and 48 per cent at Stages 3/4. Nearly one-third of total delay was attributed to the time spent establishing the diagnosis. The authors conclude that:

> *Much of the delay in diagnosis could ... be avoided if general practitioners referred patients promptly for investigation and a sense of urgency was imparted to the hospital's diagnostic process.* (p.469)

A study of delays in the treatment of lung cancer (Billing and Wells, 1996) found that the average delay from presentation to surgery was three to four months, made up as follows:

> *Of this period, there is on average one month spent before specialist referral and almost two months spent being investigated by a chest physician. This latter period of delay can be attributed as follows: two weeks waiting for an out-patient consultation and two weeks waiting for each investigation such as bronchoscopy, CT scanning, and percutaneous needle biopsy. The mean time from being referred to a thoracic surgeon to the date of pulmonary resection is then 3–4 weeks.* (p.905)

There is a limited amount of evidence bearing on the reasons for delay occurring at the diagnostic stage. A study of delays in the diagnosis of breast cancer in a New York Hospital (Wallis *et al.* 1998) found very wide variations in the delays experienced. In part these could be attributed to organisational factors within the hospital such as inconvenient clinic timings and the speed of feedback of results, which combined to produce extended diagnostic processes, and in part to primary care physicians referring patients with problems they should have been able to deal with themselves, and hence wasting hospital time.

This admittedly limited evidence suggests that significant delays, i.e. delays which affect the likely outcome of treatment or which leave patients untreated when they ought to have been treated, occur at all four stages of the care pathway. The variations found for cataract cannot be explained by reference to any lack of resources to carry out cataract surgery but rather by differences in patient behaviour and in clinical practice. In the case of cancer, the delays in reaching rapid diagnoses might be attributable to lack of resources, but it might also be attributable to poor management of the process resulting in the lack of urgency referred to above.

Thus the factors that explain delays early in the care pathway may be different from those which apply later, and the policies which might bear on them are also quite different. While delays at Stage 3 might well be reduced by extra resources being allocated to hospital diagnostic facilities, these are less likely to have a direct effect at Stages 1 and 2. As we noted, both individuals and professionals may be influenced by the existence of delays at Stages 3 and 4, but it seems implausible to explain the delays observed in these case studies in this way. If, as it seems, other factors are at work, then these factors need to be tackled directly, e.g. by making patients aware of the significance of the symptoms they experience, helping GPs assess their significance more quickly or taking proactive measures, such as screening programmes.

3.3 Why delay matters

Despite the enduring commitment to reduce waiting lists and waiting times, the amount of research into the costs that waiting and delay impose on patients, the

Service and the economy at large is tiny. It is, therefore not possible, on the basis of existing information, to provide a comprehensive account of what these costs are. Many of the studies cited are small-scale or focused on a particular procedure, or both. Furthermore, some of the costs concerned are inherently hard to estimate and the methods used open to question. What follows can, therefore, only be regarded as illustrative of the nature of the costs involved.

Clinical outcomes

The main clinical risks arising from delay are that the patient's conditions will deteriorate so that the chances of their effective recovery are reduced, or that they will die while waiting for treatment. Although there is evidence for a number of conditions that these risks are considerable, for other conditions they appear small.

In the case of patients suffering from severe heart disease, the risk of dying while waiting for a CABG has been put at 1 per cent per month of delay (Rachlin *et al.* 1991). Moreover, according to an editorial in *Heart* (Bridgewater, 1999), drawing on research in New Zealand (Seddon *et al.* 1999) and The Netherlands (Plump *et al.* 1999), the risks of any one person dying are heard to predict, so that the overall risk cannot be reduced by careful patient selection. It concluded that some 500 people a year died in the UK while waiting for treatment. In contrast, a Dutch study of patients waiting for PTCA (Koch *et al.* 1997) found that success rates fell after 12 weeks for type C lesions, but not for types A and B, while a Swedish study (Arbman *et al.* 1996) found that a short delay at diagnostic stage was more important for rectal than colonic cancer. In the case of breast cancer in Denmark (Chistensen *et al.* 1997) it was found that both patient delay (Stage 1) and doctor delay (all other stages) 'had significant independent prognostic value' (p.350).

In some instances, some delay may be necessary to maximise the benefit from treatment. Co-morbidity, for example, may jeopardise the likelihood of success from an intervention – it may be ill-advised to undertake a hip replacement on an elderly patient with high blood pressure, for example. And in others, a period of waiting may do little or no harm to the prospects of success at the treatment stage. For example, a study of larynx cancer (Barton *et al.* 1997) found that longer waiting times were not a significant predictor of relapse after treatment.

Recent review papers on delay and breast cancer, however, have revealed not only the poor quality of much of the data used in many studies of the impact of delay but also the complexity of the relationships involved and the presence of confounding factors. Although a systematic review of the literature (Richards *et al.* 1999) concluded that delays of three to six months were associated with lower survival, a study based on the Yorkshire Cancer Registry found that delays longer than 60 days did not impair survival chances (Sainsbury *et al.* 1999). This latter finding casts doubt, as the authors point out, on the Government's drive to ensure that all women referred for a hospital appointment are seen within two weeks. Moreover, a further review (Ramirez *et al.* 1999) of a large number of studies of delays and survival found that most were of poor quality and had to be disregarded. It concluded that existing evidence was insufficient to formulate specific strategies for shortening patient or provider delays. An earlier review (Faccione, 1993) also concluded that provider delay was under-researched, while Caplan *et al.* (1992) also concluded that 'little research has been done on system delay' and much of that was of poor quality. Similarly, a review of studies relating to oral cancer (Allison *et al.* 1998) 'demonstrated the weak nature of the available evidence' relating to the link between delay and outcome.

We have not found similar reviews for other conditions. However, a small-scale study of patients waiting for varicose vein surgery (Sarin *et al.* 1993), found considerable deterioration on their condition while waiting for surgery – more limbs required surgical treatment or deterioration showed within the limbs originally identified as requiring treatment.

Clinical costs

Over and above clinical outcomes from treatment, delay may also give rise to other clinical costs if deferring treatment means that its ultimate cost is greater and if patients waiting for admission require care while they are waiting.

A study of hip fracture repair in Canada (Hamilton *et al.* 1996) found that delays led to greater pre-surgery costs and more patient discomfort – although it did not identify any impact on post-surgery outcomes. Hip fracture represents an emergency and is typically treated rapidly. Another study of hip repair (Saleh *et al.* 1997) also concluded that, over a range of assumptions, the cost of immediate surgery was lower than deferred surgery, in part because delay reduced the chances of a successful outcome and increased the chance of complications arising and in part because of the home care costs incurred while patients were waiting for surgery. We have not been able to find similar studies of conditions such as joint replacement where treatment is typically delayed for long periods and which may give rise to the need for continuing care.

The individual's quality of life

From the individual's viewpoint there are three considerations: pain or discomfort; limits on activity arising from their condition during their time waiting; and the financial and other costs of waiting.

Pain or discomfort

Despite the apparent inevitability of waiting for elective care from the NHS, there have been only a few studies of what patients considered to be acceptable waits. Research carried out for the Royal Commission (Royal Commission, 1978) found that a large proportion of those interviewed did not find waits distressing or inconvenient but a substantial minority did.

As Table 3.1 shows, 61 per cent of people did not consider that waiting for admission for more than six months caused them distress or inconvenience. Over half did not mind waiting for over six weeks for an out-patient appointment: see Table 3.2. However, significant minorities did find waiting distressing or inconvenient.

Table 3.1: Proportion of patients caused distress or inconvenience by waiting for hospital admission

Whether waiting for admission caused distress or inconvenience	**Time between first being told and being admitted**					**All**
	less than 2 weeks	2 weeks – less 1 month	1 month – less 3 months	3 months – less 6 months	6 months or longer	
	%	%	%	%	%	%
wait caused distress/ inconvenience	9	12	19	31	39	20
wait did not cause distress	91	88	81	69	61	80
Base: all who had a bed booked for first admission = 100%	75	57	64	59	41	301

Source: Royal Commission, 1978.

Table 3.2: Proportion of patients who minded waiting for their first out-patient appointment

Whether minded waiting for first appointment	**Time waited for first out-patient appointment**					**All**
	8–14 days	15–21 days	22–28 days	over 4–6 weeks	over 6 weeks	
	%	%	%	%	%	%
Patient minded waiting	7	11	20	22	44	21
Wait didn't really matter	93	89	80	78	56	79
Base = 100%	147	130	106	102	157	642

Source: Royal Commission, 1978.

In October 1999, the results of the first National NHS Patients Survey became available (NHS Executive, 1999) – the first such survey since that by the Royal Commission. This revealed that a considerable proportion of those on waiting lists perceived themselves as being in pain and that their condition had worsened while waiting.

Table 3.3: Whether the condition got worse between being referred and being seen by the specialist or hospital doctor by age and sex

Male	**Age**						
	18–24	*25–34*	*35–44*	*45–54*	*55–64*	*65–74*	*75+*
	%	*%*	*%*	*%*	*%*	*%*	*%*
Did the condition get worse?							
It got a little worse	34	25	28	27	30	29	29
It got a lot worse	9	8	9	8	8	6	4
It stayed the same/got better	49	56	51	57	52	51	53
Not sure	9	10	11	8	10	15	15
Female	**Age**						
	18–24	*25–34*	*35–44*	*45–54*	*55–64*	*65–74*	*75+*
	%	*%*	*%*	*%*	*%*	*%*	*%*
Did the condition get worse?							
It got a little worse	31	23	23	29	30	29	31
It got a lot worse	7	7	8	7	8	8	8
It stayed the same/got better	50	55	58	53	53	51	45
Not sure	12	15	11	10	9	12	16

Source: NHS Executive. *National surveys of NHS patients: General practice 1998*. London, 1999.

Table 3.4: Whether the condition caused pain between being referred and being seen by the specialist or hospital doctor by age and sex

Male	**Age**						
	18–24	*25–34*	*35–44*	*45–54*	*55–64*	*65–74*	*75+*
	%	*%*	*%*	*%*	*%*	*%*	*%*
Did the condition cause pain?							
A lot of pain	13	15	16	14	16	10	9
Some pain	46	45	45	42	40	38	38
Not in any pain	38	38	37	42	43	50	49
Can't remember	3	3	2	2	1	2	4
Female	**Age**						
	18–24	*25–34*	*35–44*	*45–54*	*55–64*	*65–74*	*75+*
	%	*%*	*%*	*%*	*%*	*%*	*%*
Did the condition cause pain?							
A lot of pain	13	12	13	16	17	17	17
Some pain	36	33	36	39	43	40	39
Not in any pain	48	53	48	43	38	41	40
Can't remember	3	2	2	2	2	3	4

Source: *National surveys of NHS patients: General practice 1998*. London, 1999.

Other studies reveal a tolerance for waiting, albeit for periods shorter than is typical in the UK. A study of cataract surgery (Hadjistavropoulos HD *et al.* 1998) found that few patients were concerned with the waits they experienced for cataract (average 10 weeks). Most had developed a series of coping strategies to deal with the difficulties they did experience. A cross-national study (Dunmetal *et al.* 1997) of Canada, Spain and Denmark found that most people were not prepared to (hypothetically) pay more in charges or taxation to reduce the time they waited.

People's concerns about waiting may relate to how long they expect to wait. A Canadian study of people waiting for CABG (Llewellyn-Thomas *et al.* 1999) found that people were reluctant to wait for more than six months even if they were told that this increased their chances of a successful outcome. However, their attitudes seemed to be determined by attachment to their scheduled dates and, possibly, an exaggerated view of the risks of delay.

Although these findings suggest that (reasonable) delays may be acceptable to patients, a number of studies have reported that those on waiting lists had a poorer quality of life than the general population and that those with more severe symptoms were, naturally enough, more concerned about length of waits. For example, a study of New Zealand patients (Derrett *et al.* 1999) waiting for prostatectomy, or hip or knee joint replacement found that they experienced 'significantly poorer health-related quality of life that the general population', and that some reported 'severe symptoms'. However, neither appeared to grow worse with time spent waiting. A study of people waiting for CABGs in Iceland (Jónsdóttir and Baldursdóttir, 1998) found that 86.8 per cent were dissatisfied with their health status and 39.1 per cent described their health as bad or very poor. A study of patients waiting for arthritis surgery (Roy and Hunter, 1996) found that nearly all were in pain and suffered from a wide range of disabilities. Other studies are cited in McDonald *et al.* (1998).

Limits on productive activity

We have found no estimates of the scale of this, but it has been reported that 5–10 per cent of people on waiting lists are on sick leave from work, and the Icelandic study cited above found that the unemployment rate rose from 12.5 per cent at the time of diagnosis to 44.4 per cent on treatment (the mean waiting time was five to six months). A similar finding emerged from a Finnish study (Kontinnen and Merikallio, 1990), which found that working ability declined sharply before surgery. A Norwegian study (Lundborn, 1992) found that post-operative return to work was associated with waiting time for treatment, i.e. the longer the latter, the slower the former.

The 'costs' of waiting in their own right

In general, people prefer to have beneficial things now rather than in the future – there is what economists term a 'positive time preference'. Given the choice, we would rather have a holiday now than wait until next year, other things being equal. We may fear that something might happen to us in the intervening period that prevents us from enjoying the beneficial thing, but the mere fact of having to wait for something we want also represents a cost. As far as we are aware, no estimates have been made of the scale of pure time preference but Propper (1990) has estimated the disutility of having to wait for treatment. The scale of these costs was put at £40 per month.

Box 3.3: Time preference

Consider an elderly person who needs a hip replaced. Let us assume that the actual marginal health gain (change in health status) for the individual is the same whether the hip is replaced now or in six months; furthermore, the patient's circumstances are not going to change such that the operation will be more relevant to a future period of time (there will be no greater need for mobility predicted, for example). Assume also that the period of time for which the hip will give benefit is expected to be ten years (before it wears out). Thus, if we ignore the particular point in time when the hip is replaced, the overall health benefit to the patient is the same whether the operation is done now or in six months' time. However, by delaying the operation, we are introducing an element of uncertainty – the patient might die before the operation or something else might happen – an accident, or another illness – which prevents the operation from taking place.

Because of all these possible eventualities, we prefer to have something that will benefit us now rather than in the future, *ceteris paribus*. This is 'pure' time preference and should not be confused with a simple denial of the benefit for the first six months. It is also possible that our tastes will change in the future, diminishing the value of the intervention. And, there may also be other, less rational reasons for having pure time preference such as myopia. The rate of pure time preference is normally taken as being fairly low such that a delay of a year only reduces value by some 3 or so per cent. But for some people the rate may be much higher just as the subjective costs of delay will vary from one person to another. Some have argued, moreover, that health benefits are valued differently from the financial benefits to which discounting is usually applied and that usually the 'devaluing' effect of discounting does not apply in the same way.

See Loewenstein, 1988; Redelmeier and Heller, 1993; Gafni, 1995; Mackeigan *et al.*, 1993; and Gafni and Torrance, 1984 for further discussion.

Very closely linked to time preference is confidence on the part of the patient about when they are likely to get treated. In other words, *knowledge* of how long a wait may be is as important as the length of that wait: not only does it provide reassurance that treatment will in fact be available, it also allows plans and commitments to be made. That such knowledge has a value is the assumption underlying the Government's initiative in relation to booked admissions. It appears to be confirmed by the NZ study cited above, which found that because of uncertainty about when they would receive treatment people were reluctant to take holidays or make other commitments. Propper (1990) estimated the costs of the uncertainty attached to not having a booked admission date as £20 – the only such estimate we have identified.

Overall

Despite the limited evidence, it seems reasonable to conclude that the costs – in all senses – vary a great deal from one condition to another and between different people. However, there can be no assurance that the various categories of cost vary in the same way – the personal costs of delay for some conditions may be high, but not the clinical costs (i.e. the prospect of poorer outcomes).

Hence the main conclusion to emerge from this section is that the costs of delay – clinical, financial and personal – have not been systematically examined across even the main types of elective treatment for all the potentially relevant categories of care. As Naylor and Slaughter have pointed out (1994), 'we need to know more about the hidden burdens of delayed treatment' (p.233).

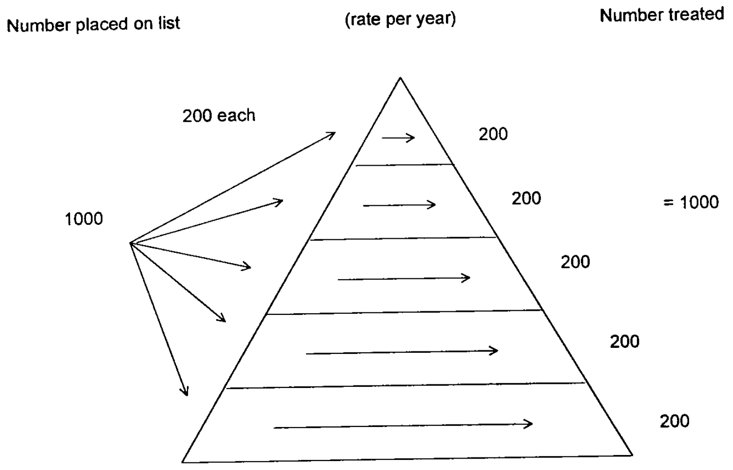
3.4 How the system as a whole functions

We have set out above the individual access pathway for a single person who receives in-patient hospital treatment and have considered the impact of delay within it, focusing on the individual and the cost of treating them. In this section we broaden the focus to consider the significance of delay to the elective care system as a whole. Our central argument is that delay – however undesirable from the patient's viewpoint – is crucial to the working of the elective care system as a whole. It is also important for a proper understanding of the possibilities for reform.

Waiting in the NHS does not in general operate on a first-come, first-served basis. At the GP's surgery, people are generally treated in the order they make appointments, although even here there are emergency surgeries and the booking process may prioritise those who ring for an appointment according to some rough and ready assessment of the person's situation. Similarly at the out-patient stage: the GP will make some indication to the consultant about how quickly someone should be seen for first appointment, and this will depend on factors other than the time at which they were referred. The consultant may – very rarely – decide to make an immediate, emergency admission, or, more commonly, an urgent one, i.e. within a few days. If the need for treatment is not considered urgent, the patient will be placed on a waiting list with or without a specified date.

Thus patients enter the elective care system 'unsorted', only having a limited conception of the severity of their condition. The next two stages 'sort' their demands in terms of the broad area of care they need – the specialty – and the urgency with which they need it, as assessed by health professionals according to various, largely implicit, criteria. In these ways, an unmanaged, uncontrolled demand becomes a rather more managed, planned demand for care.

Figure 3.3: The elective care system – inflow and outflow



One way of conceiving the process is to think of patients on a waiting list constituting a 'lake' represented by the triangle in Figure 3.2. In this example, 1000 patients are treated and 1000 are placed on the list each year, so the list is in equilibrium and not growing (the size of the triangular 'lake' is constant). Each of the 'rungs' in the triangle represents an implicit threshold of priority – the higher the rung the higher the priority of the patients on the list – and the more quickly they are treated.

A waiting list is therefore in fact a number of 'lists', each with patients moving through at different rates. There is no reason why it should be five as in the diagram (in fact there may be an indeterminate number as few consultants choose to organise their lists explicitly in this way). The central point, however, is that the consultant chooses to organise progress through the 'lake' according to his or her assessment of priority and not (or only partly) on the basis of how long patients have been waiting. This is why the pattern of waits is as described in the previous chapter: if consultants did not prioritise, everyone would wait more or less the same length of time. In practice, very long and very short waits co-exist.

It should be clear from this schematic representation that there are more patients waiting on the list at any one time in the lower categories of priority than in the higher. Consequently, the time it takes for a patient to move through the list is longer for lower priority patients. This may seem odd on reflection: if the system is in equilibrium, why should the lower priority patients have to wait longer than the higher priority ones? In fact, why should anyone wait at all? Each patient who joins such a list is matched by one who is treated. Why not treat them just as they arrive?

Part of the answer is organisational, and relates to the irregularity at which people in fact join the list – demand fluctuates from one time of year to another. In other words, the '1000 per year' figure is an average; if the 'lake' of available patients did not exist then a sudden drop in demand would leave providers with no people to treat. The 'lake' in other words is a means of dealing with uncertainty.

Another source of uncertainty is that because of the way that patients identify themselves or are identified by their GPs as likely to need treatment, they are often at a late stage in their disease or condition and have to be treated quickly to improve their chances of survival or recovery. They, therefore, have to leapfrog over others. This phenomenon may also be to some degree predictable and hence 'slots' can be left open to deal with such 'latecomers' but, to the extent that it is not, having a pool that can be called on at short notice is clearly helpful from a managerial point of view.

But the existence of this pool of 'waiters' is not simply an organisational device. As we suggested earlier in this chapter, the length of the list – or rather the time spent on it – appears to act as a deterrent to joining it, and shortening it encourages others to join. Thus if capacity was increased to 1200 cases treated per year the size of the triangle would start to shrink, other things being equal. However, GPs and patients would reassess the costs and benefits of (reduced) waiting (times), the deterrent effect would be lessened and referrals would be increased. The triangular lake would increase in size (by, say, adding another rung/threshold of lower priority) until the average wait was back to its original level, deterring just enough to balance supply with demand.

We use the phrase 'appears to act' because only a limited amount of evidence exists as to the scale of the deterrence effect and how precisely it works. Some may be diverted rather than deferred.

Patients may choose the 'go private' option, perhaps use alternative medicine or live with the condition for the time being. Indeed, the knowledge that many people are waiting for treatment may act to qualify people's expectations about when they are 'ill enough' to go to the GP in the first place: see Hamblin *et al.* (1998) for further discussion.

Such evidence as does exist includes a study of the impact of the Conservatives' Waiting List/Time Initiative (Newton *et al.* 1995). This found that although an increase in admissions improved waiting times, it did not reduce list size because additions to the list tended to increase at the same time: admission from the list and additions to it were closely correlated. This study did not attempt to show why these additions were made, but given the speed of the response of additions to admissions, it would seem that short-term factors internal to the elective care system were responsible. Earwicker *et al.* (1998) found evidence, in a study of GP referrals based on realistic case notes of hypothetical individuals, that these were influenced by waiting times.

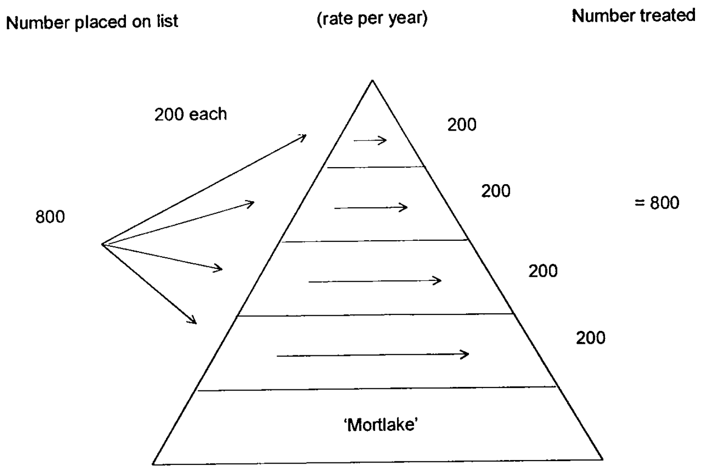
Other studies have focused specifically on the impact of waiting. Goddard and Tavakoli (1998) found that referral rates and waiting times were negatively associated, suggesting, as they point out, that waiting time guarantees may induce more referrals and hence longer lists. A subsequent study by Martin and Smith (1999) concluded that the demand effect of shorter waiting time was small and hence if extra resources reduced lists the volume of extra demand coming forward as a result would be modest.

Using a different methodology, Ackere and Smith (1997) found that waiting time did influence additions to the list but in varying degrees. They found very little effect for waits up to about three months but after that point, up to a wait of about five to six months, demand was found to be responsive. After that it became unresponsive again. The unresponsive section at the low end of the curve may indicate that people are quite content to wait up to three months; the unresponsive section at the high end may arise because the GP or consultant may feel that the condition is insufficiently serious to justify such a long wait and that other forms of treatment should be used.

The results reported here suggest that there is a deterrent effect, but not necessarily a large one. If so, then the continuing increase in numbers waiting alongside increases in elective activity must be largely explained by the external forces described above in Figure 3.1, with waiting lists acting as a 'receptacle' for increases in demand, i.e. a pure delay function. We are not aware of any work that has attempted to analyse the behaviour of waiting lists over long time periods in a way that would allow an estimate of the relative importance of the different factors at work. In what follows we assume that both external and internal factors need to be taken into account. As noted earlier, elective care covers a very wide span and it is only to be expected that the significance of delay will vary from one condition to another – as well as between individuals. None of the work cited above, however, attempts to distinguish between the different queues making up the total waiting list.

There is another twist to this tale. Assume now that resources only allow for 800 cases to be treated per year and assume also that the level of deterrence necessary to achieve equilibrium (i.e. 800 cases *added and treated* per year) is still associated with an average wait determined by a list size as large as in the previous example (this average wait is now higher than before because the same number of people on the list are being treated at a slower rate). The situation may now be something like Figure 3.4:

Figure 3.4: The elective care system – how a 'mortlake' develops



There is now a group whose priority is never high enough to get treated (if the implicit thresholds are applied consistently) but whose existence is necessary to sustain a sufficiently high deterrence effect to determine equilibrium, i.e. number passing through equal the capacity of the system. Frankel and West (1993) described these patients as occupying a 'mortlake'– that portion of a river that does not flow with the rest of the water, but forms a still 'lake' outside the normal current, usually at a bend. These patients have also been described as occupying 'never-never lists'. It may be that they die for reasons related or unrelated to waiting, become dispirited and take private surgical or alternative treatment, the condition goes away or the patient decides to live with it, or they leave the area or country (factors that apply to all those waiting to some degree). Many of those weeded out when lists are checked (as they are when governments set targets for list reduction) probably fall into these categories, thereby allowing rapid list reduction without extra surgical activity.

We do not know how many 'mortlake' waiters there are, or have been in the past, on NHS lists. In evidence to the Health Committee in 1991, the BMA (House of Commons, 1991) argued that exclusions from the purchasing bundle (which a number of authorities were then considering):

> *would have no effect upon the conservation of resources because few if any of these disorders would have reached the top of the waiting list.* (p.x)

However, Government initiatives – such as the maximum 18-month wait introduced as part of the *Patient's Charter* – impose an 'artificial' constraint on the length of time that people can wait in the lower reaches of the triangle, and hence force those who would be in the mortlake ahead of those who would otherwise command greater priority.

We noted in the previous chapter that there were signs of this effect in the way that the proportions waiting in different time bands has changed over the last two years. This suggests that the current emphasis on eliminating maximum waits is distorting priorities. But the potential importance of that effect depends critically on the way that patients are in practice allocated to the various 'streams' in the triangle. There is, however, very little information about the way that patients are sorted in this way. Although it is clear enough that consultants do make conscious decisions on degrees of urgency, how well actual waits correlate with urgency is not known across the elective care system as a whole. A number of specific conditions have been studied and, not surprisingly given the lack of agreed methods of prioritisation, considerable disparities between the actual order in which patients were treated and external assessment of their priority have been identified.

A Canadian study of people waiting for joint replacement (Williams, 1997) found that time spent waiting was unrelated to the severity of pain they experienced or their loss of function: '... for the vast majority of patients, the intra- and inter-surgeon variability in queue length and queuing practices leads to a situation where waiting times cannot be correlated with severity of pain, disability, co-morbid conditions or any other clinical factors' (p.67). This study implies that, except at the extremes of urgency, priorities may be 'distorted' already, i.e. are effectively random for some if not most patient groups.

Similar results have been obtained for serious heart disease where it might be expected that prioritisation would be more systematic, given the risks of death while waiting already noted. A study of waiting lists for CABGs carried out in three London hospitals (Langham *et al.* 1997) found that while 38 per cent were treated later than they ought to have been (using an adapted version of a Canadian urgency scoring system), 34 per cent were treated earlier.

Overview

Figures 3.3 and 3.4 and the subsequent discussion do, we believe, represent a reasonable interpretation of how aggregate waiting lists work within the English NHS. Within this interpretation, there is scope for variations between areas, between specialties and even between procedures as to the degree of user and professional responses to changes in the numbers waiting and the time they do so, as well as in the rate at which medical innovation occurs. Indeed the scale of the variations in list length and activity rates between areas would suggest that each local health economy equilibriates in its own way. But unless some mechanisms of the kind set out here are generally at work, the observed behaviour of the aggregate waiting list, as well as some of the research studies cited here, would be hard to explain: see Hamblin *et al.* (1998) for an alternative but compatible account.

What we have not been able to establish is the relative importance of the factors external to the elective care system and those internal to it. However, as we shall see in Chapter 5, most of our policy recommendations are not very sensitive to the precise balance between these factors.

3.5 Why there are geographical variations in speed of access

In Figure 3.3, we again set out the stages by which access to elective care is achieved. As we have noted, in reality the elective care system consists of a large number of largely independent mini-systems, each operating generally within its own set of rules. In such a system, it would be highly unlikely that patients with similar conditions enjoyed the same speed of access since the system as a whole does not operate to common rules and there are no mechanisms such as 'trading' between areas that are effective enough to even-out the differences that emerge from one area to another.

However, as the figure shows, many other factors are involved, over and above patient selection by consultants. On the right-hand side of the figure are a series of decisions – the first by potential users and the rest by professionals – that determine whether, or how fast, people move through each stage. Some of these factors are speculative: they are criteria that clinicians *might* take into account, although there is little hard evidence about their practical use or at which stage they tend to come into play. For example, there is no evidence that consultants consider the cost of individual cases when they make a decision on whether to admit for treatment, or whether to place on a waiting list. But they could do so, and indeed they would have to if their professed aim were to maximise the total health gain they could achieve from their budgets. Similarly, there is only anecdotal evidence that many of the non-clinical factors are taken into account, although many are plausible – particularly when other aspects of two individuals' circumstances appear very similar. What evidence there is suggests these factors are considered at the final in-patient admission stage – but there is no reason why GPs should not also take them into account when deciding whether or not to refer.

Figure 3.5: Determinants of access

Condition 'X' exists in two regions A and B

Factors affecting variations between districts A and B[*] (relating to horizontal equity)	**Access 'pathway'** (decision-points in boxes)	**Decision criteria & factors affecting relative speed of individuals' access** (relating to vertical equity)
• socio-demography • physical/biological environment (affecting degree of ill health)	Causes of ill health 'population' of illness	
• as above (including cultural factors) • number of GPs • access costs (population density & GP density)	Perceive illness and decide to seek medical advice	• perceived severity • access costs • cultural 'inhibitors'
	waiting: to visit GP	
• degree of ill health requiring specialist care • hospital provision and efficiency (both affecting availability of secondary care) • local 'choice'	GP consultation: decision about whether to refer	• extent of ill health • ability to benefit • non-clinical factors? Cost?
	waiting: for out-patient consultation	
• as above	Consultant diagnosis: decision about whether treatment is necessary	• extent of ill health • ability to benefit • non-clinical factors? Cost?
	waiting: on in-patient or day case admission list	
• as above	Consultant: decision about when to admit for treatment	• extent of ill health • ability to benefit • cost • age, dependants, in employment, time waiting, lifestyle
	treatment: health outcome	

---

[*] In addition, the priority given to the various decision criteria on the right hand side of the figure can itself vary from region to region, thus constituting another source of observed difference between the access speed of two otherwise identical individuals.

The left-hand side outlines a series of factors relating to the local geography, social and economic characteristics of the local population and the location and nature of local health facilities. These influence the aggregate pattern of demand for care at all the stages. So, for example, the local socio-demography of a region will affect the level of morbidity in that region; the number and distribution of GPs will affect how likely people are to seek advice; and the level and efficiency of secondary care provision will affect the length of waiting lists, and thus the likelihood of GPs considering it worthwhile to refer on for specialist treatment. Indeed, a consultant's own efficiency at managing the throughput of his or her waiting list – and thus the average wait – may well affect his or her *own* decision about whether to place a patient on the list in the first place. In addition, local choice will have an effect, in that some health authorities will simply decide that one client group deserves a higher priority than that assigned to others; this 'choice' factor is related to the note in the figure about differing individual priority decisions. Thus the aggregate throughput will vary significantly from region to region, as will the pattern and case-mix of those who reach each of the stages. This is entirely to be expected; indeed, it is unavoidable given the range of local factors potentially giving rise to variations. We consider each briefly.

Resources

Over the years, through changes in financial allocations and a series of measures bearing on the nature and quality of the services available in different parts of the country, successive governments have attempted to remove differences on the left-hand side. Nevertheless differences remain. In the case of GPs, for example, there remain, despite sustained attempts to even-out their availability, significant shortfalls and surpluses in some parts of the country. Exactly how large and exactly which areas are perceived as being under or over-resourced depends critically on how relative needs are estimated. According to a study by Maynard and Bloor (1995), a fair distribution of the existing resources within general practice would require substantial changes relative to the current position: see Table 3.5.

Table 3.5: Changes in finance required to equalise primary care resources

Region	% difference
Northern	-13.31
Yorkshire	-4.47
Trent	-7.71
E Anglian	7.29
NW Thames	15.11
NE Thames	5.21
SE Thames	1.33
SW Thames	5.53
Wessex	6.07
Oxford	8.98
S Western	15.34
W Midlands	-5.94
Mersey	-11.28
N Western	-18.26

Source: Maynard and Bloor, 1995.

Local policy

The financial resources allocated at health authority level are not tied to elective care and consequently there is scope for district health authorities (or GP purchasers) to make their own decisions as to the volume of elective activity they purchase. The differences in levels of access set out above may reflect different priorities in different areas. The CSAG report on specialist services (CSAG, 1996b), for example pointed out that in part the existence of variations could be ascribed to deliberate policy; different authorities took different views as to priorities:

> *Patients may be treated differently depending on where they live as a result of purchasers negotiating different service standards. This is perhaps an inevitable consequence of the internal market where purchasing, and hence resource allocation, decisions are taken at local level. HAs and fundholding GPs have to make difficult choices and inevitably they will not all decide to invest equal amounts in the same clinical areas. However, to many observers, the idea that similar patients will be treated differently depending on where they live, or the characteristics of their GP, is unacceptable given that individual contributions to the system are not dependent on these characteristics.* (p.46)

The evidence cited in Chapter 2 supports the existence of 'policy' differences. However, such 'policy' differences may not reflect any explicit decisions to favour one type of activity over another. The Group's report also suggested that variations reflected a deep-rooted bias in the way that the NHS functions as much as a set of consciously determined priorities:

> *There remains a tendency for the least needy groups to receive the most treatment. The 'inverse care law' has been described in other health care systems and was known to operate in the NHS prior to the reforms (Hart, 1971). The issue here is not that the [1990] reforms have caused a problem, but that they have failed to tackle a problem which should be amenable to purchasing based on needs assessment.* (p.46)

Patterns of provision

There remain significant differences in the way that services are provided in different parts of the country, affecting both the way that patients gain access to treatment and the outcome of that treatment. These may arise from the performance of individual clinicians to the overall design of the service they offer.

At the individual level, the Audit Commission has shown (Audit Commission, 1997) that the workload of individual surgeons varies a great deal, even after training and other non-clinical claims on time are allowed for. It has also shown (Audit Commission, 1990) that the proportion of elective care carried out by day surgery varies considerably, although overall it has risen rapidly since the Commission first examined the potential for developing it (Audit Commission, 1998). Against this background it is only to be expected that the efficiency of different hospitals will vary widely.

At a broader level, studies of particular services have found wide variations in the way they are provided. In the case of cleft lip and/or cleft palate, CSAG concluded (1998) that the existing pattern of service organisation was poor and went a long way towards explaining why results were well below the level of those obtained in other countries. Only two out of 17 centres visited had the range of experience and facilities that would allow them to be ranked among the best in Europe: in 1999 the Government took steps to reorganise these services.

The Group found that, for this particular service, the NHS had not succeeded in developing the most effective degree of specialisation and, as a result, the quality of service on offer came up to the best standards in only some parts of the country. The same conclusion underpinned the Calman-Hine report on cancer services (Expert Advisory Group, 1995) and the London Implementation Group reports on the structure of specialist services in London (London Implementation Group, 1993), and now forms the basis of the Government's programme of work on national service frameworks.

Precisely what is the proper degree of specialisation in elective care remains a contentious issue. A review of all the available evidence by the NHS Centre for Reviews and Dissemination (Ferguson *et al.* 1998) concluded as follows:

> *Overall, the literature on links between volume of activity and clinical outcomes suggests that for some procedures or specialties there may be some quality gains as hospital or clinician volume increases. In other areas the research suggests an absence of significant volume gains.* (p.19)

It follows that as far as the scale of the individual hospital is concerned, for many procedures the current pattern of provision is generally adequate or, more precisely, would not be improved simply by increasing the level of activity in each.

Access

A related consideration is physical access. Put simply, the greater the degree of specialisation, the greater the access costs – in terms of travel and time costs to the individual – are likely to be. If access costs deter use, then achieving higher quality in general may disadvantage some sections of the population. Moreover, it is inevitable that people in sparsely populated areas have further to travel than those in urban areas.

The available evidence bearing on the impact of access costs is very limited and not always mutually consistent. Summing up all the available evidence, the NHS Centre for Reviews and Dissemination study found in respect of in-patient care:

> *There is no evidence of a general relationship between accessibility and the utilisation of acute hospital services, although there is evidence in the literature of a negative association for some specific services (such as CABG and other discretionary surgery, treatment for colorectal cancer and coronary revascularisation). The work on the determinants of utilisation in England and Northern Ireland which suggests a positive effect of distance on the utilisation of non-acute beds and elective surgery is intriguing, if not altogether intuitive.* (p.46)

Personal behaviour

With the exception of a small number of preventive programmes (e.g. cancer screening) and a small amount of health promotion, individuals make the decision about whether to seek care according to their own views and experience. A number of conceptual frameworks have been devised to 'describe' how these decisions are taken. They identify a number of reasons why decisions to utilise the health care system may differ, some of which may not be systematically related to ethnic or socio-economic groups (Faccione, 1993). The King's College study of cancer patients (David, 1997) found these conventional groupings not to be significant and other studies of delay in seeking treatment have also identified personal rather than social factors as being significant. Whether these vary in different ways in different parts of the country is unknown.

Clinical behaviour

In large measure, how decisions are taken through each stage have been left to personal or professional discretion, and where research has been carried out it has often been inconclusive. General practitioners' referral rates differ widely, reflecting variations in training and professional judgement. As the National Audit Office study of cataract study (NAO, 1997) showed, the latter varies widely even where the indications for referral are relatively straightforward.

The King's Fund studies (Ham, 1988; Sanders, 1989) cite a large body of work in this country and elsewhere which reveals significant variations in treatment rates for specific conditions, most of which are attributed to professional uncertainty as to the appropriate indications for treatment.

Variations between social groups

A review by Goddard and Smith (1998) found some evidence of bias against (and some in favour of) women, ethnic minorities and older people. Overall, they found it impossible to prove conclusively whether any population group was systematically discriminated against, largely because there were no reliable measures of variations in need between the groups they examined. Analysis of Scottish data (Pettinger, 1999) revealed that those in the higher deprivation categories received less elective care per head than those other groups, despite the fact that their morbidity could be expected to be higher. As the author points out, it may be that the former receive more emergency treatment, but whether that compensates for the lower elective rate is an open question. A study in Cheshire (Johnstone *et al.* 1996) found that in some specialties treatment rates were higher in more deprived areas, but in others – cardiology and cardiothoric surgery – they were lower. There is evidence for some conditions that women are less likely than men to be considered eligible and that older people in general are less likely to be offered treatment (Age Concern, 1998).

3.6 Conclusion

Elective care waiting lists are a mechanism for rationing health care. This mechanism is not centrally run, and yet serves to assist in the crucial function of managing the pressures made on limited NHS resources. Waiting lists have typically been viewed as a form of 'rationing by delay'; this analysis has suggested that their influence extends well beyond that to rationing by deterrence, deflection and, in some cases, *ipso facto* denial.

We do not infer from this that waiting is to be avoided at all costs. We argue below that trying to remove waiting lists from the NHS by simply increasing expenditure is almost certainly not a sensible use of the nation's resources. For now, though, we simply suggest that the current system for gaining access to elective care performs this important task in ways that are not commonly recognised, but which must be if reform is to be successful.

This chapter has also shown that the elective care system is poorly understood. This is true at a simple descriptive level and also with respect to its underlying dynamics. In the past, the expectation may have been that delays would be eliminated and hence to carry out research on the costs of delay would be a waste of time. But to hold that view now would represent a triumph of hope over experience. We therefore make a number of suggestions in Chapter 5 for improving knowledge and understanding of how the elective care system works.

This section has also shown that a large number of factors bear on the way that people gain access to the elective care system. Despite the continuing efforts made to put all parts of the country on the same financial footing, the way that resources are used continues to be typified by a large degree of local clinical discretion, which in itself suggests that there is no reason to expect that similar people with similar conditions in different parts of the country gain similar levels of access with the same level of delay.

Thus if there were to be equal access for equal need, then a number of conditions must be satisfied which would in effect bring the various local elective care systems into line with each other. First, national agreement would be necessary about what counts as 'need', and in which circumstances; second, similar agreement would be required on how different need is prioritised; and third, the objective circumstances – the level of provision and the efficiency with which it functions – would have to be the same in each region relative to the (agreed) level of need. Furthermore, decisions on the definition of need must, in principle at least, be taken centrally if patients with similar degrees of need are to be treated with the same level of priority. This may, of course, be a practical impossibility and may offend other objectives of the system as a whole. We discuss this further in the next chapter.

Despite the introduction in 1991 of an explicitly identified purchasing function, there has been little systematic investigation of the reasons why variations occur even where, as with cataract, there is a large measure of professional agreement about the condition and its treatment. This in turn reflects the fact that despite decades of effort, the elective care system has not been managed either centrally or locally. Instead, successive governments have assumed they could control it, but in fact lacked the knowledge and understanding to do so.

4. Policy objectives

At the beginning of this report, we set out both the broad policy goals the present Government has set for the NHS and a series of more specific targets. These appear to provide a set of criteria against which judgements can be made about its performance in respect of elective care. But in practice they do not.

The analysis of this and the earlier King's Fund work (Hamblin *et al.* 1998) suggests that these targets are based on an inadequate appreciation of how the elective care system as a whole functions. In particular, they overemphasise Stages 3 and 4, when patients have been referred for consultation and treatment and neglect Stages 1 and 2, before they gain access to secondary care. They also ignore the broad forces, such as medical technology, which impact on the system as a whole. Moreover, by concentrating on a single national total, the Government is in effect treating a reduction in numbers as of the same importance, no matter what the local situation is, i.e. whether access levels and waiting time are above or below average. But if the Government targets are inadequate, what should replace them?

Before considering specific targets – which we turn to in Chapter 5 – we must consider how the broad criteria of efficiency, equity and accountability, against which any part of the NHS has to be judged, apply to the elective care system. Only when this is clear, can specific targets be set. We take these citeria in turn and consider how each might apply to the elective care system.

4.1 Efficiency

Technical efficiency

Efficiency in this sense means minimising the cost of producing a given volume of care. The notion of a technically efficient elective care system encompasses a range of matters about which there is little controversy. At Stages 3 and 4, the aim is clearly: to 'do the right things right' in clinical and cost-efficiency terms. This aim comprises a wide range of matters such as clinical performance, in terms both of quality and quantity, theatre management, administration and all the other aspects of efficient provision of care including diagnosis as well as treatment. All these are clearly important but they would be so even if there were no waiting lists. We therefore discuss them no further in this chapter: we return to these issues in Chapter 5.

Our system perspective, however, involves not simply focusing on the appropriate intervention – 'doing the right things right' clinically and cost-effectively. It also involves doing the right thing to the right people at the right time. This extends the notion of technical efficiency to effective and timely identification of need for treatment, and its timely performance.

As far as the individual is concerned, it might be argued that as the NHS poses no financial barriers to access, it is up them to decide when to see their GP. At Stage 1, therefore, an efficient elective care system would simply imply that everyone had rapid access to a GP. However, where individuals are not well informed about the risks they face or when they are deterred by other obstacles, the choices they make may not be 'efficient', i.e. they may delay access until their own chance of effective treatment is reduced or the costs of treatment are higher than if they had sought treatment earlier, or they may not seek access at all.

In respect of some forms of cancer care, screening programmes are in place, supported by call and recall administrative processes to help ensure that those with the disease receive treatment at an early stage. How many such programmes there should be and over which age groups they should extend remains controversial. In the present context, we need only note that an efficient elective care system may require an active process of identifying who needs care, particularly where individuals may not be effective judges of whether they should seek treatment.

As far as referral from GP to specialist is concerned, the UK system, unlike some other health care systems, deliberately imposes a delay in the interests of the technical efficiency of the Service. In other words, the NHS works on the assumption that the direct access to specialist advice that other countries allow, sometimes from more than one source, is wasteful. Any advantage a particular individual may enjoy through direct access is, it is assumed, offset by the greater delays others will suffer because the specialist's time is absorbed by a larger number of patients who turn out not to need treatment.

The efficient operation of the gate-keeping function within primary care might be defined as immediate and accurate determination on first presentation of the appropriate course of action. In practice, because it is not always clear what is required, referral decisions may be deferred, and hence delays arise, some of which can give rise to higher ultimate treatment costs and/or lower chances of complete recovery. Such delays may occur in a system with very short waiting lists simply because of diagnostic uncertainty.

At Stages 3 and 4, delays may also arise in any health care system, e.g. when patients cannot be treated immediately because of their general state of health or again there may be diagnostic uncertainty. These circumstances apart, timeliness for the individual would not involve any delay beyond that required to fit in treatment with other commitments. As we shall see in the rest of this section, however, the wider system view suggests that delays may have to be imposed in the interests of the NHS as a whole.

Allocative and social efficiency

The allocation of resources to elective care reflects a series of choices:

- the wider context – the NHS and other public services such as education
- within the NHS – elective versus other uses of resources
- within the elective care system.

NHS and other services

We noted above that the main reason for the existence of waiting lists for elective treatment is that that capacity of the NHS to provide treatment is limited. Although it has grown steadily, it has never grown fast enough to match the growth of the potential for treatment. The resulting 'rationing by delay' implies a trade-off between the extra benefit to be obtained by adding to the volume of elective activity and the value of those resources in other users.

As things currently stand, there is no separate budget for elective care, so this trade-off is not made explicitly. The new Government's policy of using targeted funds designed to achieve reductions in waiting lists is moving the NHS in this direction. If the Government were to move towards a system of specific allocations, then it could decide whether it wanted to buy more elective care by increasing activity levels across the NHS as a whole. Whether that was an efficient allocation of resources would depend on the value placed on each of these 'outputs' relative to other users of public spending, such as social care or education.

Elective and other NHS uses

Within a given budget for the NHS, any such increase would be at the expense of other health care uses. The resulting trade-off requires some elucidation.

To an increasing degree, the way that the elective care system works turns on the need to accommodate emergency pressures. As a consequence, neither high level nor operational objectives can be set for elective in-patient care in isolation.

In the short run, the nature of the interconnections between elective and emergency care in practice depends on the specific organisational and physical arrangements within each hospital. In all cases, because of common funding arrangements, there are links between the services, but the nature of the physical connections varies considerably. Because of the development of day surgery in dedicated facilities and the maintenance of elective care in small hospitals now closed to emergency admissions, there may be no 'competition' for physical resources. But competition for human and financial resources may remain.

In terms of planning for particular levels of service for elective care on a year-to-year basis, the critical interconnection stems from the unpredictability of emergency care demand. The demand for emergency care as a whole shows, as might be expected, a more or less predictable pattern of seasonal variation round a winter peak, although demand for some forms of care peak in the summer. Moreover, the scale and precise timing of the winter peak varies. Recent reports (Department of Health, 1997; 1998) have suggested that elective care is scheduled through the year so as to free up resources during the winter months – that in itself means that some elective care delay is imposed to meet the requirements of the NHS as a whole. But we are not aware of any estimates of the scale of this effect.

The significance of these unpredictable variations depends in turn on the existence or otherwise of spare capacity within the hospital system or of means to cope with sudden variations in demand outside it, e.g. through greater use of community resources or through bank nursing or other facilities such as nursing homes.

For a variety of reasons such as financial pressures and nursing shortages, the English hospital system appears currently to have very little spare capacity, even outside the winter period. As a result, emergency care requirements interfere with elective care provision, as evidenced for example by the cancellation of operations (Mallender *et al.* 1997).

Given more financial resources, the hospital system could plan to run at a lower level of capacity utilisation so as to ensure that elective care provision was isolated from external 'shocks' and cancellations. If this were done, then the volume and sequencing of elective care could be carried out on a fully planned basis.

But there would be a cost attached to doing so arising from the need to invest in extra capacity, or alternatively to reduce the quality of service offered to emergency patients. Assuming a fixed annual workload, the higher the budget, the greater the probability that a given pattern of planned activity can be realised. At current budgets and workloads, it seems that this probability is low. That may, if the Government placed a high value on reliability – as its proposals for booking elective admissions suggest – reflect an inefficient, i.e. too low, allocation of resources.

However, official statements in recent years have explicitly emphasised the importance of hospitals being able to cope, without undue delay, with the winter peaks of emergency medical admissions. This is just one aspect, however, of the quality of service hospitals may offer to emergency patients. At one extreme, the aim could be to ensure that no one who is brought to hospital with cardiac arrest dies if he or she could be saved if sufficient resources were on standby for this eventuality. These would often lie idle, waiting for the arrival of the next person who could have their life expectancy improved, no matter by how little, by treating them as soon as the diagnosis is made. Such a high standard might well be considered a poor choice when there are so many other unmet needs in society – indeed implicitly it is so judged by the existing allocation of resources. But in other sectors, such as public transport, the value attached to reducing mortality implicit in safety measures appears much higher.

Within elective

If the Government, on behalf of society at large, decides that lengthy waits for elective care are acceptable, then an efficient system requires a means of judging the order in which patients are treated. The elective care system must also operate to deter 'low' value activity either from entering at all or from entering and proceeding through it ahead of other more valuable activity.

The range of costs considered in the previous chapter would be potentially relevant to assessing the appropriate order. In addition, considerations of fairness and equity can clearly play a large part in such decisions – indeed they could be paramount. What equity considerations may be relevant we consider in Chapter 5, where we discuss the scope for introducing scoring systems to assist 'ordering' judgements.

Overview

Looking across all four stages of the elective case system, an efficient system would have the following characteristics:

First, the various elements making up the elective care system would be efficiently provided, i.e. technically efficient in themselves.

Second, the allocation of resources between the various stages would be the best that could be achieved, i.e. the elements taken together would be technically efficient. This would mean, for example, that if there were insufficient resources to eliminate delays, those resources that were available would be concentrated on the links in the system where delay could be most cost-effectively reduced. For example, in the case of access to cancer care, the evidence cited above suggested that it may well be that the area to focus resources on is the reduction of diagnostic delay and delays in initial presentation.

Third, the balance between different claims on NHS resources would be allocationally efficient. If the hospital is to be efficient that may imply some waiting to ensure that there is always a pool of potential patients to keep all the capital and human resources active. Similarly, the requirements of emergency care have to be balanced with those of elective care, not simply because both lay claim to financial resources but also because of day-to-day conflicts over access to the physical resources of the hospital.

Fourth, the process of waiting would be managed so as to ensure that the mix of activity carried out is socially efficient, i.e. that the treatments funded represented the 'best' possible set. Whether waiting lists are efficient in this respect turns on how effective they are at deferring or diverting claims on NHS resources deemed of low value, which in turn depends on how they are managed and on the precise targets that are set either for reducing them or for setting limits to overall waits. There are, as we shall see, other ways of matching demand for care with the resources available. The point here is simply that some means must be found of ensuring that the 'best' mix of activity is chosen if the capacity of the elective care system is such that delays are unavoidable.

All these choices involve judgements about the relative value to be placed on different benefits accruing to different groups of people. Different people may, therefore, have different views about them, which in part would reflect their differing views as to what constitutes an equitable distribution of benefits. But although social efficiency and equity are closely related, they are not identical. Therefore, we next consider equity in its own right.

4.2 Equity

Equity is the second guiding principle, and one that is particularly important to public perception of the NHS. The sense in which equity is generally understood in relation to the NHS is that treatment should be available according to need, not ability to pay. The second half of this principle has been easier to achieve than the first.

At the level of principle, the meaning of treatment according to need has never been precisely defined by policy-makers in the NHS, nor have its various interpretations been generally recognised by those delivering care. Clinicians have generally been assumed to achieve this objective 'automatically', simply by the impartial exercise of clinical judgement. But as we have shown above, there is substantial evidence that utilisation varies in ways which cannot be related to variations in needs and a large number of reasons why utilisation rates may vary between apparently similar groups of people. Moreover, there is very little information available as to how decisions as to priority in time are made.

Thus, two broad types of potential inequity can arise:

- clinical judgements may be based in part on non-appropriate characteristics
- clinicians and managers' judgements may vary from place to place.

These two broad types of inequity have been characterised as offending vertical and horizontal equity respectively.

Vertical equity demands that two people on the same cataract waiting list who *are* different in a number of ways (degree of sight loss, age, employment status, etc.) should only have those differences that are deemed to be morally relevant taken into account when choosing who has the greater claim to treatment and, among those receiving treatment, in what order to time priority. While it may seem obvious that degree of sight may be a justifiable differentiating factor, can the same be said for employment status?

Horizontal equity demands that two people in identical circumstances should be treated identically – they should have the same level of priority and receive the same standard of care. Thus, two people (similar in all socio-demographic and clinical aspects) who live in different parts of the country but who have the same loss of vision because of a developing cataract should not wait different lengths of time or have differing chances of being operated on.

There are some further complications. Equity also demands a fair allocation of resources between different types of elective treatment – so that those waiting for heart surgery are not unfairly advantaged or disadvantaged with respect to those waiting for a cataract. Indeed the balance between elective and other forms of care (emergency), and the level of care provided at the various stages in the access pathway, needs to be fairly apportioned too.

One implication is that the level of resources devoted to a particular condition should be equally distributed across the country relative to the extent of ill health for that condition in that area. Another implication is that the relevant criteria that are used for choosing between individual patients waiting for treatment should also be standardised, and the same weight given to each criterion when decisions are made. Thus, even if differences in efficiency or fluctuations in demand mean that some people get faster treatment than others, the basis on which the decision was made – the criteria – should not vary.

Both these implications alert us to a fundamental requirement of horizontal equity in geographical terms – the centralisation of decision rules. The existing means of allocating finance from the Centre to the local NHS are intended to create the potential for horizontal equity in the actual levels of access to care available, but they do not ensure it – as the data presented in Chapter 2 on differences between areas indicates. As we have noted already, from a practical point of view it would be extremely difficult in a largely decentralised system to ensure equity for all forms of treatment. Nevertheless, the Government has proposed performance indicators for some conditions, as we have noted above.

Much debate has taken place over precisely which criteria are, and are not, relevant to *vertical* equity considerations. Government policy has typically, and rather unhelpfully, made reference to 'need'. Practical policy has focused on time waiting. But what other criteria might be relevant? These can be separated into clinical and non-clinical factors – though in practice they may interact: see Table 4.1.

Table 4.1: Possible prioritisation criteria

Clinical	**Non-clinical**
extent or seriousness of ill health	age
urgency	existence of dependants
effectiveness	time waiting
individual health gain	'lifestyle'
cost-effectiveness	economic productivity
	other socio-demographic or personal characteristics

Simply referring to 'priority on the basis of need' is insufficient, because it does not specify which need should take priority. I may 'need' care because I am seriously ill, because I am ill with a condition for which there is an effective treatment, or because I am young and have a dependent family, and so on. These all constitute different interpretations of 'need'. It follows that the apparently consensual remark that the

NHS must deliver care equitably conceals an enormous range of conflicting interpretations about what this 'equity' involves. A number of attempts have been made to formulate points or weighting systems that take some or all of these factors into account, either for specific conditions such as serious heart disease or across all conditions. But as Edwards has pointed out (1997), these do not resolve these conflicts; rather, they reveal their nature and extent.

While in principle fairness can be improved by ensuring that like patients are treated in the same way over time, and that careful consideration is given to the appropriate criteria for exercising these decisions, it will be hard if not impossible to reach unanimity on what those criteria should be or how they should be weighted relative to each other. But some degree of agreement is more likely to result from coherent reasoning after open discussion and deliberation than if the criteria remain the sole province of each clinician or of the way that those administering lists on a day-to-day basis work.

Practical considerations

A quite different set of issues arises from considering the feasibility of achieving whatever principles of equity are accepted. We set out above data relating to the variations in access observed in practice. As the discussion of Figure 3.5 showed, these variations arise for a large number of reasons – variations in clinical practice, efficiency in provision, historical factors and accidents of geography, personal behaviour and so on.

The policies the Government has announced in relation to clinical quality, national service frameworks, etc., will tend to reduce the extent of these variations, though it will be a slow process. Standardisation of the identification of needs will be a yet more difficult task, implying as it does both accurate epidemiological information and a means of ensuring that the needs identified by these means are identified in practice – i.e. that the 'right' people (those with the appropriate condition) come forward.

Finally there are a number of general reasons why the Centre will inevitably be limited in what it can achieve by means of rules and codification. As noted above, two of these reasons relate to competing principles by which public services are judged, and which often pull in different directions:

- although it is attractive to believe that variations in outcome/provision in relation to the extent of ill health should be ironed out – possibly by the imposition of centrally specified rates – *in extremis* this will deter innovation in the efficiency with which services are provided. In any dynamic system, different parts should have some scope for developing better or more efficient methods of delivery and treatment. This will inevitably lead to some inequity, strictly defined

- the sheer enormity of the information needed to impose centrally determined levels and a range of provision is far beyond the capability of current – and possibly any future – information and central management systems

- a system entirely run from the Centre will risk the complete loss of any local accountability – except for adherence to performance standards in the provision of care – and democracy, which could in turn damage public support for the NHS. Local variations may perhaps reflect local preferences. But the case for variations based on local preferences would be all the stronger if there were accepted means of determining what these in fact were. At present, despite a great deal of experimentation with public consultation at local level, no such means exists.

Thus none of the imperatives implied by equity considerations are easy to achieve – whether in practice or in terms of agreement about the principles which underlie them. We may never achieve consensus, for example, about which criteria are morally relevant when deciding the basis on which to discriminate between waiting patients. But the overall objective of achieving fairness is nevertheless crucial for the NHS and must continue to form a central part of debates about how to improve the way it is run. In Chapter 5, therefore, we identify some areas where progress can be made.

4.3 Accountability

The system by which people have gained access to elective care for most of the life of the NHS has not been subject to any genuine degree of accountability. As we have seen in Chapters 2 and 3, the way that resources are used within elective care reflects values other than purely clinical ones, and even within the clinical field, some of the key issues such as the link between delay and outcome remain poorly understood. GPs have been free to refer patients to consultants and consultants have been free to choose whom they should treat and how. Up to very recently, they have largely been also free to choose how to treat as well. The results of this extensive discretion have been significant and unjustifiable variations in access between different parts of the country.

These variations reflect the central characteristic of elective care – that the NHS exercises considerable choice as to what to provide. Effective accountability requires that the exercise of choice should be justifiable against agreed criteria.

The requirements of system-level accountability are that decisions should be taken openly; that there should be scope for debate about the terms on which they are taken by all interested parties; complaints and redress procedures for challenging decisions by those affected should be in place; and there should be mechanisms of sanction when it is believed that improper decision-making is taking place. All these considerations apply to access to elective care at the system level as well as the individual NHS trust and the individual commissioner.

Despite the Government's insistence on a national waiting list target, accountability is currently weak at the national system level. Although the Department of Health now publishes waiting list data monthly, so that 'progress' can be regularly monitored, there is no routine reporting nationally that allows judgements to be made as to the 'mix' of procedures currently being bought, the efficiency with which they are provided and whether the distribution of benefits is equitable between different groups in the population. We noted above the introduction of an equity of access indicator comprising a small number of common elective procedures. By itself, however, it sheds very little light on these issues.

Although a large volume of data – hospital episode statistics – is collected about the workload of hospitals, there is no published national level analysis available that would reveal the different bundles of care being purchased in different areas, the extent to which procedures poorly supported by evidence as to their effectiveness are declining and vice versa, or the impact on patients of those procedures that are carried out. As we noted above, however, the National NHS Patients Survey has for the first time revealed some of the costs of waiting – but not in a way which allows those costs to be linked to particular conditions.

As far as the individual procedure is concerned, the new arrangements for clinical governance are designed to provide a means of bringing clinical activity into NHS lines of accountability. At the level of the individual patient, individuals can use the NHS complaints procedure (and if that fails the Health Service Ombudsman) or take legal action, if things go wrong. Although both the complaints procedure and the powers of the Ombudsman to investigate are new, they appear to offer to individuals appropriate means of seeking non-financial redress – and in principle the Ombudsman could be a route to financial redress as well. In this respect, the arrangements now in place are much stronger than those that have been in place for most of the life of the NHS. As for financial redress, a number of steps have been taken recently to simplify the legal process and make it easier and cheaper to use.

In between the national system and the individual, it is less clear how accountability should work. This Government, like its predecessor, has preached the desirability of local choice while imposing central priorities, but its national targets for waiting list reductions have effectively overridden this. As Alan Milburn told the Health Committee (House of Commons, 1998):

> *The first point to make is that primary care groups, primary care trusts when they come on line, just like everywhere in the National Health Service, will be subject to the need, as everywhere else, to ensure that waiting lists and waiting*

> *times come down. That is the Government's policy, that is what we want to see and we believe that it will happen. I think that because you are putting primary care groups in charge of budgets and getting a better deal for patients, what you will actually see is a further level developing to get waiting lists and waiting times down, because my guess is that the primary care group is going to be pushing pretty hard for the individual hospital and the clinical teams within individual hospitals to provide the quickest and the highest possible quality of care for the individual patient. We should see the primary care groups developing as a further means of getting these waiting lists down and consequently waiting times down.* (p.19)

Furthermore, the NHS performance assessment framework now coming into operation (NHS Executive, 1999) could lead in the direction of a 'national standard elective treatment package', i.e. a defined set of intervention rates for all common procedures.

It could stop short of that and confine itself to technical efficiency – including clinical quality – in the production of care, thereby allowing genuine local choice over the 'mix' of what is bought within the elective care budget and, by extension, what are acceptable delays and waiting times. Unless there is clarity as to where discretion is to reside, then the appropriate means of accountability cannot be determined. But however this issue is resolved, it is clear enough that existing arrangements do not provide adequate accountability:

- the means by which prioritisation between patients with a given conditions is carried out are implicit and hence not subject to critical scrutiny
- the balancing of priorities between conditions typically takes place without overt justification
- differences between different parts of the country remain substantial, largely unexplained and in no way justified

4.4 Concluding comments

Although the three broad criteria discussed in this chapter are familiar, our discussion has shown that they raise a number of difficult issues of practice and principle. At the level of practice, our discussion has shown that any systematic attempt to ensure that the elective case system is efficient, equitable and accountable requires more information and understanding of how it does and might work than currently exists.

At the level of principle, there are potential tensions between the three broad criteria. At the broadest level, the goal of equity may conflict with the goal of efficiency. At the NHS trust level, efficiency might require innovation which all providers are unlikely to carry through at the same rate – as the Government has recognised with its incentive payments to those reducing their lists the most. Such differences militate against horizontal equity, unless care commissioners can offset poor performance by 'shopping around' and buying from better performers. At the individual level, emphasis on efficiency may suggest that economic status (and loss of output while waiting for care) should be taken into account; equity may suggest it should not.

Second, local choice – as and where that is allowed to develop – will tend to conflict with equity as between different parts of the country. As we have argued, an overriding concern with equity across the NHS as a whole inherently implies nationally agreed rules for determining priorities within elective care and between elective care and other NHS activities. But if the local choice implied by the introduction of primary care groups is to mean anything, then it should lead to different preferences being expressed for one kind of patient or procedure over another, or different access thresholds.

In what follows, we do not aim to resolve the tensions between the criteria considered in this chapter; instead we set out two broad policy options that could be adopted within either a localist or a centralist NHS.

5. Broad policy options

In the first part of this report, we asserted that at no time since the formation of the NHS had a systematic policy towards elective care been developed. We noted that there has been a persistent failure to acknowledge the links between all the elements considered here, i.e. the elective care system as a whole. Despite the present Government's expressed belief in 'whole systems' approaches, in practice policy has continued to be overwhelmingly aimed at parts and single elements of the NHS rather than the system as a whole as defined in Figure 2.1.

One central message of the systems approach adopted in this report, however, is that the way that the individual parts work turns on their mutual relations and that the way that the system as a whole works can only be seen by looking at it in its entirety. From this observation, two main points follow:

- While waiting undoubtedly poses costs on individuals, and possibly the NHS as well, the existence of waiting lists performs a critical function – to contain demands within limited resources. As we argued in Chapter 3, the waiting that characterises the UK elective care system is a means by which care is rationed by deterring, delaying or otherwise managing, access to some forms of treatment. If waiting were to be eliminated or substantially reduced, then this function would have to be discharged in some other way unless the resources devoted to elective care were substantially increased. The system question is, can other means be found for identifying 'low value' uses of resources if waiting does not do this?

- Making improvements in one part of the elective care system may not result in improvements overall, because the responses of those other parts may counteract the impact of the original improvement. Specifically, we have argued that the historical evidence suggested that policies to reduce delays at Stage 4 could result in more demand appearing at Stages 1 and 2, thereby wiping out the initial gains.

The force of these points rests critically on the assumption implicit in the paper as a whole that we cannot spend sufficiently more to eliminate waiting lists. We consider next whether this assumption is justified and then set out two alternative policy options.

5.1 Can we spend our way out of trouble?

For the reasons set out in Chapter 2, eliminating waiting will be difficult. We know from the experience of other countries, which spend considerably more than the UK, that waiting for elective care still exists. Furthermore, there are few signs that the stream of medical innovation is drying up or that the other forces set out in Figure 3.1 are becoming less powerful. Nevertheless, it is worth asking whether such an objective is a sensible one. There are two reasons why it may not be so.

The first is that it will not avoid the difficulty of rationing, as broadly conceived. That is, clinicians and others will still need to make decisions about where to draw the line on treatment. Some people will still have to be dissuaded from treatment, maybe because their condition is not serious or the chances of success are too slim. The number of these people will be much smaller than under current resourcing, but nevertheless they will remain. For others accepted for treatment, similar decisions will remain about how intensively to treat – how many diagnostic tests, how much time spent on each case, hospital staff devoted, after care provided, etc. – all of which require judgements about when the benefits to be gained from treatment are too slight to continue. Difficult (rationing) choices will remain.

The second reason is that seeking to eliminate waiting may well require inefficiently high levels of spending. In the first place, because demand for health care fluctuates from day to day and month to month, sufficient slack will have to be built into the system so that any surges in demand will not reintroduce waiting – albeit temporary. Thus, excess resources – doctors, capital equipment, etc. – will be under-utilised during those periods where demand is lower than average. This would be a form of technical inefficiency or 'waste' unless the value placed on eliminating waiting was very high.

More fundamentally, the costs of treating patients with minor ailments (who would previously been deterred from entering the system because of the existence of waiting) or whose conditions have not yet developed significantly may be greater that the benefits derived from them. This will inevitably occur if gatekeepers are not encouraged to consider the cost of accepting patients – for them and their patients (who, in the NHS, face no costs at the point of use). This is an opportunity cost, borne by others in terms of what can no longer be done for them, whether they are potential patients needing treatment in other sectors of the NHS, or people needing housing, education or welfare benefits.

If resources are guaranteed to keep pace with referrals, then any benefit, no matter how small or probabilistic, will appear to be worth pursuing. The outcome will be a system that is inefficient in allocative terms, i.e. too many resources are diverted to elective health care in comparison to other uses. As noted in Chapter 4, an efficient health care system requires some form of rationing. By their very existence, waiting lists have a useful effect in controlling utilisation and restraining expectations and hence demand. Whether they are the only way of carrying out these functions is another matter – which we consider below.

These general arguments, however, do not justify the conclusion that the existing pattern of waiting is acceptable and that the level and composition of elective care in England are adequate. In fact, some of the evidence set out in Chapter 3 suggests they are not, particularly for serious heart disease – as the new Secretary of State, Alan Milburn, appeared to accept with his first policy announcement (Department of Health, 1999). But, as we have argued already, the elective care system is so poorly understood that the basis for arguing for a substantial increase in the budget does not exist. Both the options that follow and the recommendations at the end of this chapter are intended to help the NHS determine what the level of spending on elective care should be.

5.2 Two broad options

We now set out two broad strategies for reforming the present system and resolving some of the issues raised in the previous discussion. In each case we look at all the stages of the access pathway set out in Figure 2.1.

In presenting policy options we will make a distinction between two sets of strategies that are commonly discussed – and often conflated – under the single heading of *demand management*. The first of these – and the proper use of the term – focuses on how individuals decide whether or not, and in what way, to make use of health care services. It includes any policy aimed at changing both the level of demand on the formal system and the distribution of that demand between different parts of the Service. It also includes measures that divert demand from formal to informal care. The essential point, however, is that these policies – such as NHS Direct online and health promotion in general as well as proactive screening programmes – aim to affect decisions about seeking care made by individuals independently of the availability or otherwise of the services required to treat them. At the most general level, they might also include policies to mitigate the increasing belief that medical treatment is the best method of addressing what is perceived as poor health, or even the perception of what 'needs' consist of. More specifically, they include measures to fully inform potential users of the risks and benefits of different treatment options. Demand management bears largely, but not exclusively, on Stage 1 (of Figure 2.1) and it may lead to an increase or a decrease in the utilisation of services.

On the other hand, what has been termed *utilisation management* (Robinson and Steiner, 1998) refers to how patients actually make use of the system, i.e. how demand for the services of the elective care system is translated into system use. It applies to Stages 2 and 3 and includes measures that bear on professionals as well as users. Thus the GP gatekeeper system and clinical guidelines bearing on referral to hospital are not designed to affect the user's initial decision, but only their subsequent use of the system. Utilisation management also includes the need to define what is on offer to patients, either by removing certain services or procedures from the scope of the NHS or by ruling out certain treatments on the grounds that they are not clinically or cost-effective.

We have argued that some aspects of the system – such as the existence of long waiting lists – affect demand indirectly if people judge that their condition is not going to be treated with urgency and they go elsewhere. But the existence of long lists is essentially a utilisation management device. It allows the key decision-makers, principally hospital consultants, to create a pool of people from which they can choose the most 'deserving'.

The fundamental point is that any system of health care delivery operating within a fixed budget must embody these two kinds of measure, i.e. both demand and utilisation management, albeit in different combinations. It follows that if either one is relaxed, for example by encouraging access for some disease groups, the other must be strengthened, unless a corresponding increase in resources is forthcoming. Thus, the new emphasis on heart and cancer patients can only be achieved by discouraging other uses or erecting higher barriers to utilisation for other conditions.

We also make use of the terms *system design* and *service design.* It is apparent from the evidence cited in Chapter 2 that there are disparities between different parts of the country, arising from the way that services are delivered, at the level of the hospital or the specialty and linked together to form systems of care. These have arisen for a large number of reasons, of which the most important has been the lack throughout most of the life of the NHS of systematic attempts to devote resources to the conscious design of services and systems of care. This gap has been recognised by the present Government in the introduction of national service frameworks (Department of Health, 1998).

In what follows, we take first a strategy involving piecemeal improvement – essentially a development of the policies currently in operation. While this may offer improvements over existing arrangements, it risks making the situation worse. We, therefore, go on to consider a second option involving a more radical set of measures.

For each we set out the main components and their likely impact both at each stage of the access pathway and system-wide.

Piecemeal improvement

As discussion of Figure 3.5 has shown, the elective care system might be improved, in terms of both efficiency and equity, by taking action on each of the factors influencing the way in which access to care is attained and each of the stages which comprise it. Equally, for the reasons set out above, many will be hard to implement and some may prove impossible in practice. A practical set of policies would include the following:

- **Stage 1:** demand management measures targeted at groups identified as low or slow self-referrers and systematic attempts to inform potential users of the overall benefits and risks they face from treatment.
- **Stage 2:** utilisation management through development of agreed referral criteria for a range of conditions such as cataract where diagnosis is normally straightforward and clinical opinion less diverse than for other conditions.
- **Stage 3:** demand management through better information on user preferences, and utilisation management through definition of thresholds for treatment.
- **Stage 4:** system and service design involving 'standardisation of service delivery' and utilisation management through development of scoring and booking systems.

Stage 1: Demand management

It would not be feasible to try to ensure that potential users present at the same stage in the progression of their condition or disease. But there is scope for targeted intervention where there is reason to believe that access to care is delayed too long, particularly for conditions where the chances of full recovery are greater, the earlier that the condition is identified.

In the case of breast cancer, for example, the King's College Hospital study cited in Chapter 2, concluded that:

> *Patient delay contributes the bulk of total delay. Effective health promotion strategies for reducing delay should therefore include approaches which encourage* ***all*** *symptomatic women to present early.* (p.5)

It then goes on:

> *By isolating factors associated with long delays, high risk groups (e.g. patients with negative attitudes to cancer) could be identified and specially targeted with health promoting material.* (p.5)

In practice, there is no obvious, generally applicable, way of doing this. The evidence cited in Chapter 2 suggested that it is hard to identify groups who were particularly likely to delay presentation. A recent systematic review concluded that:

> *The strength of the current evidence is inadequate to inform the development of specific strategies to shorten delays by patients [or providers].* (p.1127)

But a broad brush approach aimed at the population at large may be inefficient. Only a small proportion of those seeing a specialist turn out to have a condition requiring or justifying treatment, so the risks of generating a large increase in presentation with only a small increase in numbers identified for treatment are obvious.

With the National Screening Committee, there is now a clear focus for advising on screening programmes. The Government has also recently announced a number of initiatives, backed by targeted funding, to encourage the use of health care among groups such as the learning disabled (Department of Health, 1997a) who are thought to under-use services. In principle, the health check for elderly people should have the same role but, as evidence cited above indicates, it cannot be assumed to be effective in its present form.

Another potential route is to identify and follow-up on low or late presentation on a district by district basis. If such a process could be defined, then locally determined programmes could be mounted. In principle, the way to do this would be to use waiting list and hospital activity information combined with epidemiological information, to estimate whether a local elective care system was identifying the 'right' number of people to be treated. In practice, all the steps in such a process would be hazardous, as Sanderson and colleagues (1997) have shown for prostatectomy. However, examination of the condition at which people come forward, i.e. the stage of their disease, would be feasible at least for those conditions where progression is steady and there are accepted means of measurement. In this way, failures to identify in a timely fashion people who would benefit from treatment could be identified. In practice, however, they are not: see Box 5.1.

Box 5.1: Scotland – variations in referral

The Public Accounts Committee – to which the National Audit Office report on cataract surgery was submitted – questioned the NHS Executive Scotland about the reasons for variations in referral but found that:

> *27. ... They could not explain this precisely. They pointed out that it was the job of the local health board to look at individual practices, different levels of need and different levels of uptake. However, they outlined a range of variables at play: the extent to which patients came forward; whether they visited their general practitioner or optician; whether they complained of a problem with their eyes; whether the general practitioner routinely checked for visual acuity; and what happened when a patient arrived at hospital.*

In other words, despite centrally imposed targets, the Scottish elective care system was not centrally managed.

Source: House of Commons, 1998.

Stage 2: Utilisation management

In principle agreed referral guidelines should produce greater uniformity in the number of patients referred from each practice, but as we have noted already it would be unwise to expect rapid progress. In the case of cataract, the National Audit Office reported (National Audit Office, 1997) that some GPs themselves considered that referral guidelines would be of limited benefit unless they were accompanied by improved basic and continuing education in eye medicine for GPs.

The same may well be true of other conditions. If so, the best route for improvement may be, as at Stage 1, a targeted approach either on conditions where under-referral is suspected or, on a geographical basis, where the quality of primary care is poor. Either way, the basis for any intervention should be a prior analysis of the conditions and sections of the population suffering from under-referral. Guidance (NHS Executive, 1995; 1999) has already been issued for breast cancer referrals.

Stage 3: Demand management

A second demand management strategy is to provide information to patients at the point where treatment options are being considered. Some early work in this field (Kasper, 1992) suggested that patients might choose 'watchful waiting' rather than surgical intervention for benign prostate hyperplasia when the implications of the options available were made clear to them. However, such studies have been carried out for only a small range of conditions and so there is currently no way of estimating the total potential impact of shared decision-making on demand for care. As Coulter has pointed out (1997), there are other important justifications for shared decision-making, so it should in any case form part of the incremental strategy.

Utilisation management. As noted in Chapter 1, there have been frequent calls over the years for nationally agreed treatment protocols and for the definition of thresholds at which treatment may be offered. The Clinical Standards Advisory Group report, *District Elective Surgery*, lists a large number of guidelines as well as primary studies bearing on particular procedures. It also records that these are generally ignored and that better co-ordination is required. The establishment of the National Institute for Clinical Excellence is explicitly aimed at rationalising guidelines and at promoting their implementation. It seems reasonable to expect that a greater degree of uniformity will be achieved, albeit slowly, in this way.

Stage 4: System and service design

We suggested in Chapter 2 that in general the structure of the hospital system was satisfactory for elective care, in that the volumes reached for most procedures were adequate. But there are exceptions to this, as CSAG has shown for a range of services and as Calman-Hine argued for cancer care.

The New NHS proposed that the Calman-Hine proposals for cancer care should be used as a model for a series of national service frameworks covering major services.

Elsewhere (Harrison in Klein, 1998), we have argued that the proposals put forward for cancer care suffered from important weaknesses. However, the underlying principle that each service should be designed in the light of the best available knowledge is sound. The problems are those of implementation and lack of knowledge; progress will be slow, albeit in the right direction.

Utilisation management. Although there has been considerable interest in such scoring procedures for establishing clinical priorities (Agnew *et al.* 1994; Naylor *et al.* 1996), with the exception of maximum waiting times, the order in which patients are treated is currently a matter of clinical discretion. Greater equity between patients might be achieved if this process was put on a more rigorous basis. As we have argued in the previous part, determining what the 'right' ordering procedure is might in principle require taking a large number of considerations into account. In essence, a scoring system involves a systematic attempt to place a numerical score on each of those considerations deemed to be relevant for each potential person waiting for care, as a means of ranking their priority for treatment. The various elements are then quantified and weighted. The person with the highest 'score' waiting for treatment is admitted first, and so on.

It is apparent from our discussion in Chapter 4 that it will be very difficult to reach consensus on a single set of weights covering all the possibly relevant factors and to implement them in a common way. There is no accepted way of combining them into a single indicator of priority and even if the issues of principle were resolved, the empirical importance of some of the factors – e.g. the health care or employment costs of delay – has yet to be established. Even with the narrow field of clinical priorities, there is no general agreement as to how they should be determined, i.e. agreement as to how patients with different conditions should be compared. The British Medical Association have recently come out in favour of some simple measure of clinical severity but have not shown how it should be devised, tested and implemented (BMA, 1998). However, pilot work in Wales involving members of the public as well as clinicians (Edwards, 1999), the full results of which have yet to be published, suggests there are high levels of agreement in some issues – e.g. what factors are relevant and what are not in determining the order in which patients are treated – but not in others.

So, however desirable it may be in terms of equity between patients – vertically and horizontally – in practice it will prove difficult to achieve a national scale. But greater uniformity for particular conditions such as cataract or CHD, where the chances of clinical agreement appear relatively high may be achievable. It may also be achievable for the conditions singled out by the Government for greater priority if the central management promised becomes a reality and more uniform treatment thresholds are introduced as a result.

Booked admissions. Booked appointments – i.e. where the patient is given a date of admission at the time the decision to admit is made – of those waiting for NHS treatment. In some trusts, the overwhelming majority of patients in some specialties have an admission date. These are usually where there is a small list size or where demand is not rising quickly, and we have noted that such examples are the exception not the rule in the NHS – although many will be given a firm date some time after the decision to admit is made.

The Government has recently announced its intention to extend the proportion of booked admissions. This currently represents just over one-third of the total and mainly comprise urgent and diagnostic admissions, with shorter than average waits. The proportion for orthopaedics of booked admissions – where as noted above, waiting lists are long – is much lower than average.

The scope for extending booking turns on the way the hospital system as a whole works. If the characteristics of future patients accepted into the elective care system were known accurately in advance, then in principle all could be assigned fixed dates, which would not need to be changed. But the condition of those waiting, as well as those not yet requiring admission, can change suddenly, thus requiring a re-ordering of priorities. Could this be accommodated satisfactorily within a booking system whilst retaining an increased level of confidence in patients in their bookings?

The answer turns, as with the balance between emergency and elective care, on the cost of keeping resources under-used to deal with the unpredictable. Given the current variations between different hospitals, there may well be some scope for extending booking simply by adopting better administrative procedures. But this would not resolve the problem of how to deal with the persistence of insufficient capacity to deal with all the demand that might come forward.

So what might happen if the current system of limited 'booked' admission dates were expanded incrementally, coupled with increasing activity, with a view to providing all patients information about when they were likely to receive treatment? At its simplest, this would involve consultants being encouraged to assign a date for admission, even where there are long lists, and try to stick to it. It would involve making a judgement about the workload/activity that is possible over a period, the number of people on the waiting list and the relative numbers in particular categories of 'need' or urgency, and the rate at which people are joining the list. From this, the consultant could make a 'guesstimate' about when people would get treated, according to his or her priority-setting criteria.

However, there are a number of serious complications inherent in operating such a booking system in the current NHS. If the rate at which people join the list is faster that the rate at which treatments are undertaken – generally true historically – then the number on the list will grow and, on average, the length of time until admission can be booked will also grow. We have seen that the downward trend in average waiting times between 1989 and 1995 has now gone into reverse as the very long waiters have been removed from lists. Thus, if all categories of 'need' are to be given bookings then, depending on the relative weight given to each category, some low category patients may require dates for admission years ahead if high or middle priority patients are not to be unduly delayed. This may occur even if the list size is in equilibrium (not growing), but is still large relative to throughput: the periods of time into the future would be unwieldy. Such booking systems may not even achieve the advantages of greater certainty – the length of time to wait is such that some patients would become dispirited and seek care elsewhere, the condition would resolve itself, or they might die or become unsuitable for treatment. For all these reasons, cancellations would be high and the benefits of booking watered down or removed altogether, even if the practical problems could be overcome.

Figures 3.3 and 3.4 represented our thesis that utilisation is managed in part by the very existence of lists, which deter increased referral for these low priority patients. If this is true, the effect would almost certainly be reduced through the introduction of booking systems without further reforms. GPs and patients would perceive that, once referred, they would have a 'guaranteed' admission date.

Notwithstanding the fact that bookings would stretch far into the future, people might perversely become *more* inclined to join the system because treatment would now have a documented date of occurrence. The crucial utilisation management function of waiting would have been weakened.

It is hard to see how such a system could operate for long if our conception of the underlying processes is correct. In terms of Figure 3.4, if high category patients were not to suffer (if their sub-list was not to grow to unacceptable lengths), the low category sub-list would have to grow to massive proportions with bookings extending into the distant future. We conclude that if booking is to become the typical means of gaining admission, fundamental change is required – as we set out below.

Assessment

The measures described above would help to produce a system of access to elective care that would be more equitable, horizontally and vertically, than the current situation and which would be potentially more accountable as well. In most of the areas we have considered, however, improvements will be slow to emerge in part because they involve change in clinical practice across the whole of the NHS and in part because in many areas, more research and analysis is required before progress can be made.

Judged politically, the incremental approach is unattractive. It would not eliminate long periods of waiting, unless access at Stages 1 and 2 proved unresponsive to the improvements made at Stage 4. Indeed some of the measures suggested for Stage 1 could add to the numbers seeking access to elective care. Politics apart, that in itself would not matter. But if an increase in the numbers coming forward with less serious conditions occurred within the existing policy framework, with its commitment to waiting time maxima, there would be a serious risk of the system as a whole becoming less efficient since low value treatments could displace those of higher value. An increase in resources would reduce that risk, but not eliminate it unless that increase was very substantial.

It follows that although the incremental option may lead to benefits in terms of (some kinds of) efficiency, equity and accountability, it could also lead to disbenefits. What the balance would be turns on how the elective care system *as a whole* responds to the various measures that make up the incremental option. And, as we have pointed out already, the information and understanding required to predict that is limited.

A radical option

The starting point for this option is the position outlined in the preceding chapters. Waiting lists exist and grow only partly as a means of organising and prioritising those who need elective treatment. They, or more accurately perceptions of the time patients have to wait, also act as a deterrent to utilisation and possibly, by way of influencing expectations, demand. Increasing activity is in general accompanied by increasing referral as the system constantly seeks equilibrium. The virtual impossibility of satisfying all elective need, at least within the financial prospects for the NHS in the UK, means that a focus on simply reducing numbers or the average wait is misguided.

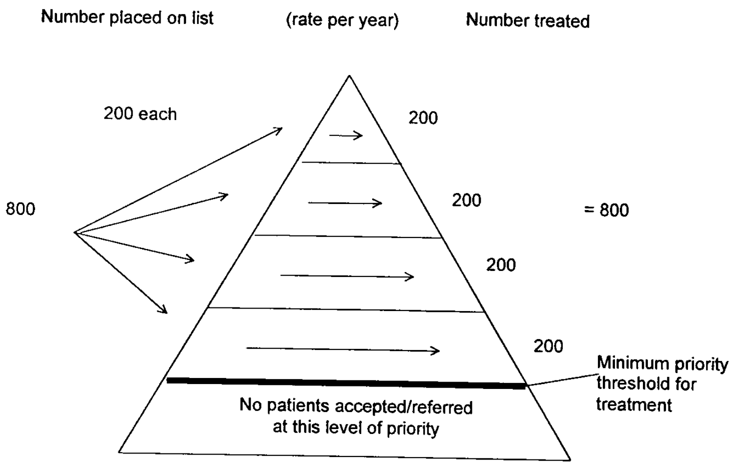
Moreover, we have argued that the deterrent effect of waiting lists performs an essential rationing function, which all health systems require. It should not in this context be seen as a 'problem' but as an important way of managing scarcity. *Any* reform to the system of access to elective care that seeks to remove waiting lists, as they are currently constituted must find away of replacing this function.

The essence of the radical option lies in the use of scoring systems of the kind discussed in the previous chapter to replace the open access waiting list with a 'closed' access system based on overt priority thresholds defined by a minimum points score. In effect, it means putting into practice the proposals cited above, which were put forward by the two House of Commons select committees in the early 1990s for treatment thresholds (pp.11–12 above).

The principal rationale for such a system is to give the patient greater confidence and more information about when they might receive care, whilst simultaneously treating the user of the service with greater honesty about what it can and cannot achieve. This option would not of itself necessarily result in greater activity, or even more technical efficiency in the delivery of care. Neither would it remove the rationing question; indeed it would openly acknowledge that rationing had to be carried out. It could, however, significantly improve accountability by making access thresholds explicit. It would require fundamental change at many points in the system, and has high political risks, but it also offers the possibility of profound benefits to the system as a whole.

Such a system might work along the lines currently being adopted in New Zealand (Hadorn and Holmes, 1997; Feele *et al.* 1999). Here the problems outlined above are tackled by explicitly acknowledging that not everyone can be treated and by overtly denying access to low priority patients. This system, in broad outline, works as follows.

Figure 5.1: Priority thresholds



Consider once again our schematic representation of the UK's waiting list system. Instead of the development of a mortlake consisting of all those patients who will never be treated, or at a very low likelihood, a hard 'floor' or priority threshold, is introduced, determined by the provider's treatment capacity: see Figure 5.1. Those with conditions that do not 'score' sufficiently to cross this threshold are told that they cannot be treated, because there are insufficient public resources for their level of 'need'. They are returned to primary care for management and may return if their condition becomes more serious.

The lower threshold is calculated with reference to two variables: activity over the past six months – as a means of predicting future demand – and available resources determined by the national 'purchasing' commission, of which there is only one in New Zealand. Those who satisfy the threshold criteria are guaranteed treatment within six months – the precise timing depending on their level of priority.

The initial assessment of a lower threshold score is calculated specifically to match supply to demand; thus the rate of acceptance for treatment will be the same as that for treatment itself – the system will be in equilibrium. In theory, then, there is no need for any waiting: with perfect information, patients with scores above the threshold could be treated immediately. However, information is not perfect, and there is always some unpredictable variability in demand. A maximum six-month lead-in to treatment was chosen as a reasonable compromise between the system being too 'tight' and inflexible, and too 'loose' with bookings taken too far ahead and thus liable to cancellation. In practice, the system approximates to the triangle in the figure: high priority patients at the top are treated within a month; low priority patients, but above the threshold, in six months.

Such a booking system must operate with a scoring system if it is to be consistent and based on a systematic assessment of what constitutes priority. The New Zealand scoring system includes clinical effectiveness measures, as well as social factors such as the existence of dependants, and some consideration is also given to time waiting. Cost, however, is not considered an appropriate criterion.

The key feature of the NZ system is that it defines an explicit line between those who are accepted for treatment and those now referred back to general practice. In fact, two lines are drawn. The first is a 'clinical threshold', determined by an assessment of the clinical factors which would 'ideally' indicate to a clinician that treatment should take place. The second is the 'financial threshold' below which people are referred back and which is determined by the financial settlement. Not surprisingly, there is typically a gap between the two, with the clinical lower than the financial – otherwise 'normal' clinical decision-making would be enough to bring the system into balance. In fact, it is unlikely that the clinical threshold is actually 'objective': it probably reflects in part the context of financial constraints and medical norms. There is also an incentive to pitch it at a politically realistic level, thereby encouraging politicians to raise the financial threshold until it matches the clinical. Such clinical thresholds, however, have a habit of becoming moving targets.

In this way, the radical option combines 'utilisation management' coupled with a scoring system. Waiting lists are not abolished, but they are organised to include only those for whom treatment can be guaranteed. There is no need for a mortlake, or sheer list size, to act as a deterrent and no separate determination of maximum waiting times as in the UK. Thus the risk of the upper limit distorting clinical priorities is removed.

The system operating in New Zealand is essentially based on a belief in openness and honesty about what a health care system can achieve under budget constraints. The political 'deal' involves offering the patient confidence about when treatment will occur, instead of the (benign) deceit inherent in a refusal to admit that rationing exists. In return, citizens need to accept that not all health care needs should automatically be satisfied by public resources.

5.3 Issues for consideration in the UK

The system just described brings with it a whole range of issues that are as yet unresolved, even in New Zealand where the system is currently in operation. The first set of issues relates to equity: New Zealand scoring systems have been developed locally and thus vary from area to area, potentially leaving identical patients with different priority scores and thus different chances of treatment. In the UK, such horizontal inequities would be even more difficult to resolve because we have multiple purchasers as well as multiple providers.

Further, there is the issue of equity *between* disease groups: could generic scoring systems be developed which assist the allocation of resources between service areas? This would mean purchasers as well as providers operating scoring systems explicitly. Some degree of generic scoring is also needed to properly prioritise *within* certain waiting list categories, for example between different surgical orthopaedic procedures. And the greater the need to choose between cases requiring different interventions, even within a particular surgical specialty, the more salient the issue of cost becomes. If it is not incorporated as a potential scoring criterion, then it will be impossible to maximise the aggregate benefit for a given waiting list (unless cases are very similar in cost terms, as with cataract surgery).

Another set of issues relates to the degree to which the system can be rolled out to all elective specialties, and this depends in part on whether demand for care is predictable. This is not just a question of variability, which is not a problem as long as the degree of variation is predictable. The issue is to what extent unforeseeable fluctuations in demand occur – random statistical variation – particularly amongst the rather more urgent or high priority categories. If an unexpectedly high number of people present with serious cataract in one month, for example, then there are two choices open to providers: either cancel some of the less serious cases booked in from a few months previously or delay treating the more serious cases. In the UK, the effect of variable and unpredictable emergency work will make this problem more serious, since both categories of work share many of the same facilities.

Over and above these technical issues, the system may pose difficulties for those required to operate it. There will, for example, be increased pressure on primary care practitioners, since it will be their task to explain why publicly financed treatment is being (if only temporarily) denied. This will be difficult for them, since they will have to balance their desire to provide care for the individual with the imperatives of the system. Should they encourage those who can to use the private sector? Should they encourage those whose prognosis is predictable (and thus will cross the financial threshold at a given point in the future) to return at that date, building up future pressures, or leave it to the patient to decide? Will GPs go along with such a policy or simply reject it and disrupt it from within, perhaps by repeatedly sending back patients for assessment and clogging up out-patient clinics?

In principle, the approach could be adopted at national level if thresholds and scoring systems could be agreed for a wide enough range of conditions. But the decentralisation of (most) purchasing decisions means that even with standard thresholds, differences might emerge for the reasons set out in Chapter 3, stemming partly from the supply side of care and partly from local preferences. There would be nothing to stop a PCG, for example, taking its own view as to the scale of the gap between the financial and clinical lines, or indeed determining its own thresholds.

Although, as we have seen, the current arrangements are very far from achieving equity, the nature and scale of the variations between areas are not generally appreciated. A move towards an explicit system, with possibly neighbouring PCGs taking different approaches, would make these apparent to the public at large and hence politically controversial.

The effect on social spending pressures could go either way. The perceived gap between clinical and financial thresholds could provide pressure on the Treasury to increase spending. On the other hand, if the cases referred back are perceived to be of low 'value' socially (and if many of these cases make greater use of the private sector) pressure on spending may be mitigated. In the UK, the private sector carries out large number of procedures, such as joint repair and CABGs, which generally are presumed to be highly cost-effective. The radical approach could lead to more of these being paid for out of public funds and hence add to the pressures on the NHS budget. Either way, this option has the merit of identifying what public funds are being used for. It does not, in itself, solve the problem of determining how much should be spent but it provides the framework within which the issue can be debated.

The effect of a change of this kind on public opinion is highly uncertain. Will the public accept a new political 'contract', or will they reject the NHS in favour of the private sector? The answer to this depends on precisely what is valued in publicly funded health care systems. Is it, on the one hand, their (apparently) all-inclusive nature, and the external benefits this provides in terms of reassurance, social solidarity and sense of community; or, on the other hand, is it their ability to provide the maximum benefit for those who have the most pressing claims, regardless of ability to pay, in a reliable, predictable and democratically accountable way?

Finally, if a New Zealand-type system is to work, it is clear that political support will be essential: hospital management and GPs must be backed-up by politicians, publicly acknowledging that not all 'need' can be satisfied (that rationing is inevitable) and that denying – or deferring – care to some people is fully justified by economic realities. As things currently stand in the UK, this may be the toughest requirement of all, given the Government's current stance on waiting lists and its continuing reluctance – Viagra apart – to acknowledge the limits to what the NHS can do.

5.4 Overall conclusions and recommendations

Looking back over the historical record, it is tempting to conclude that no real progress has been made on waiting lists since the NHS first became aware of a backlog of cases awaiting hospital treatment.

- an even bigger 'backlog' remains
- the way that waiting lists really operate in practice has been persistently ignored by policy-makers – the language used for the latest policy initiative ('targeting hospital black spots') has scarcely changed since the 1960s
- data on access to care remains poor, making it hard to establish the degree of inequity as between different groups of the population
- treatment thresholds have not been defined, despite clinical and political endorsement of their value
- variations in treatment rates remain between different parts of the country
- attempts to adopt a systematic approach to prioritisation remain local and small-scale.

The measures announced in *The New NHS* and now being gradually implemented give some grounds for optimism. Many of the elements for a coherent policy are beginning to emerge, particularly those concerned with standards of care, service and system design, and performance measures bearing on access as well as efficiency. In these areas the Government has moved far beyond any previous administration and, while progress will be slow, it is aiming to move the NHS in the right direction.

Unfortunately, in other respects, the Government's performance has fallen short of what is required. It continues to try to impose on the NHS a target generally agreed to be inappropriate, not simply because it emphasises numbers rather than the times most people have to wait, but also because it ignores the other elements of the whole system of which hospital waiting lists form part. While it has introduced measures of access into the NHS performance assessment framework, these comprise only a small number of procedures and will require considerably more investigation before they can be used to identify which are offering too little treatment and which too much. It has begun to commission research into waiting lists and booked admission systems, but has failed to set this work in a proper strategic context. Despite its rhetorical emphasis on a whole systems approach, the Government has failed to adopt it for elective care.

The first recommendation of this report, therefore, is that the Government should aim to integrate within one coherent framework the various strands of policy that bear on access to elective care. The framework adopted in this report represents a possible starting point. But we have not considered:

- the effectiveness of current preventive and screening policies, and possible extensions to them
- the administration of waiting lists, including the weeding out of those who longer seek care
- clinical performance in the execution of care
- the scale of 'over' or inappropriate treatment.

All these should form part of a coherent policy framework as indeed the Government has recognised in its proposals for national service frameworks.

To help bring these various strands together we suggest that the NHS should develop:

- a new form of reporting waiting times and numbers, using existing information but presenting it in a way which reveals more clearly what is happening in different parts of the country

- monitoring frameworks which cover the overall processes governing access to elective care and build on the growing potential for record linkage as patient records are put on an electronic basis
- equity audits for selected conditions where under, or late, referral is believed to occur – that judgement would in turn have to be based on systematic analysis of variations in access.

The data resulting from these various strands of work should be used to produce an annual report on the elective care system, which would bring together the impact of all the policies bearing on its performance as well as the results of external audit. It should include, on a selective but systematic basis, studies of the patients passing through the elective care system, monitoring their conditions pre- and post-treatment and the time taken to pass through all relevant stages. On the basis of this information, weaknesses at all stages of the system could be identified in general and in particular areas. Against this background, targets could be set for particular procedures or population groups that are more likely to promote the Government's broad objectives than those currently in place.

Our second recommendation is that substantial improvements in monitoring should be made to support the management of the elective care system, locally and nationally. Although existing sources of data could be exploited to better effect to shed light on the way that resources are being deployed within elective care, the earlier parts of this report have identified a number of areas and functions where new approaches are required. These will comprise new forms of data collection as well as research:

- an audit framework, which comprises complete care pathways, at least for some conditions, and which is based on systematic research on the experience of patients waiting for care and in the subsequent post-operative period, and on the factors giving rise to diagnostic and other delays within providers
- greater understanding of the interactions between the various demands falling on hospital resources, particularly the balance between elective and emergency work

- greater understanding of the behavioural links between the various elements of the elective care system, in particular the nature and scale of the deterrence effect of delay.

Over and above these requirements, a vast body of work is required – the detail of which will vary from condition to condition and procedure to procedure – which bears on the importance of delays and the prioritisation of different claims on the resources available for elective care. In part this is a technical matter, involving clinicians and others in analysis of the effects of different access times and pathways on patient outcomes. In part it is a political task, engaging professionals and public in debate about the general principles involved in making prioritisation and threshold judgements, as well as how these principles are reflected in specific scoring procedures.

Our third recommendation is that national waiting lists as currently defined should disappear from the national policy agenda to be replaced by a focus on access to hospital care. The emphasis needs to shift from what we have termed Stages 3 and 4 – waiting for hospital consultation and treatment – to the elective care system as a whole. The hospital waiting list should be seen as a means of monitoring how the whole system is performing, not as a policy target in its own right. Analysis of the characteristics and the condition of patients on entering and leaving the system, together with epidemiological information, can provide the basis for effective management of the elective care system as a whole. Thus, work of the kind carried out by the National Audit Office in the case of cataract should be carried out as a matter of routine for all common conditions.

Our final recommendation is that the role of private sector care should be explicitly defined. Its present role allows a small section of the population to escape 'rationing by waiting list' for both major and minor procedures, a possibility that other health care systems with waiting lists do not allow on the very equity grounds that the Government is explicitly committed to. Within the incremental option set out above, it would remain in this anomalous position. The radical option would, however, open the way to a principled division of roles in which the NHS was a universal provider for all those conditions and degrees of severity that were judged to merit the allocation of public funds, and the private sector focused largely on the rest. Such a division might be said to exist now. The difference, however, would lie in the evidence and debate on which public and private roles were determined.

Appendix 1

Literature searches

This appendix sets out the literature searches carried out by Lucy Johnson (Research Librarian in the King's Fund Library and Information Service) to support the preparation of this report. A series of searches were undertaken on a variety of databases to attempt to identify the literature on the subject of delay in relation to health care use. The literature relating primarily to waiting lists themselves was largely derived from the sources cited in the introduction to the bibliography.

The following searches were performed on the databases indicated, using the keywords shown. The searches were developed over a period of time as the research into the subject increased and more keywords became obvious. The searches were limited only by the dates of the databases, in that the most recent versions of each database were used (for example, Medline was not searched from 1966 onwards, but, rather, in most cases, from 1995 onwards) and by language, in that only references in English, French, German or Italian were retrieved. The searches are listed below, in the order in which they were undertaken:

Time to treatment

Database	**Search terms**
CINAHL	(time adj2 treat$).tw. and (emergencies.de. or emergency nursing.de. or exp accidents.de. or exp "wounds and injuries".de. or exp emergency service.de.) *tw=textword (found in title, abstract or subject terms)* *exp=instruction to "explode" CINAHL subject heading to include narrower terms as well* *de=CINAHL subject heading*
HMIC (containing DHdata and Helmis)	((time adj treat*) or (time near treat*) or (time to treat*) or (emergenc* in de) or (trauma* in de) or (wound* in de) or (injur* in de) or (stab* in de) or (accident* in de) or (casualty in de) or stab or stabs or stabbing) and (delay* or (wait* in de)) *de=subject terms*
King's Fund Unicorn Database	((time adj treat$) or (time near treat$) or (emergenc$ or trauma$ or wound$ or injur$ or stab$ or accident$ or casualty){SU} or stab or stabs or stabbing) and (delay$ or wait${SU}) *{SU}=subject terms*
Medline	(time adj2 treat$).tw. and (emergencies.de. or exp emergency service, hospital.de. or emergency nursing.de. or exp accidents.de. or exp "wounds and injuries".de.) and (english or french or german or italian) *tw=textword (found in title, abstract or subject terms)* *exp=instruction to "explode" MeSH term to include narrower terms as well* *de=MeSH term*

Health belief model and other models of health behaviour

Database	**Search terms**
CINAHL	(health adj belief adj model).tw. or (reasoned adj action).tw. or (protection adj motivation).tw. or (expected adj utility).tw.
HealthSTAR	((health adj belief adj model).tw. and (exp attitude to health.de. or models.hw.)) or (reasoned adj action).tw. or (protection adj motivation).tw. or (expected adj utility).tw. *hw=word included somewhere in the MeSH terms attached to a record*
HMIC	(health belief model*) or (reasoned action) or (protection motivation) or (subjective expected) or (expected utility theory)

King's Fund Unicorn Database	(health adj belief adj model*) or (reasoned adj action) or (protection adj motivation) or (subjective adj expected) or (expected adj utility adj theory)
Medline	((health adj belief adj model).tw. and (exp attitude to health.de. or models.hw.)) or (reasoned adj action).tw. or (protection adj motivation).tw. or (expected adj utility).tw. *hw=word included somewhere in the MeSH terms attached to a record*

Booking systems

Database	**Search terms**
HealthSTAR	(booking adj system$).mp. or ((booking$ or appointment$).mp. and (elective or refer$).mp. and exp great britain.de.) or (* "appointments and schedules".de. and system$.mp.) *mp=MeSH term, title word, word found in the abstract or source* **=focused MeSH term or major MeSH term*
HMIC	(booking system*) or ((appointment* or booking*) and (refer* or elective))
King's Fund Unicorn Database	(booking system*) or ((appointment* or booking*) and (refer* or elective))
Medline	(booking adj system$).mp. or ((booking$ or appointment$).mp. and (elective or refer$).mp. and exp great britain.de.) or (*"appointments and schedules".de. and system$.mp.) *mp=MeSH term, title word, word found in the abstract or source* **=focused MeSH term or major MeSH term*

Factors affecting the public's access to/take-up of emergency services

Database	**Search terms**
HealthSTAR	(((health adj belief adj model$).tw. or (exp health services accessibility.de. not health care rationing.de.) or choice behavior.de. or uptake.tw. or (utilisation or utilization).tw.) and exp emergency medical services.de.) or exp *emergency medical services/ut) and exp great britain.de. *ut=utilisation (a subheading)*
HMIC	(emergency in de or casualty in de or out-of-hours) and ((patient* adj access*) or (patient* adj choice*) or (patient* adj decision*) or uptake in de or access in de or choice in de or utilisation in de or (patient* adj (use or utilisation)))
King's Fund Unicorn Database	(emergency or (out hours)){SU} and ((patient$ adj access$) or (patient$ adj choice$) or (patient$ adj decision$) or uptake{SU} or access{SU} or choice{SU} or utilisation{SU} or (patient* adj (use or utilisation)))
Medline	(((health adj belief adj model$).tw. or (exp health services accessibility.de. not health care rationing.de.) or choice behavior.de. or uptake.tw. or (utilisation or utilization).tw.) and exp emergency medical services.de.) or exp *emergency medical services/ut) and exp great britain.de. *ut=utilisation (a subheading)*

Effects of treatment at different stages of disease

Database	**Search terms**
CINAHL	(((early or late) adj stage$).mp. or (early or late).ti. or (present$ and (stage$ or time or date)).tw.) and (exp *lens diseases.de. or *deafness.de. or *colonic neoplasms.de.) or (stage$.tw. and (exp disease.de. or presentation.tw,de. or *cataract.de. or *deafness.de. or *carpal tunnel syndrome.de.) or (stage$.tw. and exp disease.de. and presentation.tw,de and exp diagnosis.de.)
HealthSTAR	((exp *lens diseases.de. or *deafness.de. or (*colonic neoplasms.de. and ((course adj2 disease).tw. or ((early or late) adj stage$).mp. or present$.tw. or (stage$ adj2 disease).tw.)))) and (early or late).ti.) or (stage$.tw. and (exp disease.de. or presentation.tw,de. or *cataract.de. or *deafness.de. or *carpal tunnel syndrome.de.) or (stage$.tw. and exp disease.de. and presentation.tw,de and exp diagnosis.de.) *ti=word in the title*
HMIC	stage* and (disease* or illness* or condition* or cataract* or deaf* or rsi or (repetitive strain))
King’s Fund Unicorn Database	stage$ and (disease$ or illness$ or condition$ or cataract$ or deaf$ or carpal or rsi or (repetitive adj strain))
Medline	((exp *lens diseases.de. or *deafness.de. or (*colonic neoplasms.de. and ((course adj2 disease).tw. or ((early or late) adj stage$).mp. or present$.tw. or (stage$ adj2 disease).tw.)))) and (early or late).ti.) or (stage$.tw. and (exp disease.de. or presentation.tw,de. or *cataract.de. or *deafness.de. or *carpal tunnel syndrome.de.) or (stage$.tw. and exp disease.de. and presentation.tw,de and exp diagnosis.de.) *ti=word in the title*

Delay versus help seeking behaviour

Database	**Search terms**
CINAHL	((delay$.tw. or denial$.tw. or lagtime$.tw. or (lag adj time$).tw. or (lag or lagging or lags or lagged).tw. or ((help or care) adj seek$).tw.) and (*attitude to health.de. or *knowledge, attitudes, practice.de. or exp *health behavior.de. or *”appointments and schedules”.de.)) or ((unreport$ or ignor$) adj symptom$).tw. or (delay$ adj (help$ or care or seek$)).tw.
HMIC	((delay* or denial* or lag or lags or lagging or lagged or survival or worry*) and (patient-waiting-time in de or access* in de or belief* in de or behaviour* in de or psycholog* in de or treatment in de) or ((help or care) adj seek*) or (lag adj time*)) or ((access* in de or belief* in de or behaviour* in de or psycholog* in de or treatment in de) and patient-waiting-time in de) or ((unreport* or ignor*) adj symptom*) or delay* adj (help* or care or seek*)
King’s Fund Unicorn Database	((delay$ or denial$1 or lagtime or (lag adj time) or survival or ((help or care) adj seek$) or worry$ or (health adj journey$) and (access or belief$ or behaviour$ or psycholog$){su}) or (delay$ or denial$1 or lagtime or (lag adj time) or (survival adj instinct$) or ((help or care) adj seek$) or (health adj journey$))
Medline	((delay$.tw. or denial$.tw. or lagtime$.tw. or (lag adj time$).tw. or (lag or lagging or lags or lagged).tw. or ((help or care) adj seek$).tw.) and (*attitude to health.de. or *knowledge, attitudes, practice.de. or exp *health behavior.de. or *”appointments and schedules”.de.)) or ((unreport$ or ignor$) adj symptom$).tw. or (delay$ adj (help$ or care or seek$)).tw.

Time to treatment

Database	Search terms
CINAHL	(time adj treat$).tw. and (delay$ or wait$).tw.
HMIC	((time adj treat*) or (time to treat*)) and delay*
King's Fund Unicorn Database	(time adj treat$) and delay$
Medline	(time adj treat$).tw. and (delay$ or wait$).tw.

Health diaries

Database	Search terms
CINAHL	*diaries.de. or (dean-k$.au. or bentzen-n$.au. or christiansen-t$.au. or pedersen-k$.au.) *au=author name*
HealthSTAR	(health adj diar$).tw. or (dean-k$.au. or bentzen-n$.au. or christiansen-t$.au. or pedersen-k$.au. *au=author name*
HMIC	(health diar*) or self-record* or (self record*)
King's Fund Unicorn Database	(health adj diar$) or self-record$ or (self adj record$)
Medline	(health adj diar$).tw. or (dean-k$.au. or bentzen-n$.au. or christiansen-t$.au. or pedersen-k$.au. *au=author name*

Public expectations

Database	Search terms
HealthSTAR	((health or public or patient) adj expectation$).tw.
HMIC	expectations in de
King's Fund Unicorn Database	expectation$
Medline	((health or public or patient) adj expectation$).tw.

Late hospitalisation or diagnostic delay

Database	Search terms
HMIC	(late hospitali*) or (diagnostic delay*)
King's Fund Unicorn Database	(late adj hospitali$) or (diagnostic adj delay$)
Medline	(late adj hospitali$).mp. or (diagnostic adj delay$).mp.

Delay in the quality assurance literature

Database	Search terms
HealthSTAR	delay$.mp. and (joint commission on accreditation of healthcare organizations.de. or exp quality assurance, health care.de.)
HMIC	(delay* and quality* in de) not quality-of-life in de
King's Fund Unicorn Database	delay$ and (quality not (quality of life)){su}
Medline	delay$.mp. and (joint commission on accreditation of healthcare organizations.de. or exp quality assurance, health care.de.)

Delay and hospital re-engineering

Database	Search terms
HealthSTAR	delay$.mp. and (*hospital restructuring.de. or (reengineer$ or re-engineer$ or (re adj engineer$)).tw.)
HMIC	delay* and (re-engineer* or reengineer* or re engineer*)
King's Fund Unicorn Database	(reengineer$ or re-engineer$ or (re adj engineer$)) and hospital${su}
Medline	delay$.mp. and (*hospital restructuring.de. or (reengineer$ or re-engineer$ or (re adj engineer$)).tw.)

Self care

Database	Search terms
CINAHL	(*self care.de. or *self administration.de. or *self medication.de.) and (sn.fs. or td.fs.) *fs=floating subheading (sn=statistics and numerical data; td=trends)*
HealthSTAR	(*self care.de. or *self administration.de. or *self medication.de.) and (sn.fs. or td.fs.) *fs=floating subheading (sn=statistics and numerical data; td=trends)*
HMIC	self care in de or self treatment in de or self medication in de
King's Fund Unicorn Database	((self adj care) or (self adj treatment) or (self adj medication)){SU}
Medline	(*self care.de. or *self administration.de. or *self medication.de.) and (sn.fs. or td.fs.) *fs=floating subheading (sn=statistics and numerical data; td=trends)*

Bibliography

The literature on access to elective care is now dismayingly large and we have not in this paper attempted to provide a systematic review of it. Our aim, rather, has been to draw on it selectively to illustrate and underpin the systems approach we have adopted.

Listed below are references for all sources cited explicitly in the text. It includes a small number of major overviews, by R Tudor Edwards, Frankel and West and McDonald *et al.*, which are not cited extensively but on which we have drawn and which also contain extensive bibliographies.

Afzelius P, Zedeler K, Sommer H *et al.* Patient's and Doctor's Delay in Primary Breast Cancer: prognostic implications. *Acta Oncologica* 1994; 33: 345–51.

Age Concern England. *Equal Access to Cardiac Rehabilitation.* London: Age Concern, 1998.

Agnew TM, Whitlock RML, Neutze JM *et al.* Waiting lists for coronary artery surgery: can they be better organised? *The New Zealand Medical Journal* 1994; 107: 211–15.

Allison P *et al.* The role of professional diagnostic delays in the prognosis of upper aerodigestive tract carcinoma. *Oral Oncology* 1998; 34:147–53.

Anderson G *et al.* Willingness to pay to shorten waiting time for cataract surgery: What are the preferences of people actually on the waiting list for surgery? *Health Affairs* 1997, Sept/Oct.

Appleby J. The wait goes on. *Health Service Journal* 1998; 9 April: 34–35.

Arbman G *et al.* Short Diagnostic Delay is More Important for Rectal Cancer than for Colonic Cancer. *Eur J Surg* 1996: 162.

Audit Commission. *A short cut to better services: day surgery in England and Wales.* London: HMSO, 1990.

Audit Commission. *All in a day's work: an audit of day surgery in England and Wales.* London: HMSO, 1992.

Audit Commission. *The doctor's tale: the work of hospital doctors in England and Wales.* London: HMSO, 1995.

Audit Commission. *Day surgery follow up: progress against indicators from 'A short cut to better services'.* Audit Commission Update. London: Audit Commission, 1998.

Barrett B. *Access to elective care.* Unpublished.

Barton M *et al.* Does waiting time affect the outcome of larynx cancer treated by radiotherapy? *Radiotherapy and Oncology* 1997; 44:137–41.

Bell CM *et al.* Shopping Around for Hospital Services: A Comparison of the United States and Canada. *JAMA* 1998; 279: 1015–17.

Billings JS and Wells FC. Delays in the diagnosis and surgical treatment of lung cancer. *Thorax* 1996; September: 903–06.

Black N *et al.* International Variation in Intervention Rates: What Are the Implications for Patient Selection? *International Journal of Technology Assessment in Health Care* 1995; 11: 719–32.

Breen D. Setting Priorities: A Framework for the Assessment of Health Care Priorities in Scottish Health Boards. *Health Bulletin* 1991; 49: 34–39.

Bridgewater B. Death on the waiting list for cardiac surgery. *Heart* Editorial 1999; 81: 564.

British Medical Association. *BMA says waiting times are the true measure of NHS success or failure.* Press Release. London: BMA, 26 August 1998.

British Medical Association. *BMA welcomes recognition of doctors' work in bringing waiting times down.* Press Release. London: BMA, 9 November 1998.

British Medical Association. *Waiting lists need to be properly targeted.* Press Release. London: BMA, December 1998.

BMJ Publishing Group. *The Blind may see and the deaf may hear by the year 2050!*, Press Release, 3 December 1998.

Billing JS, Wells FC. Delays in the Diagnosis and Surgical Treatment of Lung Cancer. *Thorax* 1996; 51: 903–06.

Brattberg G. Priority Setting with Regard to Placement on Waiting List to a Pain Clinic: The Feasibility of a Delegated Ranking Procedure. *Scand J Soc Med 1988;* 16: 173–78.

Bridgewater B. Death on the waiting list for cardiac surgery. *Heart* 1999; 81: 564.

Caplan LS, Helzlsouer KJ. Delay in breast cancer: a review of the literature. *Public Health Rev* 1992/93; 20:187–214.

Carroll RJ *et al.* International Comparison of Waiting Times for Selected Cardiovascular Procedures. *JACC* 1995; 25: 557–63.

Clinical Standards Advisory Group. *Access to and Availability of Specialist Services.* London: HMSO, 1993.

Clinical Standards Advisory Group. *District Elective Surgery: Access to and availability of services.* London: HMSO, 1996a.

Clinical Standards Advisory Group. *Specialised Services (1996).* London: HMSO, 1996b.

Clinical Standards Advisory Group. *Cleft Lip and/or Palate.* London: HMSO, 1998.

Coast J. The role of economic evaluation in setting priorities for elective surgery. *Health Policy* 1993; 24: 243–57.

Committee of Public Accounts Thirty-Second Report. *Cataract Surgery in Scotland.* London: HMSO, 1998.

Coulter A. Partnerships with patients: the pros and cons of shared clinical decision-making. *Journal of Health Services Research and Policy* 1997; 2: 112–21.

Courtright P *et al.* Creation of Priority Criteria for Corneal Transplantation and Analysis of Factors Associated with Surgery Following Implementation. *Revue Canadienne de Santé Publique* 1997; 88: 320–24.

Courtright P, Poon CI, Richards JSF *et al.* Creation of Priority Criteria for Corneal Transplantation and Analysis of Factors Associated with Surgery Following Implementation. *Revue Canadienne de Santé Publique* 1997; 88: 320–324.

Cox JL *et al.* Managed Delay for Coronary Artery Bypass Graft Surgery: The Experience at One Canadian Center. *JACC* 1996; 27: 1365–73.

Cullis JG, Jones PR. Rationing by Waiting Lists: An Implication. *Journal of Health Economics* 1985; 4: 199–235.

David AS. *Determinants of Delay Seeking Treatment for Breast Symptoms.* London: Institute of Psychiatry, 1997.

Davies HTO, Crombie IK and Macrae WA. Waiting in pain: Delays between referral and consultation in outpatient pain clinics. *Anaesthesia* 1994; 49: 661–65.

Department of Health. *Working for Patients.* Cm. 555. London: HMSO, 1989.

Department of Health. *The Government's Expenditure Plans 1990–91 to 1992–93.* London: HMSO, 1990.

Department of Health. *Public Expenditure on Health Services; Waiting lists. Government Response to the First Report from the Health Committee Session 1990–91.* Cm. 1586. London: HMSO, 1991.

Department of Health. *Government Response to the Reports by the Clinical Standards Advisory Group on Access to and Availability of Specialist Services.* London: HMSO, 1993.

Department of Health. *The National Health Service: A service with ambitions.* Cm. 3425. London: HMSO, 1996.

Department of Health. *The Government's Expenditure Plans 1997–98 to 1999–2000.* Cm. 3612. London: HMSO, 1997.

Department of Health. *Report to the Chief Executive on winter pressures.* London: Department of Health, 1997a.

Department of Health. *The new NHS: modern: dependable.* Cm. 3807. London: TSO, 1997b.

Department of Health. *New Health Deal for People with Learning Difficulties. Health Secretary announces new guidance and advice.* Press Release 97/240. London: Department of Health, 1997c (22 September).

Department of Health. *A first class service: quality in the new NHS.* London: Department of Health, 1998.

Department of Health. *Modernising Health and Social Services: National Priorities Guidance 1999/00–2001/02.* London: HMSO, 1998.

Department of Health. *Additional resources in 1998–99 for 'whole systems' action to support sustained reductions in NHS waiting lists.* Leeds: Department of Health, 1998.

Department of Health. *Milburn steps up war against cancer.* PN1999/0631. London: Department of Health, 1999.

Department of Health, NHS Executive. *Quality and performance in the NHS: high level performance indicators.* London: Department of Health, 1999.

Department of Health and Social Security. *Reduction of Waiting Times for In-patients Admission: Management arrangements.* HSC(IS)181. London: HMSO, 1975.

Department of Health and Social Security. *Waiting for Hospital Treatment.* London: HMSO, 1980.

Department of Health and Social Security. *Health Services Development: Resource Assumptions and Planning Guidelines.* HC(87)7. London: HMSO, 1987.

Derrett S *et al.* Waiting for elective surgery: effects on health-related quality of life. *International Journal for Quality in Health Care* 1999; 11: 47–57.

Dickson R *et al.* Management of cataract. *Quality in Health Care* 1996; 5: 180–85, based on *Effective Health Care* 2 Bulletin No. 3.

Dixon J, Inglis S and Klein R. Is the English NHS underfunded? *British Medical Journal* 1999; 318: 522–26.

Drummond MF, Yates JM. Clearing the Cataract Backlog in a (Not So) Developing Country. *Eye* 1991; 5: 481–86.

Dunn E *et al.* Patients' Acceptance of Waiting for Cataract Surgery: What makes a wait too long? *Social Science and Medicine* 1997; 44:1603–10.

Earwicker SC, Whynes DK. General practitioners' referral thresholds and choices of referral destination: an experimental study. *Health Economics* 1998; 7: 711–22.

Edwards R Tudor. *NHS Waiting Lists: Towards the Elusive Solution.* London: Office of Health Economics, 1997.

Expert Advisory Group on Cancer. *A policy framework for commissioning cancer services.* The Calman-Hine Report. Leeds: NHS Executive, 1995.

Faccione NC. Delay versus help seeking for breast cancer symptoms: a critical review of the literature on patient and provider delay. *Social Science & Medicine* 1993; 36: 1521–34.

Feek CM, McKean W, Henneveld L *et al.* Experience with rationing health care in New Zealand. *BMJ* 1999; 318: 1346–48.

Frankel S, Coast J, Baker T, Collins C. Booked admissions as a replacement for waiting lists in the new NHS. *BMJ* 1991; 303: 1257–58.

Frankel S *et al.* Population requirement for primary hip-replacement surgery: a cross-sectional study. *Lancet* 1999; 353: 1304–09.

Gafni A. Time in Health: Can we measure individuals' 'Pure Time Preferences'? *Medical Decision Making* 1995; 15: 31–37.

Gafni A. Risk Attitude and Time Preference in Health. *Management Science* 1984; 30: 440–51.

George S, Read S, Westlake L *et al.* Differences in priorities assigned to patients by triage nurses and by consultant physicians in accident and emergency departments. *Journal of Epidemiology and Community Health* 1993; 47: 312–15.

Givio. Reducing Diagnostic Delay in Breast Cancer: Possible Therapeutic Implications. *Cancer* 1986; 58: 1756–61.

Goddard JA, Tavakoli M. Referral Rates and Waiting Lists: Some Empirical Evidence. *Health Economics* 1998; 7: 545–49.

Goddard MA and Smith P. *Equity of access to health care.* York: University of York, 1998.

Gray CS, Crabtree HL, O'Connell JE. Waiting in the dark: cataract surgery for older people. *BMJ* 1999; 318: 1367–68.

Hadorn D, Holmes A. The New Zealand priority criteria project. Part 1: Overview; Part 2: Coronary artery bypass graft surgery. In New B, editor. *Rationing: Talk and action in health care.* London: BMJ and King's Fund, 1997.

Hadorn D, Holmes AC. The New Zealand priority coronary criteria project. Part 1: overview. *BMJ* 1997; 314: 131.

Hadorn D, Holmes AC. The New Zealand priority coronary criteria project. Part 2: coronary artery bypass graft surgery. *BMJ* 1997; 314: 135.

Hadjistavropoulos HD *et al.* Measuring the Quality of Performance in the Management of Waiting Lists: Using Cataract Surgery as an Example. *Journal on Quality Improvement* 1998; 24: 407–22.

Hall JC *et al.* The Quality of life of patients on a waiting list for transurethral resection of the prostate. *J Qual Clin Practice* 1996; 16: 69–73.

Ham C, editor. *Health Care Variations: Assessing the evidence.* Research Report 2. London: King's Fund Institute, 1988.

Hamblin R, Harrison A, Boyle S. *Access to elective care: why waiting lists grow.* London: King's Fund, 1998.

Hamilton BH *et al.* What are the Costs of Queuing for Hip Fracture Surgery in Canada? *Journal of Health Economics* 1996; 15: 161–85.

Hanning M, Lundström M. Assessment of the Maximum Waiting Time Guarantee for Cataract Surgery: The Case of a Swedish Policy. *International Journal of Technology Assessment in Health Care* 1998; 14: 180–93.

Hannan EL *et al.* Access to Coronary Artery Bypass Surgery by Race/Ethnicity and Gender Among Patients Who Are Appropriate for Surgery. *Medical Care* 1999; 37: 68–77.

Harrison A, Dixon J, New B and Judge K. Is the NHS sustainable? *BMJ* 1997; 314: 296–98.

Harrison A. National Service Frameworks. In Klein R, editor. *Implementing the White Paper: pitfalls and opportunities: a King's Fund analysis of Labour's plans for the NHS.* London: King's Fund, 1998.

Harrison S, Hunter D. *Rationing Health Care.* London: IPPR, 1994.

Harvey I, Webb M, Dowse J. Can a surgical treatment centre reduce waiting lists? Results of a natural experiment. *Journal of Epidemiology and Community Health* 1993; 47: 373–76.

Heginbotham C. Why rationing is inevitable in the NHS. In New B, editor. *Rationing: Talk and action in the NHS.* London: BMJ and King's Fund, 1997.

Hippisley-Cox J. Are patients who present late with cancer registered with low referring practices? *British Journal of General Practice* 1997; 47: 731–32.

Ho E *et al.* Ontario Patients' Acceptance of Waiting Times for Knee Replacements. *The Journal of Rheumatology* 1994; 21: 2101–05.

Holmes A, Cumming A. *Briefing on Priority Setting No. 3.* Seminar presented 5-7 October 1998. Birmingham: University of Birmingham, HSMC.

House of Commons Health Committee. *Public expenditure on health services: waiting lists. Volume 1: First report [of the] Health Committee, together with the proceedings of the Committee.* London: HMSO, 1991a.

House of Commons Health Committee. *Public expenditure on health services: waiting lists. Volume 2: Minutes of evidence.* London: HMSO, 1991b.

House of Commons Health Committee. *Public Expenditure on Health Matters.* London: HMSO, 1991c.

House of Commons Public Accounts Committee. *Cataract surgery in Scotland: 32nd report [of the] Committee of Public Accounts.* London: TSO, 1998.

House of Commons Health Committee. *Primary care groups: 2nd report, together with the proceedings of the Committee and the minutes of evidence.* London: TSO, 1999.

House of Commons Health Committee. *Public expenditure on health and personal social services 1999.* London: TSO, 1999.

House of Commons Welsh Affairs Committee. *Elective Surgery, Volume I.* London: HMSO, 1991.

Howell GP *et al.* Long distance travel for routine elective surgery: questionnaire survey of patients' attitudes. *British Medical Journal* 1990; 300: 1171–73.

Hughes D, Griffiths L. 'But if you look at the coronary anatomy ...': risk and rationing in cardiac surgery. *Sociology of Health & Illness* 1996; 18:172–97.

Independent on Sunday. Doctors ration treatment for heart patients. Editorial, 21 June 1998.

Institute of Hospital Administrators. *Hospital Waiting Lists.* London: The Institute of Hospital Administrators, 1963.

Ivan Williams J, Llewellyn-Thomas H, Arshinoff R *et al.* The burden of waiting for hip and knee replacements in Ontario. *Journal of Evaluation in Clinical Practice* 1997; 3: 59–68.

Iversen T. A theory of hospital waiting lists. *Journal of Health Economics* 1993; 12: 55–71.

Johnstone F *et al. Deprivation and health in North Cheshire: an equity audit of health services.* Liverpool: Public Health Observatory, 1996.

Johnstone F *et al.* Delays in stroke referrals. *Lancet* 1999; 354: 47–48.

Jónsdóttir H, Baldursdóttir L. The experience of people awaiting coronary artery bypass graft surgery: the Icelandic experience. *Journal of Advanced Nursing* 1998; 27: 68–74.

Kasper JF *et al.* Developing shared decision making programmes to improve the quality of health care. *Journal of Quality Improvement, Quality Review Bulletin* 1992; 18: 183–96.

Kee F *et al.* Urgency and priority for cardiac surgery: a clinical judgement analysis. *BMJ* 1998; 316: 925–29.

Kee F *et al.* What is a safe waiting time for coronary artery bypass surgery? *Q J Med* 1997; 90: 669–76.

Klein R, editor. *Implementing the White Paper: pitfalls and opportunities: a King's Fund analysis of Labour's plans for the NHS.* London: King's Fund, 1998.

Klein R, Day P, Redmayne S. *Managing scarcity: priority setting and rationing in the NHS.* Buckingham: Open University Press, 1996.

Koch KT, Piek JJ, David GK *et al.* Does a waiting time for elective coronary angioplasty affect the primary success rate? *Heart* 1997; 77: 432–36.

Kontinnen M, Merikallio E. Return to work after open heart surgery: impact of long waiting lists? *Theoretical Surgery* 1990; 5: 192–95.

Langham S *et al.* The cardiac waiting game: are patients prioritized on the basis of clinical need? *Health Services Management Research* 1997; 10: 216–24.

Le Grand J, Mays N and Mulligan J-A. *Learning from the NHS Internal Market.* London: King's Fund, 1998.

Lenaghan J. *Rationing and Rights in Health Care.* London: IPPR, 1996.

Light D. *Effective Commissioning: Lessons from purchasing in American managed care.* London: Office of Health Economics, 1998.

Lindsay CM, Feigenbaum B. Rationing by Waiting Lists. *The American Economic Review* 1984; 74: 1984.

Llewellyn-Thomas HA *et al.* In the queue for coronary artery bypass grafting: patients' perceptions of risk and 'maximal acceptable waiting time', *Journal of Health Services Research & Policy* 1999; 4: 65–72

Loewenstein GF. Frames of Mind in Intertemporal Choice. *Management Science* 1988; 34: 1988.

London Implementation Group. *Report of an independent review of specialist services in London.* London: Department of Health, 1993.

Lundborn J *et al.* Factors influencing return to work after aortacoronary bypass surgery. *Scandinavian Journal of Cardiovascular Surgery* 1992; 26: 187–92.

Lundström M, Stenevi U, Thorburn W. Assessment of Waiting Time and Priority Setting by Means of a National Register. *International Journal of Technology Assessment in Health Care* 1996; 12: 136–40.

Lundström M *et al.* Vision related daily life problems in patients waiting for a cataract extraction. *British Journal of Opthalmology* 1994; 78: 608–11.

MacKeigan LD *et al.* Time Preference for Health Gains Versus Health Losses. *PharmacoEconomics* 1993; 3: 374–86.

Madhok R. Equitable Service Provision: A Feasibility Study of the Common Waiting Lists Model. *International Journal of Health Care Quality Assurance* 1994; 7: 30–31.

Mallender J, Elek R, Goswami N. Stern approach. *Health Service Journal* 1997; 107: 28–29.

Martin IG *et al.* Delays in the Diagnosis of Oesophagogastric Cancer: a consecutive case series. *BMJ* 1997; 314: 467–70.

Martin S, Smith PC. Rationing by waiting lists: an empirical investigation. *Journal of Health Economics* 1999; 71: 141–64.

Maynard A and Bloor K. *Equity in primary care.* CHS Discussion Paper 141. York: Centre for Health Economics, 1995.

McDonald P *et al.* *Waiting lists and Waiting Times for Health Care in Canada: More Management!! More Money??* Ontario: Queen's University, 1998.

McGarry J, Arthur A. Can over-75 health checks identify unmet need? *Nursing Standard* 1999; 13: 37–39.

Mills RP, Heaton JM. Waiting list initiatives: crisis management or targeting of resources? *Journal of the Royal Society of Medicine* 1991; 84: 405–07.

Ministry of Health. *Admission to Hospital: Hospital Waiting Lists and the Role of Community Services.* Circular 15/62. London: HMSO, 1962.

Ministry of Health. *Admission to Hospital and Hospital Waiting Lists.* HM(62)45. London: HMSO, 1962.

Ministry of Health. *Reduction of Waiting Lists.* HM(63)22. London: HMSO, 1963.

Mobb GE, Pugh F, Peeling B. How long is your waiting list? Experience of a urological waiting list initiative. *Journal of the Royal Society of Medicine* 1994; 87: 140–42.

Mordue A *et al.* Thresholds for treatment in cataract surgery. *Journal of Public Health Medicine* 1994; 16: 393–98.

Mordue A. New Zealand priority criteria project. *BMJ* 1997; 314: 1130.

Morgan CD, Sykora K, Naylor CD *et al.* Analysis of deaths while waiting for cardiac surgery among 29,293 consecutive patients in Ontario, Canada. *Heart* 1998; 79: 345–49.

National Audit Office. *Cataract Surgery in Scotland.* HC 275. London: HMSO, 1997.

Naylor CD. A different view of queues in Ontario. *Health Affairs* 1991; 110–28.

Naylor CD, Baigrie RS, Goldman BS *et al.* Assessment of priority for coronary revascularisation procedures. *The Lancet* 1990; 335: 1070–73.

Naylor CD, Ivan Williams J *et al.* Primary hip and knee replacement surgery: Ontario criteria for case selection and surgical priority. *Quality in Health Care* 1996; 5: 20–30.

Naylor CD, Levinton CM, Baigrie RS. Adapting to Waiting Lists for Coronary Revascularization: Do Canadian Specialists Agree on Which Patients Come First? *Chest* 1992; 101: 715–22.

Naylor CD, Levinton CM, Wheeler S, Hunter L. Queueing for Coronary Surgery During Severe Supply-Demand Mismatch in a Canadian Referral Centre: A Case Study of Implicit Rationing. *Social Science and Medicine* 1993; 37: 61–67.

Naylor CD, Slaughter PM. A stitch in time: case for assessing the burden of delayed surgery. *Quality in Health Care* 1994; 3: 221–24.

New B on behalf of the Rationing Agenda Group. The rationing agenda in the NHS. *BMJ* 1996; 312: 1593–1601.

New B, Mays N. Age, renal replacement therapy, and rationing. In Harrison A, editor. *Health Care UK 1996/97.* London: King's Fund, 1997.

Newton JN, Henderson J and Goldacre MJ. Waiting list dynamics and the impact of earmarked funding. *BMJ* 1995; 311: 783–85.

NHS Executive. *Guidelines for referral of patients with breast problems.* Sheffield: NHS Breast Screening Programme, 1995. (Second edition 1999)

NHS Executive. *Modernising Health & Social Services: National Priorities Guidance, 1999/00–2001/02.* Leeds: NHS Executive, 1998.

NHS Executive. *NHS National User Survey.* Leeds: NHS Executive, 1999.

NHS Executive. National surveys of NHS patients: General practice 1998. London: NHS Executive, 1999.

Nord E. Reducing sick leave costs by shortening waiting periods for elective surgery. *Medical Decision Making* 1990; 10: 95–101.

Nordberg M, Keskimäki, Hemminki E. Is there a relation between waiting-list length and surgery rate? *International Journal of Health Planning and Management* 1994; 9: 259–65.

North West Anglia Health Authority, King's Lynn and Wisbech Hospitals NHS Trust. *Orthopaedic Slots Initiative: Resource Pack.* Norwich: NW Anglia Health Authority, 1998.

Office of Health Economics. *Hip Replacement and the NHS.* London: OHE, 1982.

Pettinger N. Richly deserving. *Health Service Journal* 1999; 9 September: 20–21.

Plomp J, Redekop WK, Dekker FW *et al.* Death on the waiting list for cardiac surgery in the Netherlands in 1994 and 1995. *Heart* 1999; 81: 593–97.

Propper C. Contingent Valuation of time spent on NHS Waiting Lists. *The Economic Journal* 1990 (Conference); 100: 193–99.

Propper C. The Disutility of Time Spent on the United Kingdom's National Health Service Waiting Lists. *Journal of Human Resources* 1995; XXX (4): 678–97.

Rachlis MM *et al.* The vital risk of delayed coronary surgery: Lessons from the randomized trials. *Iatrogenics* 1991; 1: 103–11.

Ramirez AJ *et al.* Factors predicting delayed presentation of symptomatic breast cancer: a systematic review. *Lancet* 1999; 353: 1127–31.

Richards MA *et al.* Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet* 1999; 353: 1119–26.

Robinson R and Steiner A. *Managed health care: US evidence and lessons for the National Health Service.* Buckingham: Open University Press, 1998.

Roy CW, Hunter J. What happens to patients awaiting arthritis surgery? *Disability and Rehabilitation* 1996; 18:101–05.

Royal College of Surgeons of England. *Guidelines for the Management of Surgical Waiting Lists.* London: Royal College of Surgeons, 1991.

Royal Commission on the National Health Service. *Patients' attitudes to the hospital service.* London: HMSO, 1978.

Royal Commission on the National Health Service. *Royal Commission on the National Health Service: Report.* Chair: Merrison A. London: HMSO, 1979.

Sainsbury R *et al.* Effect on survival of delays in referral of patients with breast-cancer symptoms: a retrospective analysis. *Lancet* 1999; 353: 1132–35.

Saleh KJ *et al.* Immediate Surgery Versus Waiting List Policy in Revision Total Hip Arthroplasty: an economic evaluation. *The Journal of Arthroplasty* 1997; 12: 1–10.

Sanders D *et al. Variations in Hospital Admission Rates: a review of the literature.* London: King's Fund, 1989.

Sanderson CFB *et al.* Limitations of Epidemiologically Based Needs Assessment: The Case of Prostatectomy. *Medical Care* 1997; 35: 669–85.

Sarin S, Shields DA, Farrah J *et al.* Does venous function deteriorate in patients waiting for varicose vein surgery? *Journal of the Royal Society of Medicine* 1993; 86: 21–23.

Sarkies N *et al.* Indicator-Based audit of cataract surgery in four neighbouring hospitals in East Anglia. *Eye* 1995; 9(Suppl): 13–21.

Scottish Office. *£44 Million Drive to Cut Waiting Lists Revealed: New Team Will Target Hospital Black Spots.* News Release 0553/98. Edinburgh: The Scottish Office, 1998 (March 19).

Secretary of State for Wales. *Elective Surgery.* The Government's Response to the Sixth Report of the Welsh Affairs Committee, Session 1990–91, Cm. 1863. London: HMSO, 1992.

Seddon ME, French JK, Amos DJ *et al.* Waiting times and prioritisation for coronary artery bypass surgery in New Zealand. *Heart* 1999; 81: 586–92.

Sherlaw-Johnson C. Managing waiting lists for cardiac surgery. *Heart.* Editorial. 1999; 81: 563.

Stewart M, Donaldson LJ. Travelling for earlier surgical treatment: the patient's view. *British Journal of General Practice* 1991; 41: 508–09.

Tennant A *et al.* Prevalence of knee problems in the population aged 55 years and over: identifying the need for knee arthroplasty. *BMJ* 1995; 310: 1291–93.

Umeh HN *et al.* Impact of a Waiting List Initiative on a general surgical waiting list. *Annals of The Royal College of Surgeons of England* 1994; 76(Suppl): 4–7.

van Ackere A and Smith PC. *Making static models dynamic: the case of the NHS.* University of York Centre for Health Economics, 1997.

Varekamp I, Meiland FJM, Hoos AM *et al.* The meaning of urgency in the allocation of scarce health care resources; a comparison between renal transplantation and psychogeriatric nursing home care. *Health Policy* 1998; 44: 135–48.

Vissers JMH, van der Bij JD, Kusters RJ. Waiting Lists: A Multi-Level Demand-Supply Modelling Approach. Unpublished paper.

Wall P *et al.* Diagnostic in Breast Disease: A System Analysis of a Public Urban Hospital *Arch Surg* 1998; 133: 662–66.

Williams B. Utilisation of National Health Service Hospitals in England by Private Patients 1989–95. *Health Trends* 1997; 29: 21–25

Williams JI, Llewellyn-Thomas H, Arshinoff R, Young N, Naylor CD & the Ontario Hip and Knee Replacement Project Team. The burden of waiting for hipand knee replacements in Ontario. *Journal of Evaluation in Clinical Practice* 1997; 3: 59–68.

Wordsworth S, Donaldson C, Scott A. *Can We Afford the NHS?* London: IPPR, 1996.

Yates J. *Why are we waiting?* Oxford: Oxford University Press, 1987.

ISBN 1-85717-299-X

King's Fund Health Systems series