

• SUSTAINING REDUCTIONS IN WAITING TIMES: IDENTIFYING SUCCESSFUL STRATEGIES

Final report to the Department of Health

**John Appleby, Seán Boyle, Nancy Devlin, Mike Harley,
Anthony Harrison, Louise Locock, Ruth Thorlby**

**SUSTAINING REDUCTIONS IN WAITING TIMES:
IDENTIFYING SUCCESSFUL STRATEGIES**
Final report to the Department of Health

John Appleby, Seán Boyle, Nancy Devlin, Mike Harley, Anthony Harrison, Louise Locock,
Ruth Thorlby

This work was undertaken by the King's Fund who received funding from the Policy Research Programme of the Department of Health. The views expressed in the publication are those of the authors and not necessarily those of the Department of Health.

This research would not have been possible without the co-operation of the managers, clinicians and administrative staff of the nine trusts that agreed to take part.

About the authors

John Appleby is Chief Economist at the King's Fund and Visiting Professor at City University.

Seán Boyle is Visiting Senior Research Fellow at LSE Health and Social Care, London School of Economics, London.

Nancy Devlin is Professor of Health Economics at City University, London and Visiting Fellow at the King's Fund.

Mike Harley is a Performance Specialist at the Inter-Authority Comparisons and Consultancy, Health Services Management Centre, University of Birmingham.

Anthony Harrison is a Senior Fellow at the King's Fund.

Louise Lcock is a Lecturer at the Department of Social Policy and Social Work, University of Oxford.

Ruth Thorlby is a Health Policy Researcher at the King's Fund.

Published by
King's Fund
11–13 Cavendish Square
London W1G 0AN
Tel: 020 7307 2591
Fax: 020 7307 2801
www.kingsfund.org.uk

© Department of Health 2004

Charity registration number: 207401

Edited by Eleanor Stanley
Typeset by Lucy Latchmore

Contents

Executive summary	1
1. Sustaining reductions in waiting times	1
2. Waiting times targets and distortions to clinical priorities	7
3. A framework for system-based information requirements for the management of the supply of elective care	11
Introduction	15
Why this study?	15
Study aims	17
Overview of this paper	17
Conceptual background	18
How the elective care system works	18
What counts as a reduction in waiting times?	20
Previous studies	22
1. Sustaining reductions in waiting times	25
Research objective	25
Research design, methods and expectations	25
Findings	32
Overall findings	33
Specific factors and strategies	38
Summary	42
Conclusions	44
Key elements for managing the elective care system	45
2. Waiting times targets and distortions to clinical priorities	48
Background	48
Aims	51
Methodology	51
Survey of consultants in eight trusts	52
Indications of clinical ‘distortion’	57
Summary	79
Conclusions	81
3. A framework for system-based information requirements for managing the supply of elective care	83
Background	83
The system perspective	86
Policy environment	88
Understanding demand	92
The production of care: outpatient services	94
The demand for elective operations	95
The production of care: elective operations	97
Discussion	99
Further research	100

References	102
Appendix 1: Terms of reference, original research proposal	104
Appendix 2: Selection criteria for trusts included in the study	105
Appendix 3: Interview schedule	107
Appendix 4: Survey questionnaire	111
Appendix 5: Examples and comments from the consultants' questionnaire	116
Appendix 6: The outpatient production process	119
Appendix 7: The elective inpatient production process	122

Executive summary

This summary brings together the findings and conclusions from three strands of research that addressed different aspects of the main research topic, which is the identification of successful strategies for sustaining reductions in waiting times:

- sustaining reductions in waiting times: identifying successful strategies
- the impact of waiting times targets on clinical treatment priorities
- a framework for system-based information requirements for the management of the supply of elective care.

These three strands are detailed in full below.

1. Sustaining reductions in waiting times

Policy context and study aims

The aim of this first part of the study was to isolate factors that lead to sustainable reductions in waiting times.

Conceptual background

Earlier King's Fund studies by Harrison and New (2000) and Hamblin, Harrison and Boyle (1998) described various initiatives to tackle waiting lists and waiting times taken since the foundation of the NHS. The historical record showed that numbers waiting had risen over time, and that any improvements had been short lived. There had been some success in reducing very long waiting times but average waiting times had changed very little.

Harrison and New argued that these policies had been based on the incorrect view that waiting lists represented a backlog that was ad hoc, and that could be removed by temporary initiatives, such as those that characterised policymaking for much of the post-war period.

Sustainable reductions, as opposed to ad hoc reductions, must rest on the indefinite continuation of policies designed to respond to a range of forces – that is, to meet a level of demand that rises in response to technical change, demography, rising user expectations, and changes in clinical behaviour.

Research objective, design and methods

The main research objective was to identify strategies adopted by those trusts that appear to have been successful in not only reducing inpatient or day case waiting times, but in sustaining the reductions achieved.

In its first stage, the study opted for an in-depth qualitative analysis of short- and long-wait trusts, together with collection and analysis of a range of data concerning the performance of trusts, and other information, a priori, that was considered important.

Nine trusts were invited to participate in the research, with three trusts in each of the three following categories:

- successful – consistently low proportions of patients waiting over six months
- variable performance – some success in reducing the proportion of people waiting over six months, but not sustained
- unsuccessful – consistently high proportions of patients waiting over six months.

Through semi-structured interviews and the collection of trust, specialty and, where appropriate, consultant-level data, the study aimed to identify patterns of activity, resources, management and clinical policies, processes, attitudes, behaviours and strategies, as well as contextual factors that characterised the three groups of trusts, and thus to isolate the factors that explain sustained waiting times performance.

Findings

Five broad themes or issues were identified from the interview data, together with four more detailed factors that appeared important in separating ‘successful’ from ‘unsuccessful’ trusts:

- understanding whole systems
- the importance of sustained action over time
- catch up, keep up
- unexpected shocks
- clinical ownership and involvement.

The five broad themes are described below.

Understanding whole systems

Trusts with a poor record in reducing waiting times had a poor understanding of the way that improvement in waiting time performance depended on measures taken in other parts of the hospital, and also on the wider health economy. This relative lack of understanding also applied historically to those trusts that used to have a poor record on waiting times but had started to improve.

The situation in ‘successful’ trusts revealed the converse: that is, not only a reasonably good sense of the whole system of care but also an appreciation of the importance of such an understanding, which was reflected in the specific measures they took to achieve government targets.

The importance of sustained action over time

Successful trusts started to address the task of reducing waiting times in a systematic way much earlier than unsuccessful trusts, and had persevered with the task.

Unsuccessful and temporarily successful trusts, on the other hand, had by their own admission only really started to ‘get going’ with waiting times reductions in the previous 18–24 months. These trusts had also tended to rely on ad hoc initiatives, such as weekend working and other measures that could not be sustained indefinitely, and that often depended on time-limited injection of funds.

Catch up, keep up

Factors necessary to reduce waiting times are not always the same (or of the same importance or scale) as those involved in sustaining reductions. Catching up is not necessarily the same as keeping up.

The need to ringfence elective activity or manage demand – for example, through, say, referral protocols – is less relevant once waiting times are so low that all referrals can be quickly processed.

Unexpected shocks

Unsurprisingly, even where there is an appreciation of the whole-systems nature of the waiting times reduction issue, external shocks can upset even the best laid plans.

Reorganisations (the introduction of primary care trusts, for example), mergers, wholesale changes in senior management teams, and the need to meet financial targets can all knock a trust off its waiting times reduction course.

Clinical ownership and involvement

Those traditionally responsible for managing the workload of a hospital (and hence waiting lists and times) – consultants – are central to the job of reducing waiting times.

For example, in a number of unsuccessful and temporarily successful trusts, there were individual consultants who maintained short (six months or under) maximum waiting times. This suggests that good or bad performance depended to some degree on individuals rather than the effectiveness of the hospital management as a whole.

Pressure from above on managers to meet waiting times targets can be dissipated at local level, especially where relations between management and consultants are poor, or where consultants' objectives with regard to their work are not fully aligned with the objectives of the organisation for which they work.

Repeated use of one-off initiatives could create a culture in which medical, and some other, staff expected to be paid extra for doing waiting-list work and had come to rely on the additional income. This was proving an obstacle in trying to change people's thinking to see waiting time reduction (and in the longer term, consistently short waits in line with central targets) as a mainstream activity, and part of everyone's normal daily work.

Appealing to clinicians' motivation to do the best for their patients was more productive than top-down instruction.

Peer discussion and comparison were ways of encouraging poor performers to do better. Part of the problem was that consultants are rarely given sufficient information to judge for themselves what the problem is and how they might tackle it.

Concerns were sometimes expressed about the possible distortion of other clinical priorities as a result of the attention and effort given to waiting times reduction, even though the reasons for it were understood and felt to be worthwhile. (*See the second part of this research for further information*). However, sites with consistently low waiting times showed clear differences. In these sites, there was no need to delay care for some individual patients in order to keep up with waiting time targets for less urgent cases,

as everyone could be seen within a reasonable time. Any concerns about the priority given to waiting times in these sites were more in terms of other service developments that might have been foregone as a result of expenditure on waiting times.

Some trusts noted the positive impetus to achieving waiting times reductions through 'quick wins'. For example, in one trust, successfully reducing very long waits in ophthalmology clinics through a combination of extra staff and levelling out the workload across three clinics demonstrated that a longstanding problem could be tackled and helped foster a 'can-do' attitude.

More detailed factors were grouped under the following four headings:

- analysis, forecasting and planning
- organisational focus and persistence
- capacity
- efficiency of the production process.

These are described in detail below.

Analysis, forecasting and planning

An overwhelming consensus was apparent, both from successful and from temporarily or partially successful trusts (and less so from poorly performing trusts) of the need for information that was reliable, detailed, comparative and continuous (daily or even hourly).

To characterise it bluntly, successful trusts would not find it hard to produce waiting times information for a named patient, while unsuccessful trusts found it hard to know whether to trust their own total waiting list figures.

Successful trusts know (and others are beginning to realise) that tracking individual patients through the hospital system is vital.

There were a number of examples where managers had collated comparative waiting times and other performance data at the level of individual consultants. The first step in persuading consultants to change their working practices has been discussing the variations that such data reveals.

The need for information has been a strong driver for centralising waiting list management in successful trusts. All trusts recognised the need for much greater control over the flows and movements of patients into, around, and out of their hospital.

'Centralisation' does not mean a complete management takeover of the referral and operating list processes. In one case, it meant having one computerised office for admissions clerks, a standard 'Dear Doctor' referral letter (to help even out outpatient clinic workloads), and a version of 'earned autonomy' for consultants (with those managing their lists in a way that met their targets being autonomous, and others agreeing to have their lists managed and 'profiled' on their behalf, using, for example, CheckList).

In trusts where waiting times are consistently low and consultant workflows are already well managed, the need for centralisation may be less relevant or may take different forms.

Planning in successful trusts meant being ahead of the game – in particular, looking further ahead than the next looming waiting times milestone, and engaging in detailed capacity planning for the subsequent target.

Successful trusts used capacity planning models and ensured that they had access to the right information to plan for changes in demand and consequent changes in capacity. Successful trusts also gave examples of how they not only tried to match capacity prospectively with planned workload but also undertook retrospective reviews of what had actually happened, and analysed reasons for any discrepancies.

Organisational focus and persistence

Commitment and everyday involvement from the very top of the organisation was seen to be absolutely necessary in making progress on what all have noted are very tough targets.

Organisational focus and persistence includes the need to attract and retain experienced and skilled managerial staff – particularly directors of operations, or others with the main operational responsibility for meeting waiting times targets.

Managerial tactics to persuade clinicians to own and internalise a commitment to reducing waiting times were almost as varied as the number of consultants. However, the use of comparative consultant-level waiting times and performance data (shared with consultants) was important, as was the argument that reducing waiting times was not just a government target but was what patients wanted (and was good for their health).

Capacity

Having the resources to increase capacity (where it was identified as necessary) was clearly important – with the converse (in other words, not having the resources) being seen as almost a guarantee of failure.

All trusts stated that previous ad hoc or one-off uses of such resources had not led to sustainable reductions.

Temporary increases in capacity were essential as a short-term strategy to meet targets, but were often wasteful and expensive, and prevented the same money being invested in permanent capacity.

Efficiency of the production process

Some trusts had made use of short-term initiatives and had come to accept that they could not be sustained in the long run and were, in any case, expensive in terms of cost per case.

In contrast, successful trusts had begun to look in detail at the logistics of their hospital's care processes. This involved looking at the patient pathway and attempting to simplify and shorten it, identifying bottlenecks and pinch-points for the individual pathway, and then using the whole-hospital system perspective to work out, for example, the best way of handling the interaction between elective and emergency flows.

Within these broad strategies, successful trusts employed a host of smaller measures bearing on efficiency, including tight bed management, maximising day case activity,

ensuring full utilisation of theatres, and effective discharge planning, including possible investment in step-down facilities where the local private sector was inadequate.

Conclusions and recommendations

While this part of the research has been able to identify sets of factors that appear to be more common among successful trusts, we cannot attribute any magnitude of effect to each. We do not know the importance of that factor individually, nor whether it is likely to be effective on its own, or only in interaction with other factors in the set.

In summary, there appear to be four particularly important factors that account for variations in achieving and sustaining reductions in waiting times:

- a sustained focus on the task
- an understanding of the nature of waiting lists
- detailed information, analysis, forecasting, monitoring and planning
- development of appropriate capacity.

These factors are explained in detail below.

A sustained focus on the task

A clear and unambiguous message from successful trusts (and from those beginning to turn the corner on reducing waiting times) was the absolute necessity, first, to focus the organisation on reducing waiting times, and second, to sustain management and clinical effort and priorities on the task.

The energy and detailed day-to-day management (down to tracking individual patients through the hospital system) should not be underestimated. Bringing about this sustained focus requires skilled and strong, but sensitive, leadership and management at all levels of the trust.

An understanding of the nature of waiting lists

Understanding that waiting lists are not simply a backlog problem but that they are the manifestation of a more complicated, dynamic flow through interconnected parts of a whole system of care has enabled successful trusts to break down the problem, and to tackle those particular factors that, given their own circumstances, have given rise to long waiting times.

Detailed information, analysis, forecasting, monitoring and planning

Detailed, consistent and accurate time-series and cross-sectional information on waiting lists and times, as well as on key resources, provided successful trusts with a means to analyse and understand their waiting lists, to see them in context with other trusts, and to allow them to monitor progress and outcomes of changes in service delivery. This information also enabled them to plan future changes in services to meet targets, and to find the resources required to provide them.

Development of appropriate capacity

Lack of capacity can ultimately undermine efforts to reduce waiting times. Developing appropriate capacity – not just through increasing the totality of resources, but also

through more efficient use of resources and managing the demand on those resources – is essential.

2. Waiting times targets and distortions to clinical priorities

Policy context and study aims

Arising from the first part of this study was a question concerning the extent to which consultants felt that their clinical priorities for deciding the admission of patients from their waiting lists were being ‘distorted’ in attempts to meet the maximum waiting times targets.

With little previous quantitative research into this issue, either in the UK or abroad (where similar concerns have been expressed by consultants), it was decided that it would be useful to attempt to quantify any possible distortion to clinical priorities.

Research objectives, design and methods

The main research objective was to quantify the scale of any distortion to clinical admission priorities in one specialty: trauma and orthopaedics. Three investigative methods were used:

- a survey of all consultants in four specialties: trauma and orthopaedics, ear nose and throat (ENT), general surgery and ophthalmology, in nine trusts previously recruited to take part in the first stage of this research
- modelling of ‘expected’ admissions distributions of trauma and orthopaedic hospital episode statistics for all admissions during 2001/02 (a year in which the 15-month waiting time target prevailed), in order to identify ‘excess’ admissions (that is, where observed admissions exceeded ‘expected admissions)
- a ‘before-and-after’ (or ‘policy-off, policy-on’) comparison of waiting times distributions for 1997/98 (before waiting times targets were in place) and 2001/02 (after targets were introduced), in order to identify possible differences in the pattern of admissions (with the assumption that any differences could be attributed to the introduction of the waiting times target).

Findings

Survey of consultants in nine trusts

Given the small size of the sample, the following results from the survey of clinicians should be taken as indicative only.

One-third of consultants surveyed felt that they had had to treat patients out of clinical order (as they saw it). Very few, however, could quantify this apparent distortion in terms of numbers of patients affected.

Around 80 per cent (18) of those who stated that their priorities had been distorted also stated that the impact on delayed patients had been clinically negative. Around 40 per

cent (9) stated that there had been some positive clinical impact on patients being seen sooner than the consultant would have done in the absence of the waiting times target.

Only one consultant (out of 14) from trusts traditionally performing well on waiting times felt their clinical priorities had been affected as a result of the waiting times target, whereas around 42 per cent of consultants (23) from the other two groups of trusts (traditionally performing less well on waiting times) stated that their clinical priorities had been adversely affected by waiting times targets for 2002/03. These groups were also more liable than those trusts performing well to have taken a range of actions – extra sessions at weekends, and so on – to tackle long lists.

The survey confirmed the findings of much previous research that there is no standard clinical opinion concerning how long it is reasonable to expect urgent and routine patients to wait. Consultants stated that routine patients could reasonably wait between one and 18 months, with a mode at six months and sub-modes at three and 12 months. For urgent cases, there was more agreement, with the majority stating up to one month as reasonable, but the variation ranged from no wait at all to up to three months.

Consultants at trusts performing well on waiting times tended to indicate shorter times as a reasonable wait, both for urgent and routine cases, than consultants at trusts not performing so well.

Modelling 'expected' waiting time distributions

The particular pattern of waiting prior to admission in 2001/02 raised five questions:

- Can the apparent additional admissions for those waiting around 12 and 15 months be quantified? That is, what would we expect the waiting times distribution of those admitted for treatment to look like? Quantifying the apparent 'additional' admissions would provide a rough guide to an estimate of the scale of possible distortionary effects of maximum waiting times targets.
- Assuming an expected distribution can be calculated (and hence the additional admissions quantified) is it possible to determine whether additional admissions displaced other types of patients, either in terms of the types of procedures carried out, or in terms of the time patients waited?
- Given that the 2001/02 maximum waiting times target deadline was March 2002, is there a 'deadline effect', with additional patients tending to be admitted in the fourth quarter of the year? If so, this would indicate problems with the way trusts tackled the task of meeting the waiting time target – concentrating effort (and hence, perhaps, increasing the likelihood of distorting clinical priorities) into the weeks and months prior to the March target deadline.
- As the 2001/02 target was that no patients should be waiting more than 15 months, why is there also an apparent increase in admissions for patients waiting around 12 months (and to a lesser degree, 18 months)?
- How does the 2001/02 waiting times distribution compare with a period before the implementation of the 15-month target? A difference between distributions – particularly for admissions of patients waiting around 15 months – would lend support to the case that the 2001/02 maximum waiting times target affected admission behaviour.

The distribution of the length of time that patients waited prior to being admitted to hospital in 2001/02 exhibited 'blips' at 12, 15 and 18 months. Omitting data for admissions between 40 and 80 weeks and fitting 'best-fit' curves provided an estimate of the size of these blips.

This analysis suggested that trusts and their orthopaedic consultants responded to the 2001/02 15-month maximum waiting time target in part by altering the proportion of their waiting list admissions coming from different wait-time categories. Compared with an estimated expected distribution of admissions of patients waiting between 40 and 80 weeks, admissions in 2001/02 were around 38,000 higher than expected.

The composition of these extra admissions was found to differ from those who had not waited between 40 and 80 weeks prior to admission. Twelve procedures accounted for 77 per cent of 'excess' admissions, with arthroscopies and knee and hip replacements accounting for nearly half of these. These same procedures, plus soft tissue and other bone procedures (Category 1, for those aged under 70 without complications) alone increased their share of admissions in patients waiting 40–80 weeks. It did not seem to be the case, therefore, that the apparent extra admissions of patients waiting 40–80 weeks were long-wait minor cases.

Examination of the timing during the year when short- and long-wait patients were admitted revealed that the proportion of weekly admissions of long-wait patients (both 40–80 weeks and 56–65 weeks; in other words, around 15 months) rose from around October towards the end of the financial year. However, the proportions of patients waiting between one and four weeks did not change in any responsive way. This analysis also revealed that holidays – particularly the Christmas week – produced very large distortions in admissions affecting all types of patients, no matter how long they eventually waited for admission.

A 'deadline effect' was also notable, with 'excess' admissions of patients waiting around 15 months, peaking in the last quarter of the year as the target deadline loomed. While this may have contributed to an increased likelihood and feelings on the part of consultants that their clinical priorities were being distorted in the weeks before March 31, subsequent changes to the monitoring of the target may have dealt with this problem.

'Before-and-after' comparison

While peaks in admissions for 2001/02 were found at 15 months, there were also peaks at 12 and, to a lesser extent, 18 months. Comparison of waiting times distributions for 1997/98 revealed very similar peaks at these latter two dates, but not at 15 months, confirming the likelihood that it was the 15-month target in 2001/02 that was responsible for this change in admissions.

Subtracting the 1997/98 from the 2001/02 distribution suggested an estimate of the number of excess admissions of patients waiting around 15 months prior to admission of 9,333 (2.2 per cent of all admissions) – probably a more accurate estimate of the impact of the 2001/02 target than previously calculated using statistical estimates of 2001/02 waiting times distributions, as the 12-month peak in admissions cancelled out in the comparison.

Whether these apparent 'excess' admissions were at the expense of other patients – in particular, more urgent cases – is very difficult to answer with aggregate national data. Examination of the composition of these admissions did not immediately suggest either

that more minor cases were substituted for more major cases (in fact, it appears the opposite was true, with hip and knee replacements dominating these admissions) nor that very short-wait (and hence most urgent) cases gave way to longer-wait (presumably less urgent) patients.

However, for hip and knee replacements (the two procedures accounting for a majority of 'excess' admissions of patients waiting around 15 months), while admissions increased for virtually all categories of patients, proportionally fewer patients waiting between one and 44 weeks were admitted in 2001/02 than 1997/98. Whether this can be taken as proof of a substitution between long- and short-wait patients is difficult to answer, however.

By definition, of course, if the proportion of admissions of one group of patients increases, the proportion of another group must fall. But actual numbers of admissions of virtually all groups of patients rose between 1997/98 and 2001/02, which complicates any interpretation of the change in proportions of patients admitted.

When 2001/02 was compared with 1997/98, while it was found that the number of patients who had waited one week prior to admission fell, it seems very unlikely that this difference was due to the 'excess' 15-month-wait patients usurping their position on the waiting list. On balance, there are likely to be other explanations for this difference.

Quantifying the impact of the 15-month waiting times target in the absence of an admission-criteria 'gold standard' is difficult. Of two approaches used, a 'before-and-after' comparison of the waiting times distributions for trauma and orthopaedics for 2001/02 and 1997/98 probably provided a more accurate picture of the impact on admissions from the waiting list.

Using national data, it has not been possible to show unambiguously that the admission peak at around 15 months does represent clinically relevant distortions. However, the figures do provide grounds for suggesting that these were not of major importance.

Urgent cases do not appear to have been displaced – holidays had a greater impact. If cases have been deferred to meet targets, they are likely therefore to have been less urgent ones (for example, from our analysis, those waiting between 33 and 52 weeks) and the scale of their deferral, relative to their average wait, modest.

The extra cases largely consisted of significant operations (hip and knee replacements, for example) rather than minor ones, and hence no evidence was found of substitution of lesser cases for more serious ones.

The form of target prevailing during 2001/02 meant that for most of the year the risk of distortion was absent.

Moreover, although the 15-month peak in admissions appears to be a new phenomenon, in other respects the distribution of waiting times does not appear to have changed greatly following the introduction of the 15-month target.

Our overall conclusion, therefore, is that serious and extensive clinical distortions are likely to have been fairly limited. However, we recognise that the use of national aggregate statistics may understate the problems at local level.

Conclusions and recommendations

This research has only been able to provide indications of the possible scale of distortions to clinical priorities as a result of attempts to meet maximum waiting-time targets. We have outlined the research that would need to be carried out from which more definitive conclusions could be drawn.

Our line of investigation could however be pursued further, to look at other specialties or sub-samples of the data (for example, divisions based on waiting list length or time, or other factors that might explain variations in reported distortions).

This latter sub-sample analysis may help answer a key question: as waiting times targets become progressively shorter, will reports from consultants of distortions become more common?

In terms of the nature of maximum targets, it may be useful to explore possibilities for local variations in targets based on experience from abroad (for example, rather than 100 per cent compliance with a target, to set low levels of 90 per cent or 85 per cent together with local review mechanisms to ensure that patients do not end up languishing on lists.

3. A framework for system-based information requirements for the management of the supply of elective care

Policy context and study aims

The final stage of this research addressed the issue of information requirements at trust level, for formulating strategies to deal with waiting times reduction.

The first stage of this research had strongly indicated that effective management of the provision of elective care required an understanding of the whole system of elective care as part of the whole hospital system, and the hospital system as part of the local health economy. The findings showed that this required an analytic approach (and implied the collection and use of information) that was not apparent in its entirety in any of the hospitals we looked at – even those whose waiting times performance was, and has remained, relatively good.

Research objectives

The main research objective was, therefore, to elaborate the ‘systems’ view of the elective care system and, as a consequence, to identify key information requirements for hospitals in support of the goal of reducing inpatient waiting times.

Findings

Understanding the whole system

From a whole-systems point of view, we considered the elective care system to be embedded in, and interconnected to, three different levels: the health economy, the hospital and the specialty.

- **The health economy** The key systems issues identified at this level were:
 - the likely developments in the supply of elective care services outside the hospital
 - how these developments may be influenced by the hospital's own decisions to expand or reduce capacity in the light of its financial position
 - how the hospital's own decisions affect demand for its services
 - how GP referral decisions affect the demand for elective surgery.
- **The hospital** Key systems issues identified at this level were:
 - the extent to which it is feasible and economic to separate the elective care system from the rest of hospital activity
 - the best means of providing for variations in demand
 - in areas where elective work is not isolated, the nature and scale of linkages: joint staff, joint facilities, cross referrals
 - how out-of-hours arrangements affect demand for emergency service.
- **The specialty** The systems issues at this level were:
 - the way in which referrals will respond to reductions in waiting times for outpatient consultations and for treatment
 - how decisions to treat (that is, the conversion ratio) respond to changes in waiting times.

Understanding demand

There are two broad categories of elective care services provided by the hospital: outpatient consultations and elective operations. Demand for the former arises from referrals to a specialist, which gives rise to an outpatient appointment. Demand for the latter results from decisions by specialists, in consultation with the patient, usually during one or more outpatient consultations.

It is only by having a clear view of the nature of the demand for its services that the hospital will be able to manage and adjust its capacity successfully, to meet the variations in this demand that will occur. Moreover, the hospital will also be in a better position to take measures that will influence the level and nature of the demand for its services (for example, through the use of referral guidelines or online or telephone consultation procedures) and also, crucially, the meeting of desirable goals, such as reducing and sustaining reductions in its waiting times.

The key systems issues with respect to demand are:

- the impact on the demand that a hospital experiences as a result of its own performance
- interactions along the care pathway – for example, how changes in performance at one stage impact on the workload at other stages.

It is important to be aware of the flow of people into outpatient services on a weekly basis. The information that this requires is crucial to managing the corresponding supply of services. At its simplest, the information required is:

- the number of:
 - GP referrals
 - consultant referrals
 - GP referrals accepted
 - consultant referrals dealt with on the ward (or in any alternative manner)
 - follow-up appointments
 - consultations that take place;
 - 'did not attend' (DNAs)
- the number waiting:
 - for an outpatient appointment with a slot
 - for an outpatient appointment without a slot
 - who cancel an appointment
 - who request different appointment
 - who lose a slot (where the hospital cancels)
- the proportion of urgent referrals from GPs
- the average consultation time required (by type of consultation)
- the average time taken by each consultation (by type of consultation).

Understanding the demand for outpatient services needs to be complemented by an understanding of the productive capacity in this area. At its simplest, information requirements for this part of the system will be:

- the number of clinics, consultants, consulting rooms, and slots per clinic
- the average time per slot
- any additional information on distinctions between clinics.

The demand for elective operations arises from decisions made by the clinician during a consultation with the patient. One outcome of the outpatient attendance is a decision by the clinician that the patient requires an elective procedure. If the patient agrees to this, then the consultant will ensure that the patient is placed in a queue for elective care.

To monitor demands placed upon the elective care system requires at a minimum the following data, on a weekly basis:

- the number of consultant referrals for an operation:
 - as an inpatient
 - as a day case
- the proportion of urgent referrals as:
 - inpatients
 - day cases
 - inpatients that are given a date
 - day cases that are given a date
- the proportion of non-urgent referrals as:
 - inpatients that are given a date
 - day cases that are given a date
- the average time required for an operation as an inpatient, by category of operation
- the average number of days in hospital required for an inpatient operation
- the number waiting for an operation who:
 - cancel
 - request different appointment
 - lose a slot (where the hospital cancels)
- the number of operations that:

- take place
- fail to occur for some reason on the day
- the number of people waiting for an operation who are suspended
- the number of DNAs that occur
- the average time of an operation.

A fundamental requirement is to be clear about the level of capacity of elective care service provision that is potentially available.

The information that is required in order to determine operating capacity is crucial to managing the supply of services. At its simplest, this includes:

- the number of:
 - operating theatres
 - theatre sessions
 - consultants
 - anaesthetists
 - nurses (or nurse teams)
- the length of theatre sessions
- the average length of an operation.

Conclusions and recommendations

We identified some of the key issues raised by taking a ‘systems’ view of elective care – that is, looking at the relationships between different systems and between parts of the whole system that need to be understood. We then looked in more detail at the demand for, and supply of, elective care, and the information that a hospital requires to plan its activity, and subsequently to know what is happening to the relevant flows and how it is performing in dealing with them.

We identified two kinds of requirement: an understanding of key relationships (including the factors making for change in the situation a hospital faces), and an ability to describe numerically the main features of the demand for, and supply of, care at a detailed level.

Although some hospitals have much of the information required for improved planning and operational management, in our original survey of nine trusts with good, indifferent and poor performance on waiting times, we found none where this was being used in a systematic way in the context of an understanding of how elective care relates to the various systems within which it is located.

Some of the relationships or systems issues will be hard for any one hospital to address fundamentally. However, other recent developments, such as the introduction of choice and payment by results, introduce new elements to be taken into account, both in the short and long terms, and potentially – at least for some hospitals – a much greater amount of potential variation in some of their planning assumptions. In effect, this means that linkages between specialties working through the financial system become more immediately important than they are now.

Introduction

Why this study?

The present government came to office in 1997, pledging to reduce the numbers waiting for inpatient (including daycase) elective care by 100,000 (King's Fund 2002). While committed to maintaining this reduction, with the publication of the NHS Plan (2000), the emphasis shifted to reductions in waiting times and an increase in the scope of targets to cover specific priority groups, such as cancer patients, as well as outpatients and access to primary care (see Table 1).

To achieve these targets, the Government has taken a range of measures, including:

- significantly increasing NHS funding
- earmarking funds for waiting list and waiting times initiatives
- introducing choice of treatment location by patients, facilitating faster treatment
- encouraging the use of private-sector capacity
- introducing a provider reimbursement system with incentives to increase activity in key high-wait health care resource groups (HRGs)
- disseminating best practice through, for example, the Modernisation Agency.

While most of these initiatives and actions are designed to increase activity, the Government accepts that simply doing more of the same is only part of the solution to meeting the waiting times targets it has set. The consultation document *Reforming the NHS: Financial Flows, Introducing payment by results* argues, 'International and previous domestic experience suggests that increasing activity alone may not be enough to improve access and reduce waiting times. PCTs and trusts need to manage referral and admission thresholds and the priorities for admitting patients' (Department of Health 2002, p 4).

As has been noted elsewhere (Harrison and New 2000), the persistent existence of wide variations in numbers and time spent waiting (between trusts, between specialties, and within specialties) does not depend solely on variations in levels of activity and throughput.

Nevertheless, the existence of significant variations in waiting times across the NHS (variations over and above that which can be explained by differences in need) in itself suggests that the current pattern of waiting times can be changed. There are also some lessons from history on this point. In 1989, around 96,000 people waited over two years for admission to hospital (over 1 per cent of the total list). But, through a combination of targeted funding and list validation, within two years this had been reduced to zero. Similarly, at the beginning of the 1980s, around one-third of all those waiting for admission had waited more than a year, but by the mid-1990s this too had been reduced substantially (to around 2–3 per cent of the total list).

There are, therefore, reasons to be optimistic about the prospects for further reductions in waiting times. However, such optimism needs to be tempered. While the very long waits of the 1970s and 1980s appear – literally – to be an historical feature of the NHS, current targets for reductions represent an increasingly difficult task.

Table 1: Waiting list targets (and milestones) announced since 1997

Target	Date
Cut inpatient waiting lists by 100,000 from March 1997 level	End of 1997 parliament
No one with suspected breast cancer to wait more than 2 weeks for outpatient appointment following urgent GP referral	April 2000
Numbers of outpatients waiting more than 13 weeks to be cut to 334,000	March 2000
No one to wait more than 4 weeks for treatment for testicular cancer, children's cancers and leukaemia following urgent GP referral	December 2001
No one to wait more than 4 weeks for treatment for breast cancer following diagnosis	December 2001
No one with suspected cancer to wait more than 2 weeks for their first outpatient appointment, for patients referred urgently	December 2000
Reduce number of people waiting over 12 months	March 2002
No one to wait more than 15 months for inpatient treatment	March 2002
No one to wait more than 12 months for inpatient treatment	March 2003
No one to wait more than 9 months for inpatient treatment	March 2004
No one to wait more than 6 months for inpatient treatment	March 2005
No one to wait more than 3 months for inpatient treatment (and an average wait of 1.5 months)	End 2008
No one to wait more than 26 weeks for an outpatient appointment	March 2002
Reduce the number of people waiting over 13 weeks	March 2002
No one to wait more than 13 weeks for an outpatient appointment	March 2005
Maintain the commitment to cut waiting lists by 100,000 from March 1997 level	Ongoing
All patients attending A&E to wait 4 hours or less, from arrival to admission, transfer or discharge	March 2004
No patients to wait no more than 24 hours for an appointment with a primary health care professional, and no more than 48 hours for an appointment with a GP	March 2004

This is partly due to the volume of patients currently waiting for longer than the target periods compared with numbers previously enduring very long waits. For example, across England (and on a hospital basis), in the third quarter of 2003/4, there were over 151,000 patients waiting over six months (and over 406,000 waiting over three months), compared with 96,000 waiting over two years at the end of the 1980s. Moreover, while these latter waits were reduced substantially through relatively simple processes such as validation, the gains from such tactics have probably now been exhausted. Pushing waiting times ever lower and, crucially, sustaining such reductions, thus represents a task of increasing magnitude and one that requires different strategies to those employed in the past.

Study aims

Given the need to achieve significant further reductions in waiting times, the existence of trusts that have been consistently successful in achieving (and in some cases, exceeding) these targets provides an opportunity to identify contextual, managerial and operational characteristics of these trusts that explain their success, and that might be transferable to other organisations. The aim of this study is to isolate the factors that lead to sustainable reductions in waiting times.

In addition, it was recognised that there were likely to be other researchable issues emerging from the first part of this study that would shed additional light on issues related to the sustainability of reductions in waiting times. Two issues emerged from the first part of the research that were followed up:

- the extent to which strategies to reduce maximum waiting times, and in particular the use of maximum waiting times targets or guarantees, clashed with clinical priorities concerning the admission of patients
- the information requirements of trusts to formulate strategies to reduce waiting times.

Overview of this paper

The next part of this paper, 'Conceptual background' sets out some further definitional, analytical and conceptual background bearing on the study. Then Section 1 details the first stage of the research, based on an analysis of a selected sample of trusts with variable performance on waiting times. Section 2 reports on the second stage of the research, and details an analysis of possible distortions in clinical priorities arising from managerial efforts to meet the maximum inpatient waiting times target of 15 months on 2001/02. Section 3 details our research into the basic information needs of trusts in order to begin to tackle reduction in waiting times. The final section relates to further research.

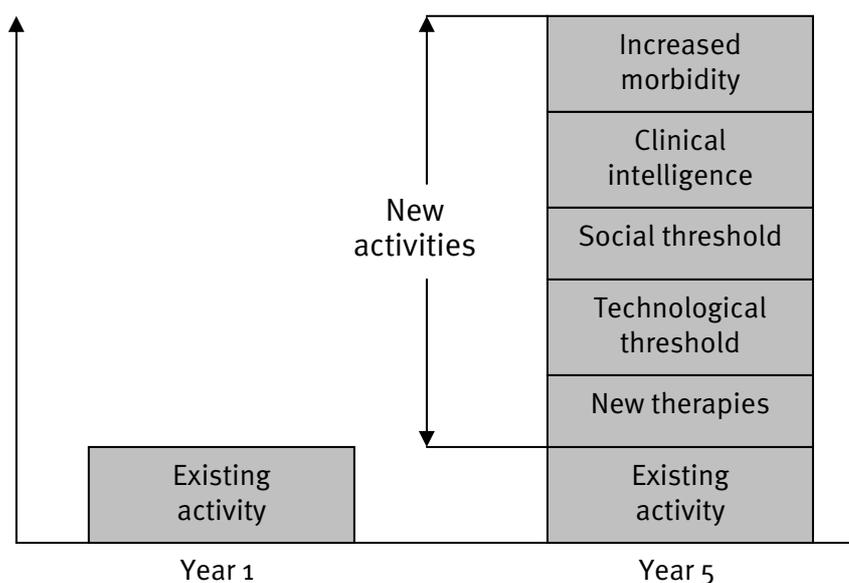
Conceptual background

This section serves several purposes. First, it sets out our view of how the elective care system works, highlighting the concept of ‘whole systems’, and providing the conceptual model underpinning our analysis of the determinants of waiting lists and times. Second, it provides a critique of existing studies that have sought to explain variations in waiting times performance. The material in this section provides the rationale for the study design of the first stage of the research and specific hypotheses described in the following section.

How the elective care system works

Earlier King’s Fund studies by Harrison and New (2000) and Hamblin, Harrison and Boyle (1998) described the various initiatives to tackle waiting lists and waiting times taken since the foundation of the NHS. The historical record shows that numbers waiting had risen over time, and that any improvements had been short-lived. There had been some success in reducing very long waiting times, but average waiting times had changed very little.

Figure 1: Factors that may lead to new elective activity



Harrison, New (2003)

Harrison and New concluded that these initiatives had been based on an incorrect view of how waiting lists worked – a view that waiting lists represented a backlog which could be removed by ad hoc and temporary initiatives, such as those that characterised policymaking for much of the post-war period. Accordingly, they went on to set out an alternative view summarised in Figure 1, above. (For fuller discussion, see Harrison and New 2000).

According to this view, sustainable reductions must rest on the indefinite continuation of policies designed to respond to the range of forces set out in Figure 1 – that is, to meet

a level of demand that rises in response to technical change, demography, rising user expectations, and changes in clinical behaviour. Although it is impossible to put precise figures on the impact of these forces, at a minimum, and in the absence of other measures, they require some sustained increase in activity.

In addition, if reductions in waiting times are achieved (for example, by short-term bursts of additional activity) that may in itself lead to responses that tend to obviate the initial success. That is to say, the demand for elective care will to some degree depend on waiting times themselves. If they are reduced, then this reduction may lead to more people seeking treatment, and to changes in clinical behaviour that allow more people into the system for treatment. Such increases in demand may come from a range of sources – for example:

- Improved access times in the NHS may encourage some of those using private facilities, particularly self-payers, to remain in the public system.
- Those who might otherwise have treated themselves or sought help from other therapists may decide to go ahead with hospital treatment.
- GPs may be more ready to refer.
- Hospital consultants may modify their treatment thresholds.
- A successful hospital may attract referrals from less successful hospitals.

To achieve sustainable reductions, therefore, requires strategies to deal with long-term trends and the response to whatever the initial impact of those strategies is, in terms of more people seeking, or being advised to seek, treatment.

The whole system

Before considering the range of strategies that might improve waiting times, we need to consider the context within which the elective care system operates. In recent years, the term ‘whole system’ has come into common use, often without careful definition. In this paper we use it to denote two concepts:

- the whole hospital system
- the local health economy in which the hospital is located, including private sector and social care facilities. This includes policies that might bear on the demand for care, as well as its provision.

The hospital system

Within the hospital itself, elective care is part of a wider system providing a number of other services, of which the most important is emergency care. In most UK hospitals, staff, beds, operating theatres and diagnostic equipment are to a greater or lesser degree shared facilities. It follows that the capacity of the hospital to provide elective care depends on the extent to which these shared resources are required for other uses.

The central characteristics of emergency care are variability, unpredictability and immediacy. Recent work from the King’s Fund (Dixon and Damiani 2002) on the factors giving rise to variations in demand suggests that the pattern of demand over the winter period can be anticipated with some confidence. Moreover, hospitals working with the Met Office (Met Office 2001) have found that variations due to the weather can also be forecast for a short period ahead. However, some residual uncertainty remains arising from random variation in the day-to-day workload.

In the past, the elective care system – particularly its bed stock – has provided the main ‘cushion’, or reserve capacity, for emergency care. With shorter lengths of stay and increasing day case treatment, this cushion has been reduced, with the result that cancellations of elective activity have become more common.

Accordingly, improvements in elective care capacity depend to some degree on the way in which the links between the elective and emergency sub-systems are managed. In addition, investment in facilities not forming part of the elective care system itself may be critical in allowing it to function more effectively and efficiently – for example, by increasing the number of beds available for dealing with emergency patients.

The wider health economy

The wider health economy is also critical to the performance of the elective care system. This is manifested in several ways:

- Some elements of the elective workload may be transferred to other parts of the wider health system (for example, simple procedures or consultations in primary care).
- Through the referral process, the wider health economy is the main source of demand for the hospital elective system.
- The efficiency with which the wider health economy accepts patients back into the community is a key determinant of the efficiency of the whole hospital system (as measured, for example, by average length of stay).
- Fundamental redesign of patient access requires contributions from all parts of the health economy.

So the elective care system is itself complex, and how it works is also determined by the wider system in which it operates. Any strategies to improve its performance must take into account the way the elective care system responds to changes in waiting times, the way in which other demands on the hospital system are dealt with, and the whole system within which the hospital operates.

What counts as a reduction in waiting times?

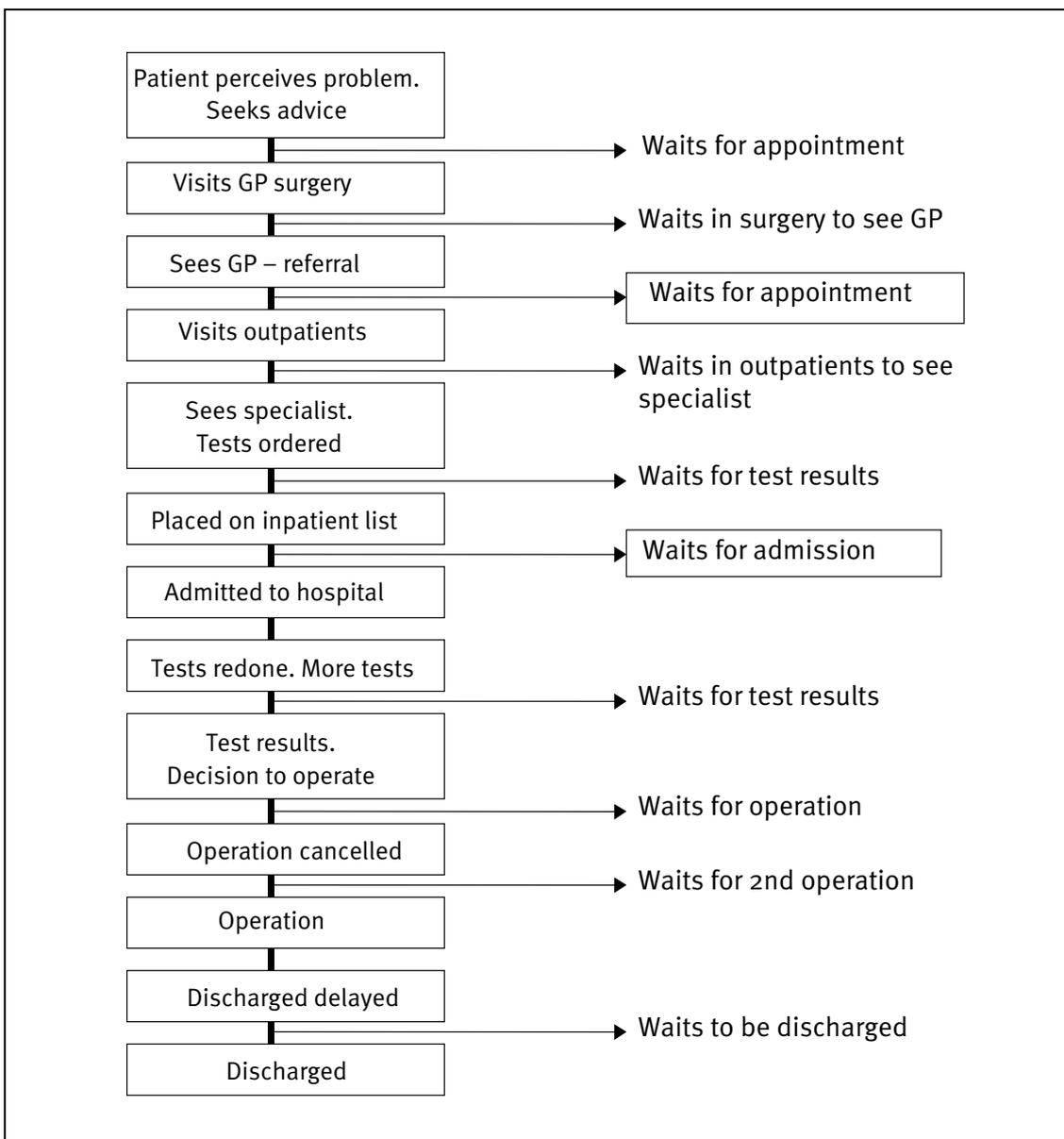
At present, the total time a patient spends waiting for treatment (in other words, waiting at all stages of the care pathway – see Figure 1) is not captured within the reporting systems required by the Department of Health. Waits at some of the stages on the pathway are usually short but there may, for example, be long unrecorded waits between the first outpatient wait and the decision to admit – particularly if the patient is referred from one consultant to another, and if initial diagnostic tests are inconclusive.

The targets set by the Department of Health for first outpatient appointments and for patients treated from the waiting list only capture two of the stages where patients may wait for treatment (see Figure 2). Accordingly, practical considerations dictate that only these partial measures could be included in this study. But for this, and other reasons outlined in the box on p 22, apparent improvements, as measured in the data held by trusts, may disguise the situation on the ground. Examining the relationship between the apparent situation and the real one is outside the scope of this report.

Even within the available data, however, the measurement of a reduction is not straightforward. In selecting trusts to be included in the study, we have taken their performance as measured by the Department of Health for its performance targets for patients waiting for admission. These focus on those who have been waiting a long time. However, elimination of such long waits does not necessarily mean that the average wait for all patients is reduced. In fact, average waits have risen in recent years even though very long (over 12-month) waits have been almost eliminated. In effect, recent policies (in a similar way to those adopted in the 1980s) have compressed the shape of the list.

For many trusts, there is scope for further compression of this kind. However, reductions cannot be achieved by this route indefinitely if they are not accompanied by increases in activity to reduce waits across the board. Accordingly, policies that may have been successful up to now may not be in future years.

Figure 2: Waiting for care: an example of a patient’s actual experience*



*Officially recorded waiting times are shown in bold

Sources of apparent improvements in waiting times: confounding factors

There are various other reasons why improvements may be recorded, which do not reflect a real improvement in the underlying situation.

Hospitals may ‘clean up’ their waiting lists by checking with all those on them whether they still wish to be treated. If this task is neglected and then carried out in a short period of time, numbers waiting will fall and those still in the queue may be treated more quickly. But once the main gains from such procedures have been achieved, they cannot be repeated on the same scale. Here there is a genuine backlog effect.

Changes in priorities within the list may lead to reductions in some categories of waits. Current policies targeting long waits may have the effect, where successful, of squeezing up the queue, but the gains to those waiting for a long time are at the expense of those nearer the top of the queue. As a result, the average wait may change very little.

Recording of waiting times is not complete. Current recording systems take into account only two periods of waiting: for the first outpatient appointment, and the period from the decision to treat to treatment. It appears, however, that waiting between these two periods may have been increasing (for example, for a subsequent outpatient appointment as a result of cross-referral). Furthermore, there is also evidence of unrecorded and lengthy queues for some investigations.

Finally, there are changes in recording. At one extreme, such changes represent deliberate ‘fiddling’ with the data (see National Audit Office 2002, Auditor General for Scotland 2002). However, there are other reasons why recording changes may legitimately be made that impact on the recorded figures. The dividing line between elective activity relevant to waiting lists and other hospital activity is not absolutely clear. In particular:

- some elective activity takes place as a planned series of admissions at regular intervals, and those waiting for planned care are not counted as waiting – however, some procedures overlap the two categories
- the distinction between diagnosis and treatment is not absolute: some procedures appear on both sides of the line
- some activity may be carried out and recorded as day surgery, or as outpatient or clinic activity, or not recorded at all.

In brief, because these boundaries are porous, and because recording systems are not complete, there is scope for improvement to appear without any change in the underlying situation.

Previous studies

As far as we are aware, the present study is the first of its type. However, other studies have considered part of the ground covered here.

While some studies (for example, Locock 2001) have provided some important indications of the multiple factors involved in achieving significant and sustained reductions in waiting times, it is hard to draw general conclusions from studies of single

trusts. First, investigating a single trust can provide only a partial analysis: it describes the characteristics of that trust that are associated with its waiting times performance. Identifying which of these factors explain that performance requires a comparison with less successful trusts.

Second, relying on qualitative analysis identifies the factors that key actors in the system believe explain waiting times performance. In practice, these views may over- (or under-) state the importance of various factors (including the practices of those interviewed), or they may be biased, or indeed wrong. For example, certain managerial and operational practices may be a feature of successful trusts, while not being unique to them. Superior performance may be explained by differences in resources, and the reduction in waiting times might be explained to a greater degree by differences in the amount of resources devoted to elective surgery and the greater use of non-NHS capacity.

This would suggest that it is important for qualitative analysis to be comparative between successful and unsuccessful trusts, and for qualitative analysis to be accompanied by quantitative analysis of key variables. This is, in large part, what we attempt in the study described in the remainder of this report.

Other studies have attempted to identify significant differences between trusts with short and long waiting times in a range of measurable demand, resource provision, resource use, and other characteristics. For example, in a study of ear, nose and throat (ENT) waiting times, Harley, Jayes and Yates (1999) found that trusts with long waiting times tended to:

- make less use of day surgery
- have a higher proportion of emergency admissions
- have higher levels of new outpatient DNAs (did not attends)
- have much lower levels of productivity per senior surgeon than short-wait trusts.

A problem with this sort of pair-wise comparison (rather than a multi-factorial approach) is, as the authors note, that while statistically significant differences can be identified in a set of characteristics between short- and long-wait trusts, there is considerable overlap in the distributions for all such characteristics. For example, some long-wait trusts have a higher proportion of day case work than some short-wait trusts yet still have longer waits than the latter.

A more recent study of access to ENT services (Audit Commission 2002) also attempted to identify pair-wise correlations between whether a trust had short or long waiting times in three areas:

- demand (GP written referrals per 1,000 population)
- capacity (whole-time equivalent consultants per 100,000 population)
- efficiency (total outpatients seen per whole-time equivalent consultant).

No significant relationships were found. Nevertheless, as the study noted, this does not mean that such factors are irrelevant. The problem here, again, is the limit of the methodology in unravelling what is a complex interaction between variables such as demand and capacity, and waiting times.

Martin et al (2003) attempted to identify the factors associated with long waits (over six months), taking the four main waiting list specialties (general surgery, ENT, ophthalmology and trauma, and orthopaedics) and examining possible correlations between long waits and various explanatory variables covering NHS capacity, independent sector activity, need for health care, and other characteristics of

hospitals, such as teaching status and star rating. In general, they found no significant correlation between long waiting and any measure of capacity or independent sector activity. There appeared to be some positive correlation with numbers of anaesthetists and an inverse correlation with measures of population deprivation (in other words, more deprived areas had fewer patients waiting over six months).

This study suffered from a number of limitations (noted by the authors). Further, it appears that the dependent variable was simply the number waiting over six months rather than, say, the proportion waiting over six months, which would give rise to size effects.

1. Sustaining reductions in waiting times

Research objective

The main objective for the first stage of this research was to identify strategies adopted by trusts that appear to have been successful in not only reducing inpatient/day case waiting times, but in sustaining the reductions achieved.

Research design, methods and expectations

Previous (quantitative and qualitative) studies to identify unambiguously factors associated with short waiting times have had their limitations. This is not to say such approaches are without merit, as it would seem extremely unlikely that observed differences in waiting times are completely unassociated with any other observed variations in a range of demand, capacity or performance measures. However, it may be that a more sophisticated, multivariate, statistical technique would be more fruitful, or, indeed, that whatever differences in key trust characteristics that are observed should be considered important whether or not such differences are statistically significant.

Given the limitations detailed above, and hence to provide clear policy recommendations to help reduce observed variations in waiting times, this study opted (in its first stage) for a more in-depth qualitative analysis of short- and long-wait trusts. It builds and expands on existing research – particularly Locock’s (2001) qualitative study of Dorset – together with collection and analysis of a range of data concerning the performance of trusts, and other information, a priori, considered important.

Given the need for comparative analysis of factors linked to waiting times performance (as noted in ‘What counts as a reduction in waiting times?’, p 20) the project approached ten trusts to take part. One (successful) trust declined due to work pressures. That left nine trusts, falling into three categories of waiting time performance:

- **successful** – consistently low proportions of patients waiting over six months
- **variable performance** – some success in reducing the proportion of over six month waiters, but not sustained
- **unsuccessful** – consistently high proportions of patients waiting over six months.

Further details of the way trusts were selected are provided in Appendix 2. Figure 3 shows historical trends in the proportion of patients waiting over six months for admission to hospital for each trust in our study, where trusts A, B and C are in the ‘successful’ category, trusts D, E and F are of ‘variable performance’, and trusts G, H and I are in the ‘unsuccessful’ group. Trust A had eliminated long waits by 1995, and has managed to more or less maintain that position since then. Trusts B and C had sustained progress, albeit with some setbacks. Trusts D, E and F had made progress but this had been reversed or stalled. In the remaining three trusts, performance had worsened.

Figures 4 and 5 show distributions of waiting times, using information from the Department of Health’s Hospital Episode System (HES). These are not strictly comparable

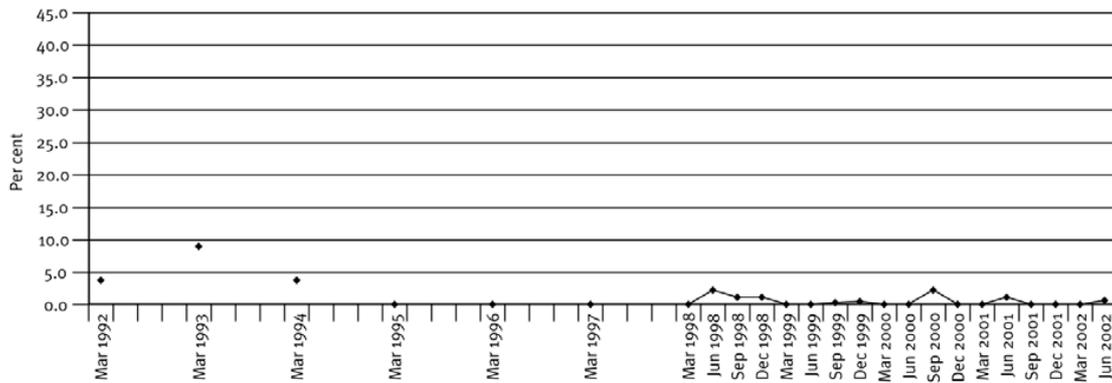
with Figure 3, which are based on Körner returns (for example, according to Figure 4, all trusts had at least 10 per cent of patients waiting more than six months for treatment). Figure 4 shows the cumulative distributions of waiting times for the combined waiting lists of four key specialties (ENT, trauma and orthopaedics, general surgery and ophthalmology), for our sample of trusts. Figure 5 provides profiles of the waiting lists for each trust in the sample for each of these specialties separately. For each specialty, the lines corresponding to each trust show the four quartiles of the distribution of waiting times (the shaded area is the middle 50 per cent range).

The general patterns in profiles and distributions reflect the groups we chose for our sample. Figure 4 shows the 'successful' trusts tend to have more 'rectangular' cumulative waiting times distributions – in other words, they have a greater proportion of patients waiting only a short period of time and the numbers waiting in the 'tail' are consequently lower. Similarly, Figure 5 shows, for the 'successful' trusts, comparatively compact distributions for key waiting list specialties. Those in the 'unsuccessful' group tend to have the opposite distributional characteristics. Both figures reveal some overlap between those in the 'variable performance' group and the other groups, particularly when individual specialties are considered.

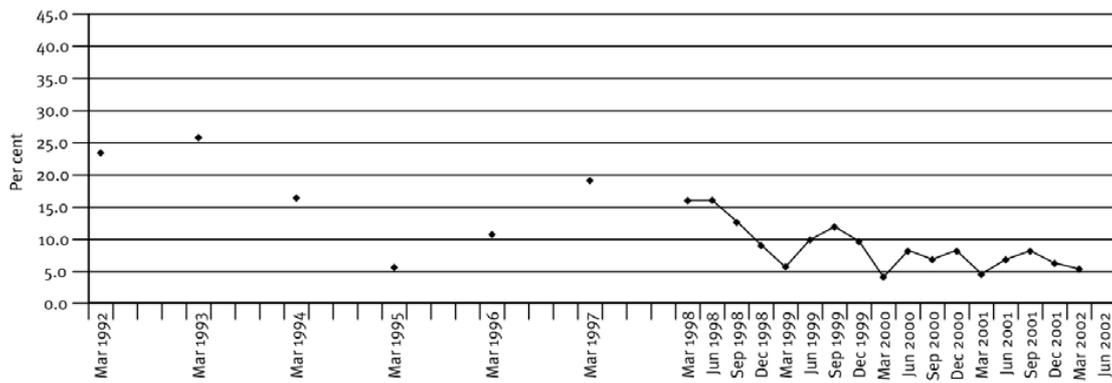
Through semi-structured interviews (see Appendix 3) and the collection of trust/specialty and, where appropriate, consultant-level data, the study aimed to identify patterns of activity, resources, management and clinical policies, processes, attitudes, behaviours and strategies, as well as contextual factors that characterised the three groups of trusts, and thus to isolate the factors that explained sustained waiting times performance.

Figure 3: Proportion of patients (of total trust list) waiting over six months

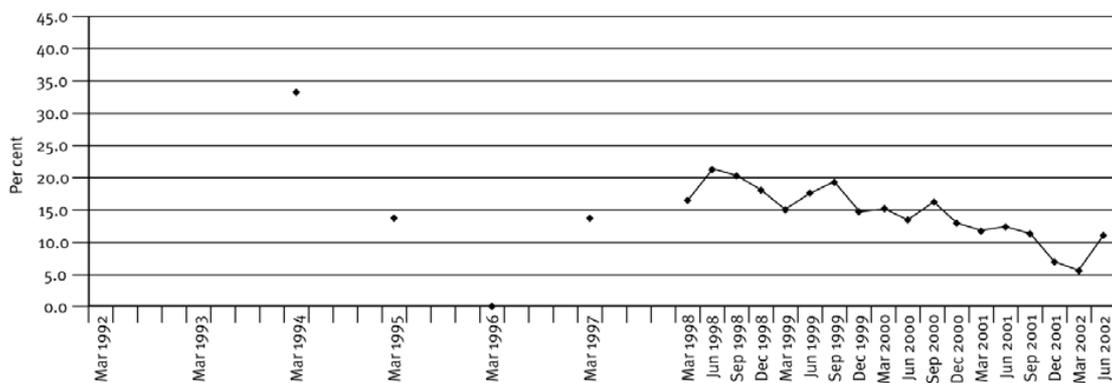
Trust A: March 1992–June 2002



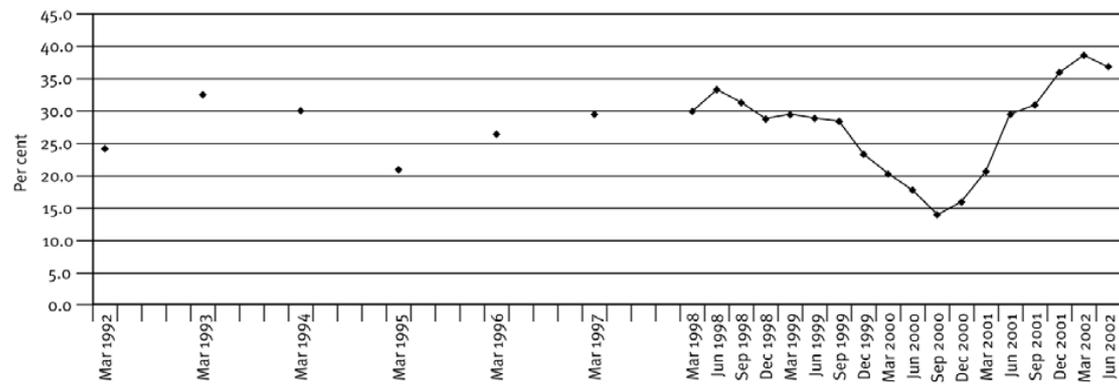
Trust B: March 1992–June 2002



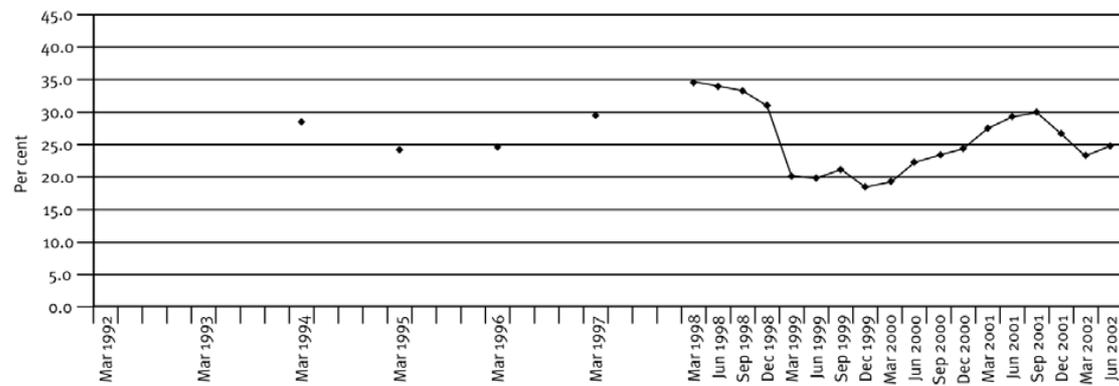
Trust C: March 1994–June 2002



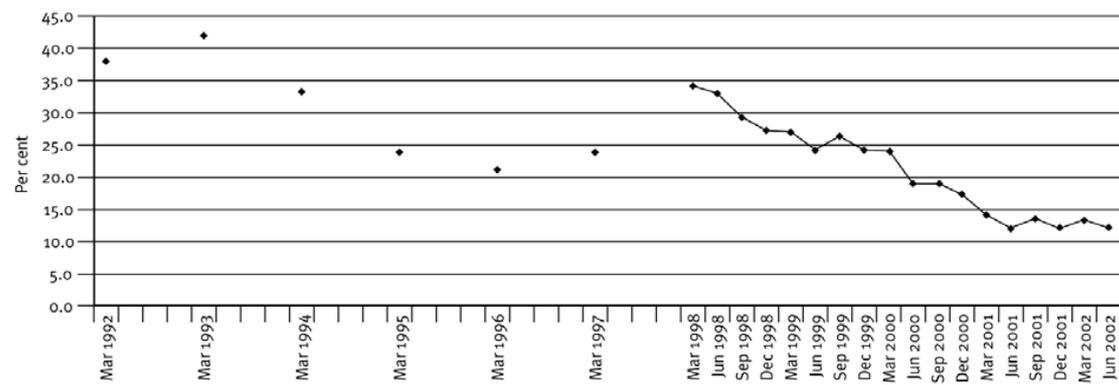
Trust D: March 1992–June 2002



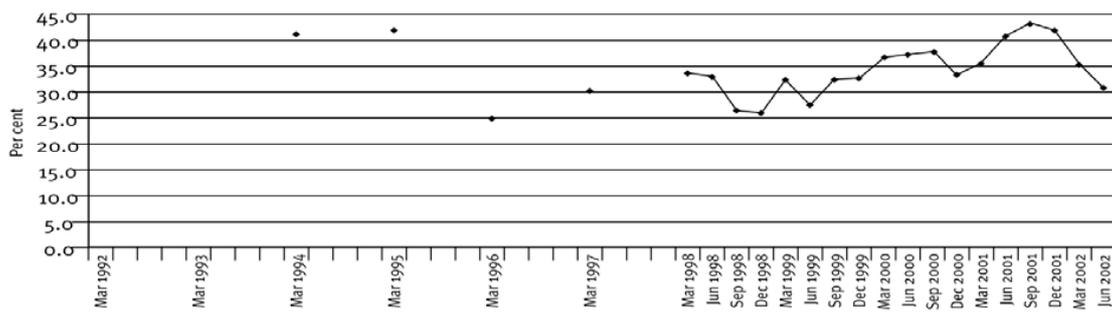
Trust E: March 1994–June 2002



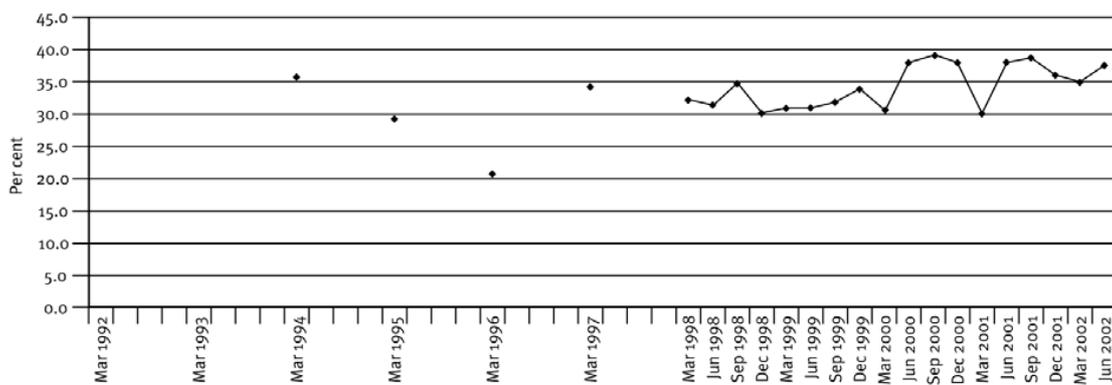
Trust F: March 1992–June 2002



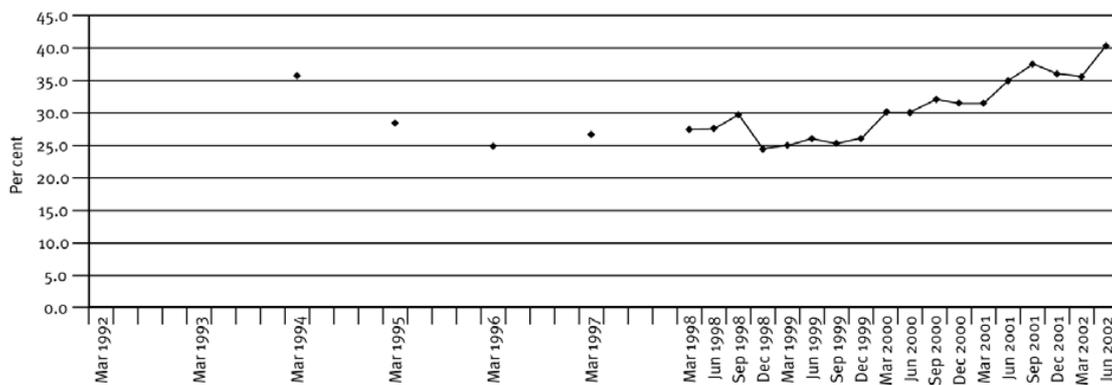
Trust G: March 1994–June 2002



Trust H: March 1994–June 2002

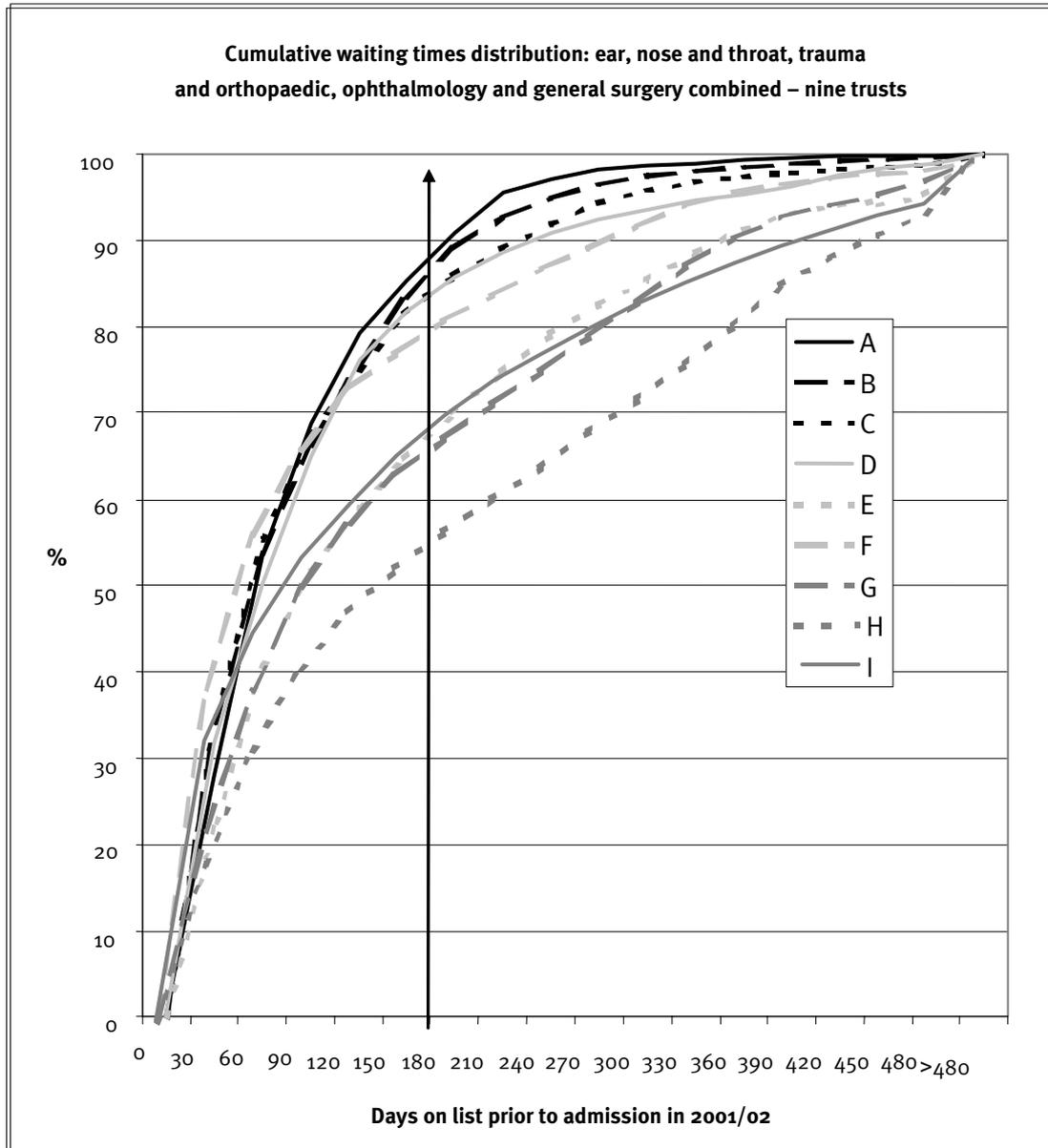


Trust I: March 1994–June 2002



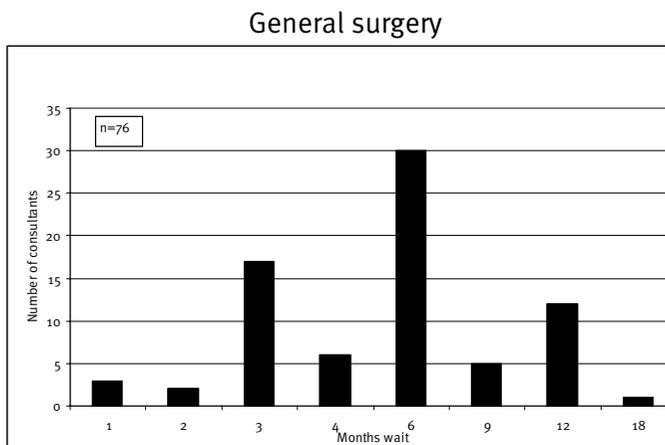
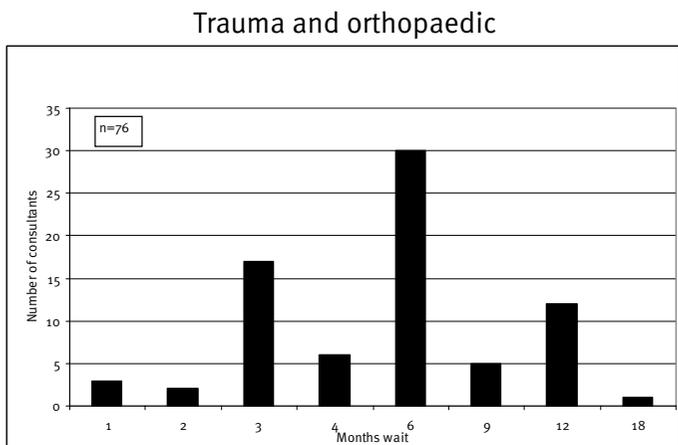
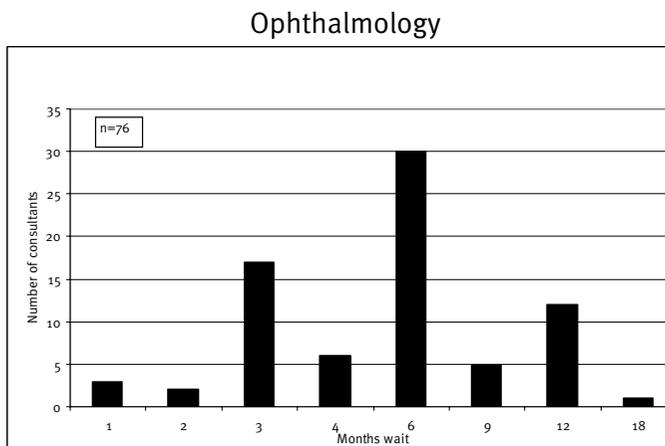
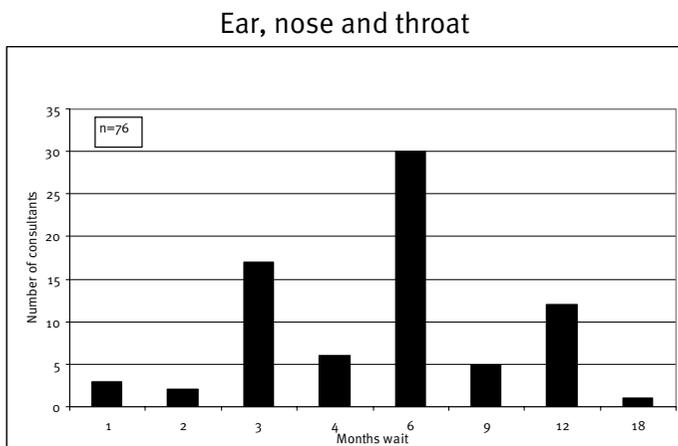
King's Fund (2003)

Figure 4: Cumulative waiting times distribution for sample trusts



King's Fund (2003)

Figure 5: Waiting times profiles for patients admitted during 2000/01



These distributions show the shortest and longest waits for each trust (excluding outliers defined as 1.5–3 box lengths from the lower or upper edge of the box). The shaded box shows the middle 50 per cent interquartile range and the line in the box shows the median waiting time.

Unsuccessful trusts tend to have greater ranges in distributions of their waiting lists (although median waits tend not to follow a particular pattern across the three groups of trusts).

NB: Trust G had no ENT and Trust D no ophthalmology service.

Methods

The original research proposal set out a two-stage approach.

Stage 1 involved:

- collecting historic quantitative data at hospital/consultant level, where appropriate, for each trust for the main waiting list specialties
- interviewing key players – consultants, managers, GPs – to explore attitudes to waiting lists in general, perceived reasons for success or failure, private work, and so on. The interviews were semi-structured (*for the interview protocol, see Appendix 4*)
- analysing quantitative and qualitative data to draw out possible common explanations and indicators for success or failure.

In Stage 2, the focus was dependent on the findings and hypotheses that emerged from the first stage. In addition, we intended to disseminate findings/hypotheses from Stage 1 to poorly performing trusts. We then planned to monitor waiting times and other indicators of success in the trusts, and undertake further selected interviews to track any changes in factors such as attitudes and behaviours.

In the event, work in the first stage raised issues that it was decided warranted further research (reported in sections 2 and 3). While a draft of the first stage of the research was circulated to chief executives of all the trusts that took part in the work, no follow up was carried out given the decision to investigate the two issues that arose from the first stage.

Findings

Figures 3, 4 and 5 show that there are wide variations in performance between trusts. (We assume that these differences are genuine, and not the result of recording differences.) The previous discussion suggests that these differences may be attributed to a wide range of factors and gives rise to a range of different explanatory expectations for each group of trusts.

Given the findings from previous studies (*see Previous studies, p 22*), there was little expectation that from our small sample of trusts we would be able to identify unambiguously a unique set of factors that wholly explained the variations in waiting times performance between the trusts. However, there was an expectation of being able, through the interviews with key managers and clinicians and appropriate data analysis, to identify broadly the right conditions for reducing waiting times and sustaining the reduction achieved. We also expected to be able to highlight particular examples of successful waiting times reduction strategies.

In total, 45 senior managers and clinicians were interviewed in the study's sample of nine trusts, including, in some cases, managers from local PCTs. In all trusts, interviewees included the chief executive, medical director and director of operations and/or the senior manager with responsibility for waiting list or waiting times targets. In some trusts, additional interviews were carried out with other staff on the recommendation of other interviewees. Interviews in each individual trust or PCT were carried out by the same primary interviewer for that site, with all interviews conducted using a common interview schedule (*see Appendix 4*). Interviews were carried out between August and December 2002.

Interviews were conducted using a mixture of face-to-face and telephone interviews. In both cases they were recorded, and summary transcriptions of the interview were completed by the interviewers. At least one visit was undertaken for each site. These then formed the basis for a meeting of all researchers, which drew together themes and explanations for the waiting times experience of each trust.

We know that the reasons that some trusts appear more successful than others at reducing waiting times are linked to a whole range of factors. What has bedevilled previous research (and this study) is not so much the difficulty of identifying the many factors involved, but rather understanding the relative importance of individual factors and their interaction with each other. The main reason for this difficulty stems from the sheer variety of circumstances and history for individual trusts. In the literature bearing on the relative performance of similar institutions, there is now considerable consensus that the context in which the organisations operate is an important determinant of success or failure, and that each context comprises a unique constellation of factors and relationships (Pettigrew *et al* 1992, Pawson and Tilley 1997). This means that we cannot expect to uncover a clear prescription for success that can be universally applied.

Nevertheless, while it may be very difficult to unambiguously identify for any individual trust not only the necessary conditions for being successful, but the sufficient conditions as well, the research has identified a number of common themes and factors – as well as an indication of a cluster of factors that appear to be associated with sustaining reductions in waiting times. These findings have arisen through iterative discussion of the interview data by members of the research team, as well as through analysis of the quantitative data. These generalisable factors will be of varying importance in different sites over time, but we can conclude that trusts should consider their relevance for their own current situation. Some NHS-wide recommendations can also be identified, and these are discussed later.

None of these factors on their own would be a sufficient condition for success. Sustained reduction in waiting times requires action across a number of fronts, and more successful trusts are more likely to have several of these factors in place, while in less successful trusts some gaps can be identified that are undermining their efforts, or have done so in the recent past. Single initiatives are unlikely to prove to be a long-term solution for the reasons outlined in *How the elective care system works*, p 18.

The framework we have adopted below to present our findings seeks to move from a broad and overarching level (the need to understand hospitals as whole systems) through to more detailed factors, such as implementing a system for accurately reporting waiting times to senior managers and clinicians on a weekly basis. However, given the multi-factorial nature of the issue, there will inevitably be some overlap between sections. Where possible, we illustrate the factors identified with anonymised examples taken from our case study sites.

Overall findings

Understanding whole systems

A very broad generalisation emerging from our interviews – and in particular, our discussion of interview findings within the research group – is that trusts with a poor record in reducing waiting times had a poor understanding of the way that improvement

in waiting time performance depended on measures taken in other parts of the hospital, and also on the wider health economy. This relative lack of understanding also applied historically to those trusts that used to have a poor record on waiting times but had started to improve. Further, interviews with key personnel in the three ‘successful’ trusts revealed the converse: not only a reasonably good sense of the whole system of care, but also an appreciation of the importance of such an understanding, which was reflected in the specific measures they took to achieve government targets.

From an understanding of whole systems – not just in theory, but in practice – flows a whole range of managerial and clinical actions and behaviours which, when brought to bear on a particular objective (reducing and sustaining reductions in waiting times) tend to lead to success, rather than failure or temporary success. For example, an appreciation of whole-systems working leads quite naturally to a focus on patient care pathways. Similarly, the need to take a view of activity across the whole hospital system leads to prioritisation of investment in timely, system-wide information systems. Other actions and behaviours are noted below under individual factors and strategies.

Similarly, recognition of the whole system, including the wider health economy, may suggest measures bearing on the demand for hospital services, such as referral protocols, or measures such as investment in intermediate care, which allow the hospital itself to operate more effectively.

The importance of sustained action over time

A second general observation is that successful trusts started to address the task of reducing waiting times in a systematic way much earlier than unsuccessful trusts, and had persevered with the task.

Example: The long-term approach

This site is perhaps an extreme example of the importance of a long-term approach, with over a decade’s history of concerted effort to reduce waiting times, with part of this time clearly spent learning what works and what does not. Interviewees at this site agreed that their success was the result of dogged persistence – reducing waiting times has been an unwavering and explicit priority.

Clearly, success could not be solely ascribed to the length of time spent on this issue, however, and interviewees noted that consistent emphasis on improving waiting times was backed up by an effective system of incentives to recognise good performance. At the same time, where problems were identified, managers and clinicians alike could expect very close monitoring and strong pressure to improve. Nonetheless, there was unanimity that other trusts could not be expected to produce similar results overnight.

Unsuccessful and temporarily successful trusts, on the other hand, have, by their own admission, only really started to ‘get going’ with waiting times reductions in the last 18–24 months. Of course, these trusts had not previously completely ignored the issue, but it is only relatively recently that they had come to place it consistently in the top three of their organisational priorities (very possibly as a result of the emphasis on waiting times targets by ministers and the Department of Health) and thus devoted the necessary management energies to the task which, in order to be successful, need to be considerable. In other words, they had until recently implicitly adopted the ‘backlog’ model of waiting list behaviour, relying on ad hoc initiatives, such as weekend working and other measures that could not be sustained indefinitely, and which often depended on time-limited injection of funds.

Catch up, keep up

A third general finding is the observation that the factors necessary to reduce waiting times are not always the same, or of the same importance or scale, as those involved in sustaining reductions. Catching up is not necessarily the same as keeping up. As we have noted above, short-term measures are inadequate as they fail to deal both with long-term demand trends and with the impact of shorter waiting times on the demand for treatment. We also believe that where measures are sustained, the nature of the problem facing trusts changes.

In our view, there is a clear demarcation between trusts that have achieved consistently low waiting times and those still aiming to do so. Once waiting times have been kept low for some time, many of the problems faced by other trusts no longer apply. For example, the need to ringfence elective activity or manage demand through, say, referral protocols is less relevant once waiting times are so low that all referrals can be quickly processed. It may be that the (feedback) effect of rising demand, as GPs, for example, reduce their referral thresholds in response to lower waiting times, is a transition problem.

Once GPs feel confident that they can obtain treatment quickly for their patients whenever they need it, they may feel less urgency to refer. As Gravelle, Smith and Xavier (2002) have shown, demand and supply of elective care does appear to be responsive to the 'pseudo' price signal of waiting times. Here, the speculation is that the demand response is not straightforwardly related to the waiting times signal but varies depending on the length of time a trust manages to keep waiting times low. That is, when lists or times are long, GPs may raise their referral threshold, in some sense, 'under' referring, but as lists and times start to reduce, they may then lower their referral threshold, in some sense 'over' referring. However, given sustained reductions in waiting lists/times, GPs may once again re-evaluate their referral threshold and settle on a rate somewhere between 'under' and 'over' referral.

The degree to which potential private patients switch back to the NHS may also be a transitional issue. Some interviewees argued that some patients would always choose to go private, so reducing the volume of switching. Nonetheless, there will be some patients who do choose to use the NHS instead of private, particularly if they are self payers.

However, trusts with very low waiting times may face their own, rather different, problems. For example, one senior doctor noted that patients may become anxious that they are being treated so quickly and believe their condition must be much more serious than they are being told, to warrant such speedy intervention. At an organisational level, trusts may find it more difficult to keep theatres working at full capacity. If there is a late cancellation, it can be difficult to find a patient willing to come in at short notice to fill the slot, if their planned date is so close that there is little advantage for them in coming in earlier, with all the disruption that implies. One trust manager reported that they work 'up to the last minute' to try to fill slots, and commented 'we've now got people's heads around the fact that if you miss that slot you might as well stand there ripping up £5 notes.' As a result, it is arguable that the need for good information systems and active management measures becomes even greater, rather than less, as performance improves.

Unexpected shocks

Furthermore, and perhaps unsurprisingly, even where there is an appreciation of the whole-systems nature of the waiting times reduction issue, external shocks can upset

even the best laid plans. The most obvious example is where there are unexpected increases in referrals, or in emergency admissions.

In some cases, disruption can be caused by policy changes imposed by government. For example, some trusts reported that the creation of primary care trusts (PCTs) was having an adverse effect – plans that had been previously agreed were now being subjected to renewed scrutiny by new PCTs that were understandably unwilling to commit resources to a development they had not helped to plan. This might include plans to invest in new admission and observation facilities to reduce inpatient admissions, or step-down facilities to improve the efficiency of discharge arrangements.

Straight trade-offs between objectives can have a similar effect. For example, some of the trusts in our sample experienced significant difficulties achieving financial balance. In some cases, these problems were potentially severe enough to jeopardise their waiting times reduction efforts – at least for a time – with management energy and financial resources being diverted elsewhere.

Example: Changing pattern of activity

Case study D is an interesting example of a trust that had been making good progress in reducing waiting times but then faced a dramatically worsening position. The trust ascribed this partly to an underlying shortfall in elective bed capacity compared to other similar trusts, but in addition identified a sudden change in emergency admissions that had knocked it off course.

During 2001/02, average length of stay rose unexpectedly by one day, which severely disrupted planned elective activity. The reasons for the change in stay were not entirely clear, but were assumed to be partly related to an ageing population with more complex inpatient needs. Despite strong joint working with social services to improve the range of facilities available, delayed discharges remained a problem, and it was noted that the total stock of nursing home beds locally was declining as homeowners decided to close in the face of rising costs to meet legal standards.

This case illustrated how knock-on consequences affect the whole system: when length of stay went up and the hospital was struggling to find beds for incoming emergencies, care became more fragmented as medical patients had to be placed as outliers in surgical wards. This, in turn, meant that discharge was less efficiently handled, leading to further pressure on capacity. At the time of interview, the trust had responded by opening a new observation unit and a step-down unit, and was hopeful that waiting times would be brought under control again, despite an increase in emergency admissions of 14 per cent in the first few months of 2002/03.

Clinical ownership and involvement

A final broad observation – again, unsurprisingly – is that those traditionally responsible for managing the workload of a hospital, and hence waiting lists and times – consultants – are central to the job of reducing waiting times. For example, in a number of unsuccessful and temporarily successful trusts, there were individual consultants who maintained short (six months or under) maximum waiting times.

This suggests that good or bad performance depended to some degree on individuals, rather than the effectiveness of the hospital management as a whole. However, in some cases, individual good performance may have been achieved at the expense of others. For example, keeping waiting times short by the simple expedient of not accepting

referrals that would increase maximum waits given resources available to carry out the work almost certainly led to an impact on the waiting lists of consultant colleagues (or, perhaps, to waiting being exported back to general practice).

It is also worth noting at this point the disjuncture between the unit traditionally responsible for waiting times – in other words, the consultant – and the unit used to measure progress towards the nationally-set target on waiting times – in other words, the trust. So, pressure from above on those nominally in charge of trusts – managers to meet waiting times targets can be dissipated at local level, especially where relations between management and consultants are poor and/or where consultants' objectives with regard to their work are not fully aligned with the objectives of the organisation for which they work.

The question of consultants' motivation and behaviour was discussed during interviews. The question of financial incentives cannot be ignored. Many interviewees, including medical directors, commented that repeated use of one-off initiatives had created a culture in which medical and some other staff expected to be paid extra for doing waiting list work, and had come to rely on the additional income. This was proving an obstacle in trying to change people's thinking to see waiting time reduction (and in the longer term, consistently short waits in line with central targets) as a mainstream activity that was part of everyone's normal daily work.

Expecting consultants to work also towards a situation that reduces patients' incentives to seek private care may be doubly difficult. This is by no means to criticise consultants, but simply to note that human reactions to change are inevitably conditioned by concerns about income and livelihood, and that this needs to be taken into account in planning a new solution.

Several interviewees commented on the potential to provide alternative incentives, such as investment in new facilities or equipment, or indeed in additional consultants. While in some cases this might be seen as further undermining the potential to maintain levels of private income by increasing competition for patients, there were several reports that consultants were very keen to have new colleagues to reduce the burden of workload for themselves and improve the experience and quality of care for patients.

This brings us on to the broader motivation of consultants' concern for patients. While we did encounter one or two arguments that waiting was a good thing (perhaps as a deterrent, perhaps to allow time to see if an operation was really necessary), most interviewees agreed that consultants did not like having to make patients wait, and that they fully appreciated that reducing waiting times was in their patients' interest. One medical director commented, 'No doctor likes having to ring up a patient and cancel their operation – it feels terrible.'

There was considerable agreement that appealing to clinicians' motivation to do the best for their patients was more productive than top-down instruction. Peer discussion and comparison were, it was suggested, ways of encouraging poor performers to do better. It was also suggested that part of the problem is that consultants are rarely given sufficient information to judge for themselves what the problem is and how they might tackle it.

Inevitably, there will be some doctors who are not fully engaged in the process. A medical director at one of the most successful sites commented that some doctors locally were still influenced by the vision of the doctor as 'a lone soldier, [with] everyone else... scuttling round supporting them... Some people are still making the journey, and

we do have some dyed-in-the-wool dinosaurs... We have a number who are not entirely signed up.’ Others at this trust also commented that the organisation’s culture was not especially innovative or radical, but that even so, high levels of investment in new capacity and consistent focus on the issue had resulted in progress. Nonetheless, the medical director argued, ‘If you involve them in planning – and clinicians have been involved here – they’re more likely to be brought in from the cold.’

Concerns were sometimes expressed about the possible distortion of other clinical priorities as a result of the attention and effort given to waiting times reduction, even though the reasons for it were understood and felt to be worthwhile. However, this is another example of how sites with consistently low waiting times show clear differences: in these sites, there was no need to delay care for some individual patients in order to keep up with waiting time targets for less urgent cases, as everyone could be seen within a reasonable time. Any concerns about the priority given to waiting times in these sites were more in terms of other service developments that might have been foregone as a result of expenditure on waiting times.

In some cases, the fact that performance fell back once an initiative was exhausted had led to the feeling that sustained progress was impossible. However, some trusts noted the positive impetus to achieving waiting times reductions through ‘quick wins’. For example, in one trust successfully reducing very long waits in ophthalmology clinics through a combination of extra staff and spreading out the workload across three clinics demonstrated that a longstanding problem could be tackled and helped foster a ‘can do’ attitude.

Specific factors and strategies

To provide some structure to the many factors associated with good and poor performance on waiting times that emerged from our interviews, factors are grouped under the following headings:

- analysis, forecasting and planning
- organisational focus and persistence
- capacity
- efficiency of production processes.

The first two groups are, in a sense, activities that support and inform various strategies that we have identified within the groups of capacity and efficiency. Of course, this simplified taxonomy does not allow for overlaps between the issues and factors we have identified. We would also make the point that those sites that have been most successful usually ascribe their success to a whole array of actions. In themselves, each of these may be fairly small, but taken together have much greater impact.

We cannot list here every single tactic that has been used in all the sites, and in any case, each site will need to find the range of actions that best suits its needs, given its own history, its pattern of relationships and its physical and organisational structure. The important point is that both the availability of high quality information to aid planning and a strong organisational focus and leadership provide fertile conditions for such tactics to be devised and implemented.

Analysis, forecasting and planning

An overwhelming consensus was apparent from the successful and temporarily or partially successful trusts alike (and less so from poorly performing trusts) of the need for information that was reliable, detailed, comparative and continuous (daily, or even hourly). To characterise it bluntly, successful trusts would not find it hard to produce waiting times information for a named patient, while unsuccessful trusts found it hard to know whether to trust their own total waiting list figures. During site visits, it was immediately apparent that senior managers in successful sites had a precise grasp of their current situation, and could lay hands on the information necessary to answer questions about waiting times without any difficulty.

Two aspects relating to information are worth highlighting. First, successful trusts know (and others are beginning to realise) that tracking individual patients through the hospital system is vital. While some managers bemoaned the amount of detailed information they were receiving and the time it took to ensure that individual patients were chased up for appointments, they also intimated that for the individual with overall responsibility for waiting lists, this change in the nature of the job was probably permanent. Second, we were told of a number of examples of managers collating comparative waiting times and other performance data at the level of individual consultants. Discussing the variations that such data reveals has been the first step in persuading consultants to change their working practices. (*See also* Clinical ownership and involvement, p 34).

The need for information has also been a strong driver for centralising waiting list management in successful trusts. In fact, all trusts recognise the need for much greater control over the flows and movements of patients into, around, and out of their hospital. 'Centralisation' does not mean a complete management takeover of the referral and operating list processes. In one case, for example, among other things, it meant having one computerised office for admissions clerks, a standard 'Dear Doctor' referral letter (to help even out outpatient clinic workloads), and a version of 'earned autonomy' for consultants (with those managing their lists in a way that met targets being autonomous, and others agreeing to have their lists managed and 'profiled' on their behalf – using, for example, *CheckList*).

In another case, it meant having a co-ordinated waiting list 'team', headed by an executive director, which met weekly, or more often if needed, to review waiting times and match capacity and anticipated demand. Again, in trusts where waiting times are consistently low and consultant workflows are already well managed, the need for centralisation may be less relevant or take different forms.

Planning in good trusts also meant being ahead of the game – in particular, looking further ahead than the next looming waiting times milestone – and engaging in detailed capacity planning for the subsequent target or, in one case, already looking beyond the six-month target. In order to do this, successful trusts used capacity planning models and ensured that they had access to the right information to plan for changes in demand and consequent changes in capacity. Successful trusts also gave examples of how they not only tried to match capacity prospectively with planned workload, but also undertook retrospective reviews of what had actually happened, and analysed reasons for any discrepancies.

Example: Modelling elective activity

One trust used a computer-based model to work out what levels of activity it had to achieve to reduce waiting times and what resources it required to achieve the projected levels. The projection depended on a series of assumptions about referrals, the conversion rate (in other words, the proportion where decision to admit is made), the rate of day case activity, and average length of stay for inpatients (both benchmarked against good practice). It also involved estimating in broad terms the number of 'simple' and more complicated procedures, with the latter requiring more resources than the former. Actual activity levels could then be monitored against the project to check whether the trust was in line to achieve its target reductions.

It is hard to overstate the importance of good, frequent and reliable information. Even if trusts are conscious of the need to understand the whole system, including the wider health economy, they cannot turn this into a reality without the information to underpin the good intentions. (For further work on this area, see A framework for system-based information requirements for managing the supply of elective care, pp 77–92).

Organisational focus and persistence

All our 'successful' trusts noted the importance of making waiting times reduction an absolutely clear, unwavering priority. This in turn called for strong leadership, both managerial and clinical. Commitment and everyday involvement from the very top of the organisation were seen to be absolutely necessary in making progress on what all have noted are very tough targets. Good leadership also engenders commitment from other staff in the organisation, provides the channel to keep everyone continually focused on the task at hand, and – almost by definition – implies good working relationships between managers and clinical staff, and good communication generally, throughout the organisation. Not everyone may agree with the priority afforded to waiting times reduction, even if they think it is understandable, but the fact that it is an organisational priority is clearly understood and acted on.

However, organisational priorities may be upset by mergers, reorganisations and changes in senior personnel, which can deflect attention from focusing on waiting time reduction. On the other hand, two of the three most successful case-study sites had faced a threat of merger and had found it a stimulus to make radical improvements in performance. This echoes the evidence on re-engineering, which suggests that such radical solutions are more likely to work when organisations face crisis.

Organisational focus and persistence also includes the need to attract and retain experienced and skilled managerial staff – particularly directors of operations, or others with the main operational responsibility for meeting waiting times targets.

A number of managers in successful and temporarily successful trusts emphasised the need to persuade clinicians to own and internalise a commitment to reducing waiting times. Here, managerial tactics were almost as varied as the number of consultants. However, the use of comparative consultant-level waiting times and performance data (shared with consultants), and the argument that reducing waiting times was not just a government target but was what patients wanted (and was good for their health), were important.

Capacity

Having the resources to increase capacity (where it was identified as necessary) was clearly important, with the converse (not having the resources) seen as almost a guarantee of failure. One of the successful trusts had been very able at attracting waiting time and waiting list initiative monies in the past, and this has undoubtedly helped it achieve substantial reductions in waiting times. Another had benefited from a new hospital build with extra capacity planned in, even if that may not keep up with future rising demand.

On the other hand, there was a realisation among all trusts that previous ad hoc or one-off uses of such resources does not lead to sustainable reductions. Use of weekend working by trusts' own staff and ad hoc use of the private sector were not seen by many trusts as a long-term solution to capacity constraints, even though they expected to have to keep using those strategies for the foreseeable future. There was a clear view that temporary increases in capacity were essential as a short-term strategy to meet targets but were often wasteful and expensive and prevented the same money being invested in permanent capacity. The latter would probably yield greater efficiency in the long run, but the Government would either need to invest in parallel for an interim period, or would have to forego some short-term reductions in waiting times.

Many trusts mentioned the difficulties they were having to address to meet the requirement for financial balance, and the trade-offs they thought they would have to make with initiatives designed to reduce waiting times (for example, funding for a mobile day case unit).

Example: The whole-system perspective

This site was aware of the importance of taking a whole-system perspective. One example was its participation in the early development of intermediate care teams, which was described as a key factor for the trust in obtaining improvements in waiting times. The trust developed intermediate care in a joint venture with social services. Teams, led by one of the trust's consultants, delivered treatments to patients in their own home. This required good co-ordination, and took 12–18 months to get going effectively. The PCT had now taken on responsibility for this, with leadership continuing to come from the trust. Good intermediate care meant that the trust had fewer blocked beds and hence was getting more out of existing bed capacity.

However, if demand patterns continue to shift, even permanent increases in capacity may not solve the problem. As noted earlier, Site D had invested in a new admission and observation unit and a new step-down unit, seeking to tackle both incoming demand and delayed discharge, but the latest waiting time figures suggested that this had not helped as much as it had hoped. In other words, additional capacity is not always sufficient.

Efficiency of the production process

We have noted that some trusts had made use of short-term initiatives and had come to accept that they could not be sustained in the long run, and were in any case expensive in terms of cost per case. In contrast, successful trusts had begun to look in detail at the logistics of their hospital's care processes. This involved looking at the patient pathway, attempting to simplify and shorten it, identifying bottlenecks and pinch-points for the individual pathway, and then using the whole-hospital system perspective to work out, for example, the best way of handling the interaction between elective and emergency

flows. This might involve some degree of ringfencing and/or scheduling of elective care over the course of the year, so as to minimise the potential conflict between the needs of the elective and the emergency subsystems.

Within these broad strategies are a host of smaller measures bearing on efficiency – tight bed management, maximising day case activity, ensuring full utilisation of theatres, and effective discharge planning, including possibly investment in step-down facilities where the local private sector was inadequate. As noted at the start of this section, overall success is likely to depend on many small actions rather than one or two magic bullets, but these we believe are likely to be more effective if they are carried out within a whole-systems framework rather than in a piecemeal way.

Ultimately, efficiency is a matter of relating the resources available, the cost of deploying them, and the resulting output. We found virtually no evidence bearing on the cost of making specific changes to the production process even in the successful trusts.

Summary

The box that follows shows how some of the characteristics that have emerged from this study can be used to profile ‘successful’ and ‘unsuccessful’ trusts. This is not intended to imply that all trusts in these categories have all of the characteristics associated with that category. However, the attributes associated with success are generally necessary if sustained success is to be achieved. The middle group of ‘variable performance’ trusts exhibited some of the characteristics of both other groups, but their failure to sustain improvement appeared to be due to external shocks and their inability to react to them, rather than a fall-off in their own efforts.

Table 2: Sustaining improvement: a summary of trust characteristics

	Successful	Unsuccessful
Analysis, forecasting and planning	<ul style="list-style-type: none"> ■ close detailed monitoring of queues ■ good information at a detailed level ■ clear responsibility for planning and alerting operational managers of need for action ■ day-to-day connection between planning and operational sides ■ many people with a view of the whole system of elective care, and trying to manage through the system 	<ul style="list-style-type: none"> ■ poor information, and only at a very high level ■ no central function for managing queues ■ nobody seeing the whole elective care system within the hospital ■ no connection between the planning process and day-to-day management of care system
Organisational focus and persistence	<ul style="list-style-type: none"> ■ strong leadership ■ use of incentives and motivation of staff ■ long-term view ■ organisational stability 	<ul style="list-style-type: none"> ■ weak leadership ■ continual change in personnel ■ poor relations between managers and clinical staff ■ short-termism, which may be associated with getting the job done and moving on
Capacity	<ul style="list-style-type: none"> ■ understanding what capacity is needed to deal with demand, both now and in the future ■ ability to vary capacity when required to meet contingencies of variations in demand or other factors, such as surges in emergency care ■ getting extra resources by having a reputation for good management and use of resources ■ keeping demand for services within manageable limits 	<ul style="list-style-type: none"> ■ insufficient resources to meet needs resulting in always trying to catch up ■ not able to work in partnership with commissioners, other providers to bring in more resources ■ distortion of priorities resulting from strength of one or more pressure groups in trust – for example, in teaching hospitals where top research specialty gets more than its fair share

<p>Efficiency of production processes</p>	<ul style="list-style-type: none"> ■ understanding the underlying production processes of the elective care system in detail ■ strong performance management ■ good discharge-planning procedures ■ booking systems ■ use of comparative information on internal performance ■ strong clinical leadership resulting in better management of key workers in the elective care system 	<ul style="list-style-type: none"> ■ poor performance management ■ no use of comparative information to improve performance ■ an individual clinical approach to care rather than a managed approach through strong clinical leadership
--	---	--

Conclusions

This stage of the research focused on the results from the qualitative analysis of interviews. Thus the picture that emerges of factors that promote or inhibit the achievement of waiting times targets is restricted to the views of those interviewed. Further, while we have identified sets of factors that appear to be more common among successful trusts, we cannot attribute any magnitude of effect to each, as we do not know the importance of that factor individually, nor whether it is likely to be effective on its own or only in interaction with other factors in the set.

A key finding from this research is that at local levels (trusts and PCTs) and the national level (the Department of Health) alike, success in sustained reductions in waiting times involves a combination of knowledge, attitudes and activities. We identify four particularly important factors:

- a sustained focus on the task
- an understanding of the nature of waiting lists
- detailed information, analysis, forecasting, monitoring and planning
- development of appropriate capacity.

These are explained in detail below.

A sustained focus on the task

A clear and unambiguous message from successful trusts, and from those beginning to turn the corner on reducing waiting times, was the absolute necessity first, to focus the organisation on reducing waiting times, and second, to sustain management and clinical effort and priorities on the task. The energy and detailed day-to-day management, down to the tracking of individual patients through the hospital system, should not be underestimated. Bringing about this sustained focus requires skilled and strong, but sensitive, leadership and management at all levels of the trust.

An understanding of the nature of waiting lists

Understanding that waiting lists are not simply a backlog problem but the manifestation of a more complicated, dynamic flow through interconnected parts of a whole system of

care has enabled successful trusts to break down the problem, and to tackle those particular factors that, given their own circumstances, have given rise to long waiting times.

Detailed information, analysis, forecasting, monitoring and planning

Detailed, consistent and accurate time-series and cross-sectional information on waiting lists and times, as well as on key resources, provided successful trusts with a means to analyse and understand their waiting lists, to see them in context with other trusts, and to allow them to monitor progress and outcomes of changes in service delivery. This information also enabled them to plan future changes in services to meet targets, and to find the resources required to provide them.

Development of appropriate capacity

Lack of capacity can ultimately undermine efforts to reduce waiting times. Developing appropriate capacity, not just through increasing the totality of resources, but also through more efficient use of resources and managing the demand on those resources, was essential.

It is interesting to note how similar many of this study's findings are to conclusions reached by the Research into Practice Team's work on sustainability and spread in the national booking system, and its analysis of sharing and learning from the cancer services collaborative (Modernisation Agency 2002a, 2002b). Clearly – and perhaps unsurprisingly – there are general management lessons to be learned from successful attempts to bring about permanent change in the way trusts organise and deliver services.

The following section shows some of the elements that would help to bring about sustainable improvements in the management of waiting lists. These are mainly concerned with the need for a systematic and overarching analytic approach that requires intelligent gathering and use of information. But also essential to success is ensuring clear links and good working relationships between those responsible for the planning of the delivery of care and those delivering care, so that there is a confluence of aims, achievements and responsibility.

Key elements for managing the elective care system

Where the following refers to outputs or services, these are the various parts of the elective care process in the hospital, such as consultations, tests and investigations of various kinds and operative procedures.

1. Produce a 'menu of services' for each specialty, with estimates of time taken per service (for example, type of operation), and the level of expertise required. Using this menu, produce a detailed profile of demand for services (operations and consultations) on a daily basis, showing levels of variability in numbers referred, the difficulty of cases, and the time taken per service.
2. Produce detailed output schedules for the current configuration of services, indicating how many slots are available, with time allocations per slot, and the nature of the work that could be produced. Thus, there might be three categories of operation

(simple, difficult and very difficult); and three time-slots for operations (less than 30 minutes, less than one hour, and over one hour). Similarly, there might be two categories of outpatient consultation (five and 15 minutes). The important thing is to encourage staff to think about how they allocate patients in queues into these categories so that optimal use can be made of the time available. This requires prior knowledge, for example, of holiday rotas for all staff.

3. Produce an analysis of the potential constraints on output (other than availability of direct-contact staff, such as operating theatre staff, and theatre slots). These would include all testing and investigation issues (both pre- and post-operation or consultation) and care availability within the hospital setting (primarily, staffed beds). This requires an understanding of the needs of individual patients in terms of the number of bed days required (the profile of demand must take this element into account), care availability outside of the hospital, and availability of other professionals for activities such as rehabilitation (such as physiotherapists, occupational therapists and dieticians). This is needed to ensure availability of other 'indirect' inputs required to produce planned outputs of elective activity.
4. Plan for the flexible availability of additional physical capacity (such as theatres out of hours) to deal with unpredicted surges in demand. Options such as the use of private facilities could also be considered in the same way.
5. Produce an analysis of patients that do not attend for one reason or another, or are excluded from treatment at points in time (such as suspensions or cancellations), so that the impact on 'real' waits and on the efficient use of resources can be derived. This should lead to changes in procedures to ensure that patients are not inadvertently missing their slots (for example, booking systems of various kinds and near-date reminders, such as texting, letters, phone calls and emails). In all cases, the relative costs of any administrative procedures used should be assessed and matched against efficiency gains in the use of health care delivery resources and improvements to patient experience.
6. Produce a transparent costing of all of the various options proposed for producing services, based on a detailed investigation of the variable inputs, plus an appropriate allocation of fixed costs. Ensure that all members of staff understand these costings and (if possible) have agreed them as reasonably accurate.
7. Finally, if all of this is in place, then at the start of each year, produce a profile of demand and supply for each individual service and how this would be met throughout the year, together with a range of variability and how this would be dealt with. The profile would be based on best estimates of detailed activity in that year. There will be considerable variation in individual procedures, but this would be reduced through the use of time-related and difficulty-related categorisations, noted earlier.
8. Introduce a system of 'notional' booking for all patients in queues so that when a patient joins a queue there is a notional slot allocated (but not necessarily given to that patient). This could be applied to all outpatient slots, inpatient and day case treatments, and tests and investigations. It should also be possible for predictions to be made of when a treatment will be necessary where the patient does not require a treatment immediately, and a notional slot allocated, even in that case. This should prevent the 'chalking the dartboard' phenomenon that is sometimes observed. Eventually, the 'notional' system could become a live system for all queues. Of

course, some hospitals already have such systems in place for some of their activities. What is required is a roll-out from 'notional' to live systems.

9. Ensure that the progress of patients through the elective system is traceable by using a unique patient identifier. This would make it possible to produce estimates of total average waits in different parts of the elective system within the hospital.
10. Give consultants responsibility for the delivery of timely services to patients on their list. This requires a common agreement as each patient enters the queue as to the maximum time they should have to wait. Eventually, this process would be routinised, and would be carried out by administrative assistants, but to start with the consultant body would need to think things through in this way. Each consultant can produce a plan of how their own patients can be slotted into the overall work of the hospital. This should result in situations where the inconsistency of current arrangements is seen ahead of time, and changes are introduced to deal with these.
11. Work out the best way of managing the interface between emergency and elective care. This will involve forecasting the likely profile of emergency demand, drawing on Dixon and Damiani (2002) for example, the scope for reducing bed and nursing requirements for elective care by, for example, scheduling operations involving short hospital stays in times of likely peak demand, and working out the most effective and efficient balance between ringfenced and general pool beds.
12. Improve the referral process with the aim of reducing unnecessary referrals and clinic visits through, for instance, the use of protocols and other forms of closer working between primary and secondary care.
13. Redesign the main care pathways with a view to simplifying and speeding up access times for patients and reducing the costs of providing each service.
14. Assess (by reference to suitable comparators) the scope for improving productivity of staff and facilities – for example, number of operations per full time consultant or per theatre session.

2. Waiting times targets and distortions to clinical priorities

Arising from the first stage in this research was an important question concerning the possible clash between a central national strategy to reduce waiting times – maximum waiting times targets or guarantees – and priorities for admitting patients from waiting lists as determined by clinicians. In this section we report on our investigation into this issue and, in particular, an attempt to quantify the scale of any distortion in one specialty: trauma and orthopaedics.

Background

As part of its investigation into inpatient and outpatient waiting in the NHS in 2001, the National Audit Office carried out a survey in which nearly 300 (52 per cent) of a sample of 558 consultants in three specialties considered that ‘...working to meet NHS waiting list targets meant that they had to treat patients in a different order in 1999/2000 than their clinical priority indicated.’ One-fifth of consultants stated that treatment of patients in a different order had occurred frequently, and of the 300 consultants, 80 per cent stated that deferring treatment of ‘urgent’ patients had had a negative impact on patients’ health (National Audit Office 2001). It would appear from the survey that urgent cases were displaced by less urgent patients in danger of breaching the then-18-month waiting times target.

This survey, and indications from the first stage of the present study, suggest that there is concern among some clinicians about the impact that reducing waiting times is having on their clinical decisions to admit patients from waiting lists.

Such concerns are not unique to England. Although a search of the literature revealed few publications addressing the specific issue of the impact of waiting times targets on clinical admission priorities, there were a number of pertinent examples from Sweden, Norway and Denmark that were followed up with informal interviews with authors.

A version of the English waiting time targets, in the form of guarantees of treatment, have been tried in a number of European countries, and have met with similar concerns from clinicians. Guarantees have varied in type between ‘unconditional guarantees’, under which all patients are promised treatment by a certain time, and ‘conditional’ guarantees, which set some limits – for instance, applying only to certain conditions, or specifying that a percentage of patients (for example, 80 per cent) have to be treated within the time limit (Hurst and Siciliani 2003).

In Sweden, a guarantee was initiated in 1992 for patients suffering from certain conditions to be treated within three months or to be offered a choice of provider for alternative treatment. This was abandoned in 1996 – partly because waiting lists, which had initially fallen, began to rise again, and partly because of a lack of support from clinicians. Much of the evidence of hostility is anecdotal (resulting from interviews with key informants in the countries concerned), but some limited research has been attempted.

A survey of Swedish clinicians conducted on two occasions (both at the start of the policy in 1992 and between 1993 and 1994) revealed that the initial support of doctors for the guarantees had increasingly slipped away (Hanning and Spangberg 2000). Some doctors saw the guarantees as an intrusion into their clinical freedom. For example, cataracts were subject to a waiting time guarantee, prompting the following comment from a senior doctor in ophthalmology:

Patients with cataract certainly get the best care, but cataract does not lead to permanent blindness – which can actually be the case in many other illnesses such as diabetes, squint and retinal ablation – if the patients do not get care on time. However, those other patient groups have been partly set aside in the maximum waiting time discussion. (Hanning and Spangberg 2000)

The growing lack of support for the principle of guarantees on the part of clinicians has been seen as a key element in the failure of the initiative in the 1990s.

A similar sense of compromised clinical freedom emerged from research in Norway, where a waiting time guarantee was attempted, modified, and has now been abandoned. The initial form of the guarantee began in 1990 and applied to a list of conditions for severely ill patients. It guaranteed that treatment would be given within six months, but no specific sanction existed for failure to meet the guarantee (European Observatory on Health Care Systems 2000). The guarantee was subsequently modified in 1997. The time limit was reduced to three months but would only apply to those patients who would, in the judgement of the doctor, benefit from treatment and who fulfilled certain criteria of need (for instance, were on sick pay). The rate at which clinicians ‘awarded’ the guarantee varied enormously, with some giving it to 80 per cent of their patients, while others guaranteed only 20 per cent, even within the same hospital.

A postal survey of over 150 physicians and nurses revealed a view that the policy had not led to a more equitable situation for patients. Seventy-three per cent of chief physicians and nurses agreed with the statement ‘the waiting time guarantee has meant that some patients who ought to get treatment don’t get it’. In addition, ‘difficulties in interpreting the rules and infringements of the guarantee in respondents’ departments led to negative attitude towards the guarantee’ (Lian and Kristiansen 1998).

Norway’s guarantee has been superseded by a Patients’ Rights Act (1999), which came into force earlier this year. It sets a time limit for an initial consultation with a specialist (30 working days from referral) but none for treatment (Norway 1999).

In Denmark, several guarantees have been attempted and abandoned. The first was in 1993, in the form of a three-month maximum wait for all non-acute surgical treatment, to be achieved by 1995. The target was accompanied by a free choice of hospital. The targets were abandoned once it was clear they would not be reached. A new guarantee was tried for two treatments (knee and ruptured disc operations), which was again revoked in 1997 when it was not going to be met. ‘Since then the political approach has been to encourage a reduction in waiting times by allowing increases in health care funding but to avoid general legislative guarantees’ (European Observatory on Health Care Systems 2001).

The experience of ‘distortion’ in the Danish case seems to be slightly different. There is anecdotal evidence that waiting time guarantees, free choice of hospitals and a large injection of funds over the past two years may have distorted the emphasis in several ways – for instance, towards surgical procedures (particularly elective surgery) where

waiting times can relatively easily be measured, and away from other areas, such as geriatrics or psychiatry. There is also a suggestion that 'out-of-county' choice patients have been prioritised over home county patients because of the way payments are made to hospitals.

On the other hand, there has been no hard evidence of clinical distortion produced in any country, and certainly no sense of whether (or how much) patients might have suffered (or indeed, for some, benefited) as a result of enforced changes in clinical ordering. Experts interviewed as part of this current research commented that it was certainly in the interests of the clinicians to claim that patients were suffering (even though it is almost impossible to prove), as a bargaining tool for a softening, modification or abolition of the guarantees.

This highlights a difficulty for the present research in identifying the presence and magnitude of any distortionary effect of waiting times targets – namely, the lack of a 'gold standard' for the clinical priorities that would, or should, prevail in the absence of the managerial imperatives created by waiting times targets. In some cases, trusts have attempted to use condition-specific measurement tools to assist consistency in prioritising patients. More often, clinical priorities are determined by individual clinicians' judgements. This creates a challenge for comparing these priorities with those that emerge under waiting times targets.

First, clinical opinion is demonstrably idiosyncratic. Clinicians often offer quite different opinions on the degree of clinical need for patients. The Audit Commission's report *Access to Care: Ear, nose and throat and audiology services* (Audit Commission 2002) highlights differences in clinical opinion concerning the urgency with which patients requiring grommet operations should be admitted. In one trust, 20 per cent of clinicians stated that grommet patients should be operated on within one month, while in another trust ENT consultants stated that all these patients could be left waiting for three months or more.

Second, non-clinical factors may occasionally influence the priority accorded to particular patients. Examples include a clinician's views of the extent to which a patient's quality of life is affected by the severity of their clinical condition, the 'deservingness' of a patient, or the 'interest value' of treating an unusual case. The way in which clinical resources are deployed may also, to an extent, dictate or 'interfere' with the clinical order of patients. For example, the length of operating sessions may mean that from a trust's point of view, it is efficient to admit a mixture of cases (and hence, perhaps, a mix of urgent and non-urgent cases) requiring different operating times, in order to maximise the use of operating theatres.

Third, where patients are admitted from the waiting list in strict order of urgency, some non-urgent patients will either be temporarily or permanently displaced. Clinically, a hip replacement is always likely to have higher priority than, say, an excision of bunion. Therefore, where resources are limited, admitting according to clinical priority alone will mean that bunions are never likely to be treated, because there will always be another case of higher priority, and it is only by interfering with clinical priorities that some patients will be treated at all. Thus, the key issue in determining the impact of waiting times targets on the order and timeliness with which patients are treated is the relative costs to patients of admission criteria based solely on clinician's judgements and those based partly on clinicians' judgements, and partly on strategies designed to reduce maximum waiting times.

The key research question could, therefore, be: ‘Which prioritisation process minimises the loss of health status experienced by patients on waiting lists?’. Answering this requires the calculation of the actual loss of health status (for example, quality-adjusted life years (QALYs) suffered by matched samples of patients under these two sets of admission criteria. A further issue that naturally follows is whether it is then possible to specify a set of admission criteria that minimises this health status loss.

However, while these questions lie at the heart of the issue, data requirements (surveys of patients over time, to establish changes in health status arising from alternative approaches to clinical prioritisation) are beyond the time resources of the study. Therefore, we proceeded with a more modest investigation of the problem, using a combination of a survey of consultants in our original sample of nine trusts and an analysis of actual waiting list data.

Aims

The main aims of this stage of the study were:

- to identify the potential scale of distortions in clinical priorities arising from waiting times targets
- to identify factors common to trusts that have high levels of clinical distortion.

Methodology

The study aimed to investigate and compare the distribution of waiting times – both overall, and for particular groups of conditions in orthopaedics. An initial survey among consultants in four specialties (orthopaedics, general surgery, ENT and ophthalmology) across all nine of our original sample of trusts was carried out in order to establish the scale of the problem, and to provide any possible indications of the nature of and context for clashes in priorities.

Second, hospital episode statistics (HES) data for all trusts in England were analysed and actual waiting times distributions in two separate periods (2000/01 and 1997/98) were compared, in order to identify differences in distributions that may be attributable to the 2001/02 waiting times target. While we would expect similarities in both periods, we would hypothesise that the more recent year would show differences concerning the proportion of admissions of patients waiting around 15 months. We then investigated further at the level of procedures (health care resource groups, or HRGs) to see if there were changes in the composition of admitted cases. We also attempted to establish the extent to which there has been any substitution of long (less urgent) for short (more urgent) wait patients.

The next section describes the findings from the consultant survey and the HES analysis. This is followed by a summary of the results and an overall set of conclusions to the study.

Survey of consultants in eight trusts

A short questionnaire (see Appendix 4) was sent to 167 consultants by post. Of the nine trusts that participated in the original research, one refused to participate (Trust B). Within the remaining eight trusts, consultants from four specialties (ENT, trauma and orthopaedics, general surgery and ophthalmology) were surveyed. A total of 78 questionnaires were returned, giving a response rate of 46.7 per cent.

Just over one-third (33.8 per cent) of the sample reported that during 2002/03 they had treated patients 'in a different order to that suggested by their clinical priority'.

These consultants were then asked what impact they felt this had had on their patients – both those patients who had been seen sooner (as a result of meeting the waiting list targets) and those they had felt had been delayed. Of those seen sooner, over half the consultants felt there had been no clinical impact on patients, but a further 40 per cent felt that there had been a minor positive effect (see Table 2).

Table 3: Clinical impact on patients seen sooner (n=21)

Clinical impact?	Number	Percentage
Major – negative	0	0
Minor – negative	1.0	4.8
No impact	11.0	52.4
Minor – positive	9.0	42.8
Major – positive	0	0
Totals	21.0	100.0

Of those who responded to the question on the impact on those patients whose treatment had been delayed, over half felt there had been a 'minor negative impact', while a further third judged the negative impact as 'major' (see Table 3).

Table 4: Clinical impact on delayed patients (n=22)

Clinical impact?	Number	Percentage
Major – negative	6.0	27.3
Minor – negative	12.0	54.5
No impact	4.0	18.2
Minor – positive	0	0
Major – positive	0	0
Totals	22.0	100.0

There was less clarity about just how many patients were affected, with numbers ranging between 6 and 200 (per consultant), with an average of 36 patients being seen 'sooner' as a result of actions taken to meet the March 2003 maximum target of 12 months, while an average of 37 were delayed during the year. (However, only nine consultants provided information on the numbers affected.)

Accepting the caveat of a small overall sample size, there were some notable differences in the behaviour of trusts as ranked by their waiting list performance. The original trusts sampled were selected primarily on the basis of their trend performance on the

percentage of their patients waiting six months or less. In Table 4, trusts have been grouped into good (Group 1), middling (Group 2) and poor (Group 3) performance.

Only one consultant from Group 1 claimed to have admitted patients (number unknown) out of clinical order, compared with 11 from Group 2 (42.3 per cent) and 12 from Group 3 (42.8 per cent). In the worst performing trust from Group 3, nine out of 10 responding consultants complained of distortion to their clinical priorities. One interpretation is that trusts that have organised their capacity and admissions systematically to meet waiting times targets can achieve those targets without distorting clinical priorities that would otherwise have prevailed in the absence of those targets, while trusts with less systematic responses in place are more likely to rely on short-term actions, including displacing urgent cases, in order to meet their targets.

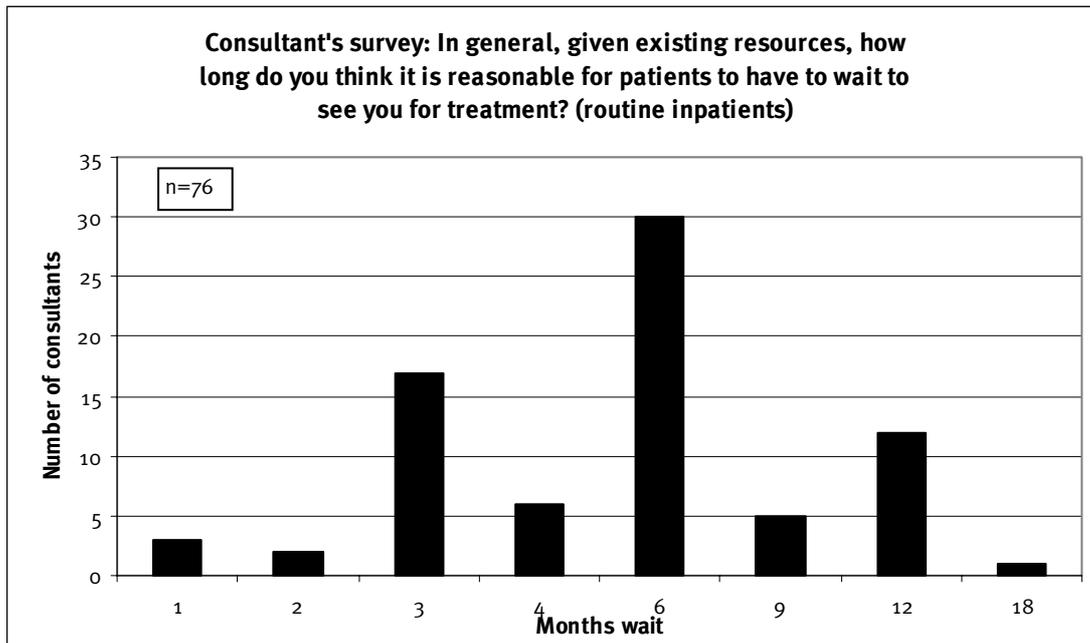
The survey also asked consultants what action they had taken to meet their waiting list targets, with possible responses ranging from 'none, as no-one waiting more than 12 months' to holding 'additional theatre sessions', both within and outside the working week. Consultants in trusts in Group 1 engaged in very little extra activity of the sort described in the table. By contrast, those in Group 2 engaged in rather more activity than Group 3.

Table 5: Actions taken to meet the March 2003 12-month waiting time target

Action	Group 1 n=14		Group 2 n=26		Group 3 n=28	
	no	%	no	%	no	%
Additional sessions within working week	1	7.1	17	65.4	12	42.8
Additional session outside working week	1	7.1	16	61.5	13	46.4
Transfer to private provider	1	7.1	18	69.2	8	28.5
Transfer to another consultant	0	-	14	53.8	18	64.3
Deferred	1	7.1	6	23.1	3	10.7
Reclassified as outpatient	0	-	1	3.8	0	-
Potential 'breachers' accommodated within existing theatre sessions	0	-	9	34.2	8	28.6

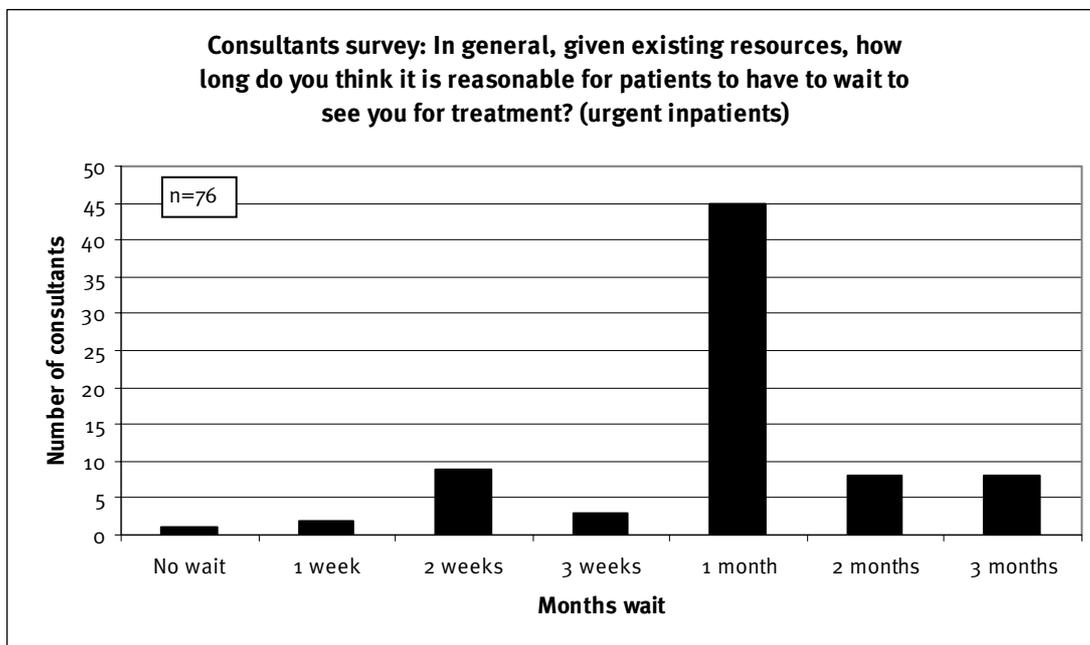
Consultants were also asked how long, in their opinion, it was reasonable for patients to wait, both for first outpatient appointments (urgent and routine) and for admission from the inpatient waiting list (urgent and routine). As expected, there was a wide variety of responses, reflecting a lack of broad consensus behind a key element of clinical decision making. However, despite the range, there were some patterns emerging in what might constitute a 'reasonable' wait for inpatient urgent and routine patients (see Figures 6 and 7). This was particularly true in the case of urgent patients, where a considerable majority felt one month was a reasonable wait.

Figure 6: Consultant survey – routine inpatients



King's Fund (2003)

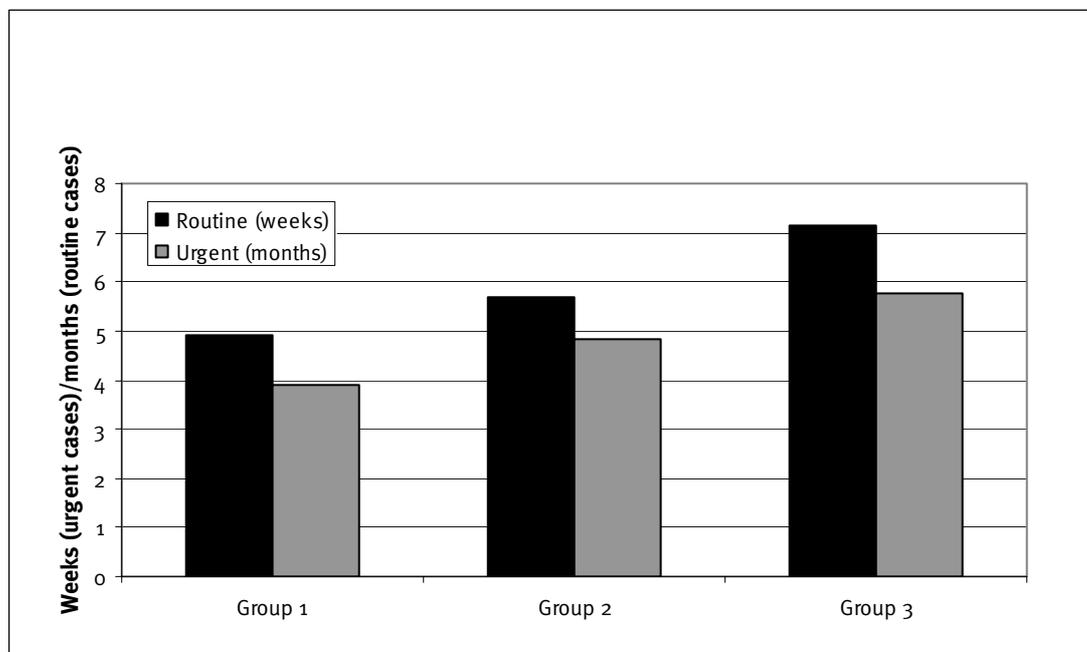
Figure 7: Consultants survey – urgent inpatients



King's Fund (2003)

Interestingly, the mean 'reasonable' urgent and routine waits by trust group reveal a significant difference between good, middling and poorly performing trusts, with average waiting times increasing from good to poor, to the extent that the mean urgent wait for Group 1 trusts is lower than the mean routine wait for Group 3 trusts (see Figure 8).

Figure 8: Reasonable waiting times for urgent and routine patients: average times by trust group



King's Fund (2003)

Finally, the survey asked consultants for examples of changes in clinical priorities, and other comments relating to waiting list targets. Views tended to be forthcoming from those who felt distortions had been taking place rather than from those who did not report a problem. The box that follows lists a selection of the comments received. The majority of the 'other comments' were hostile to the setting of waiting time targets, though not universally so. (For additional comments, see Appendix 5.)

Patients being treated in a different order to their clinical priority

Examples

- Mainly joint replacements being brought forward
- Mainly joint replacements delayed to accommodate those waiting a long time
- Intraocular surgery delayed in preference to minor operations, such as cysts, so that larger numbers could be put through
- Disfiguring and uncomfortable mucky eyelid procedures delayed to give opportunity to non-urgent cataract patients
- Removal of ingrown nail: expedited. Bunions: expedited. Total hip replacement, clinically urgent: delayed

General comments

- ‘Clinical prioritisation disappeared with arbitrary timescale imposition.’
- ‘I am appalled that management influence is allowed to alter clinical priority. We are not unpaid election agents for the Labour Party.’
- ‘I don’t feel that the above compromised patient’s care as the seriously ill/cancer patients or those in pain were treated ASAP.’
- ‘I use a booked admission system for 98 per cent of my patients. This allows me to control my workload. It should be used more widely. It must be accompanied by ringfenced elective beds because cancellations are disruptive.’
- ‘Manipulation of minor ops to meet numbers targets means that we have used up “fillers” between major operations. The overall effect is to reduce numbers for next year.’
- ‘Effort to meet targets caused immense stress to staff. Patients are often not able to tell the difference between a 12-month wait and an 18-month wait.’

Summary

The survey of consultants has revealed that around one-third (26) felt that they had had to treat patients out of clinical order, as they saw it. Very few, however, could quantify this apparent distortion in terms of numbers of patients affected. (Estimates ranged from 6 to 200 per consultant).

Around 80 per cent (12) of those who stated that their priorities had been distorted also stated that the impact on delayed patients had been negative clinically. However, around 40 per cent (9) stated that there had been some positive clinical impact on patients seen sooner than the consultant would have done in the absence of the waiting times target, with just over half (11) stating that there had been no clinical impact on patients seen sooner. While there is no intention in presenting these results to imply that it is straightforward to measure trade-offs between patients in terms of costs and benefits, ignoring the benefits to those seen sooner would also be misleading in terms of assessing the overall impact of the waiting times target.

When analysing the results by the three groups of trusts in our sample, for those consultants who stated their clinical priorities had been distorted it was found that only one consultant (out of 14) from trusts traditionally performing well on waiting times felt their clinical priorities had been affected as a result of the waiting times target. This may initially seem unsurprising given that these trusts had no one waiting over 15 months during 2001/02. Existing management arrangements and resources had been sufficient to achieve the target. On the other hand, it might be expected that as waiting times fall to the levels currently experienced by those in Group 1, pressures on clinical priorities for admission might increase. But this does not appear to be the case.

Around 42 per cent of consultants (23) from the other two groups of trusts (traditionally performing less well on waiting times) stated that their clinical priorities had been adversely affected by waiting times targets for 2002/03. These groups were also more liable than those trusts performing well to have taken a range of actions – extra sessions at weekends, and so on – to tackle long lists.

Finally, the survey confirmed much previous research that there is no standard clinical opinion concerning how long it is reasonable to expect urgent and routine patients to wait. Consultants stated that routine patients could reasonably wait between one and 18 months, with a mode at six months and sub-modes at three and 12 months. For urgent

cases there was more agreement, with the majority stating up to one month as a reasonable wait, but the variation ranged from no wait at all to up to three months.

The point at issue here is that applying this variation in clinical opinion to waiting times would produce very different distributions (or, depending on different consultants' views, 'distortions' in treatment priorities). For example, if it is assumed that around 15 per cent of all orthopaedic cases are urgent, for 2001/02 this would give a figure of around 64,000 patients who were admitted during that year. In fact, 62,000 were actually admitted who had waited up to and including three weeks (in line with the views of 60 per cent of consultants we surveyed). But if the views of the 13 per cent of our sample of consultants who stated that urgent patients should be admitted within two weeks is applied to the distribution of actual admissions, around 20,000 urgent patients breached this.

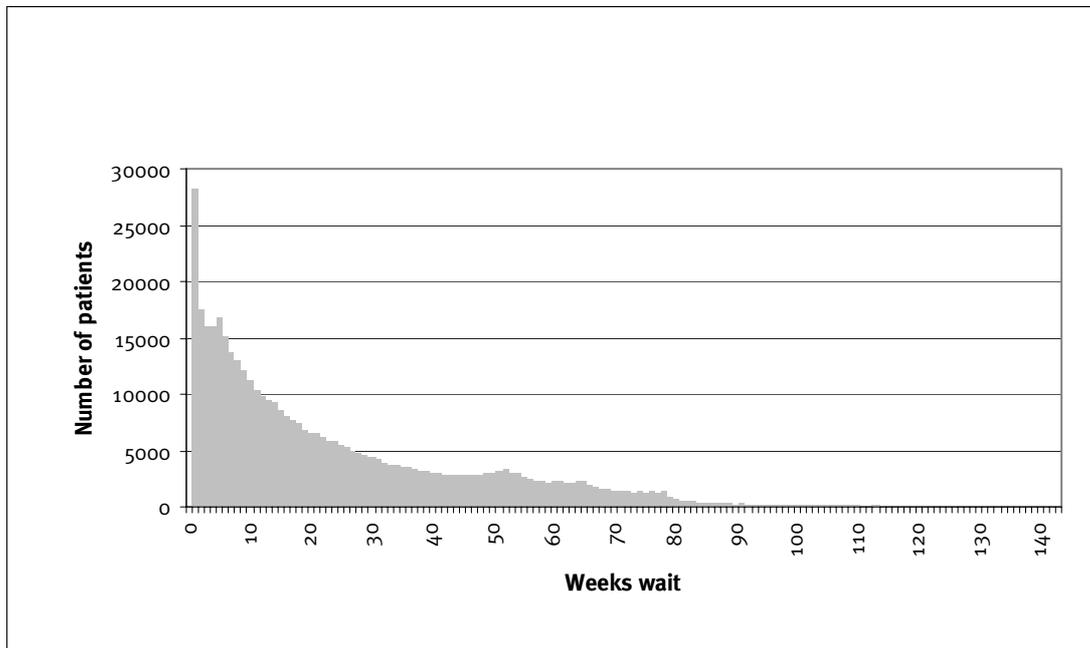
Indications of clinical 'distortion'

Hospital episode statistics (HES) are a potentially useful source of information for analysing indications of possible clinical distortions arising from efforts to meet maximum waiting times targets. Here, we examine trauma and orthopaedic admission data for England, and in particular, analyse data on the distributions of the time that patients waited prior to being admitted into hospital from waiting lists.

A priori, given the maximum target wait of 15 months in place for 2001/02, we might expect to detect some anomalies in the waiting times distribution – for example, the possibility of an increase in the numbers of patients admitted who had been waiting around 15 months and over in 2001/02, compared with previous years – with the possible implication that other, shorter-wait patients had been displaced. (If patients coming up to, and currently exceeding, 15-month waits were not admitted, they would appear on the waiting list census at March 31 and hence breach the target. In addition, there may also be anomalies in the timing with which short- and long-wait patients are admitted during the year, as a result of efforts to meet the 15-month waiting time target and/or the types of cases in terms of procedures.

An initial perusal of the trauma and orthopaedic waiting times distribution in England for those patients admitted during 2001/02 (see Figure 9) reveals a small 'blip' in admissions for patients waiting around 15 months. All the statistics include booked patients and those admitted from the waiting list. Interestingly, it also shows another blip for patients waiting around 12 months, and possibly another for those waiting around 18 months.

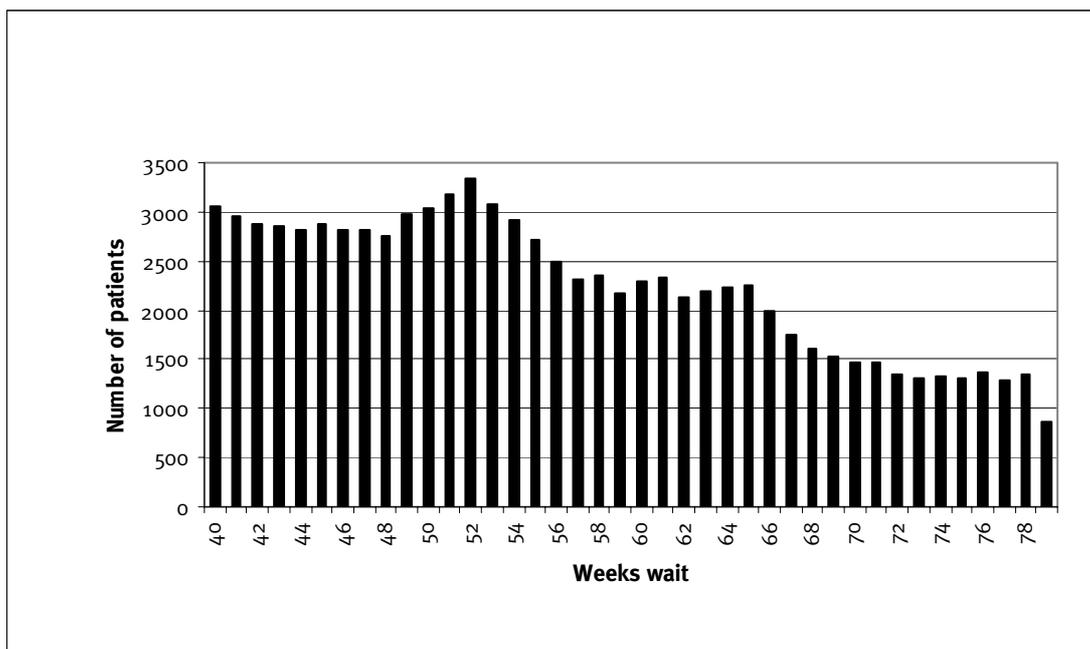
Figure 9: Trauma and orthopaedics 2001/02: all patients admitted from waiting list, weeks waited prior to admission



King's Fund (2003)

Figure 10 shows these changes in admissions more clearly.

Figure 10: Trauma and orthopaedics 2001/02: all patients admitted from waiting list, weeks waited for admission



King's Fund (2003)

The distribution of waiting times for patients admitted from waiting lists suggests that extra effort was made to admit patients who had been waiting around 15 months, but also for those who had been waiting around 12 months (and to a lesser degree, patients waiting 18 months or 78 weeks). If this is the case, then it provides, at best, an indication that in meeting the 2001/02 15-month maximum inpatient waiting time, some change may have occurred in the 'natural' order in which patients were admitted from waiting lists. (It should be emphasised that the existence of increased admissions does not in itself prove that distortions in clinical priorities occurred and that conversely distortion may have occurred even in the absence of any increase in admissions.) However, these data raise a number of further questions:

- **Expected waiting time distribution** Can the apparent additional admissions for those waiting around 12 and 15 months be quantified? That is, what would we expect the waiting times distribution of those admitted for treatment to look like? Quantifying the apparent 'additional' admissions would provide a rough guide to an estimate of the scale of possible distortionary effects of maximum waiting times targets.
- **Displacement of other patients by 'additional' admissions** Assuming an expected distribution can be calculated (and hence the additional admissions quantified), is it possible to determine whether additional admissions displaced other types of patients, either in terms of the types of procedures carried out or in terms of the time patients waited?
- **'Deadline effect'** Given that the 2001/02 maximum waiting times target deadline was March 2002, is there a 'deadline effect', with additional patients tending to be admitted in the fourth quarter of the year? If so, this would indicate problems with the way trusts tackled the task of meeting the waiting time target, concentrating effort (and hence, perhaps, increasing the likelihood of distorting clinical priorities) into the weeks and months prior to the March target deadline.
- **'Excess' admissions** The 2001/02 target was that no patients should be waiting more than 15 months, so why is there also an apparent increase in admissions for patients waiting around 12 months and, to a lesser degree, 18 months?
- **Comparison of 2001/02 and 1997/98 distribution** How does the 2001/02 waiting times distribution compare with a period before the implementation of the 15-month target? A difference between distributions – particularly for admissions of patients waiting around 15 months – would lend support to the case that the 2001/02 maximum waiting times target affected admission behaviour.

The following sections tackle these questions in turn. Given the problem noted earlier of the lack of a 'gold standard' in terms of what might be expected to constitute a 'clinically correct' distribution of patients admitted for treatment from waiting lists, we have employed two methods in an attempt to 'fix' a baseline against which to compare observed waiting times distributions. The first is purely statistical and assumes a predictable 'decay' in the number of admissions based on the time patients actually waited prior to admission. The second method takes a 'before-and-after' approach, comparing waiting times distributions before the implementation of the current target regime (1997/98) with distributions after its implementation (2001/02).

A general assumption running through the following analyses is that waiting times distributions reflect reasonable and systematic clinical judgements about admission

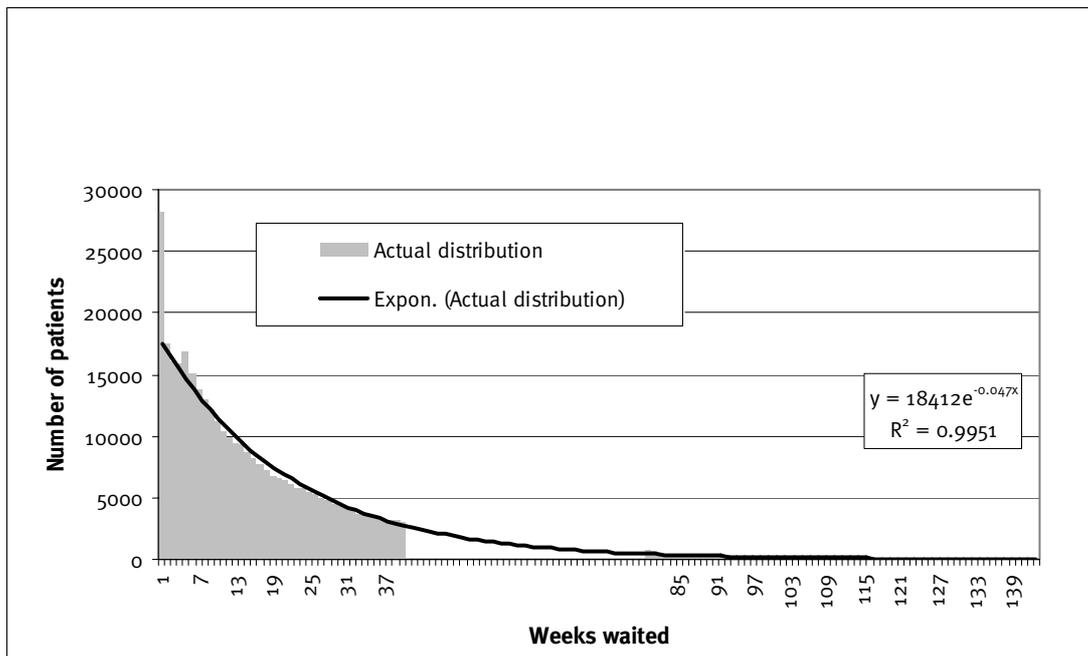
priority; that those admitted sooner are those whose conditions tend to be most severe, while those admitted later tend to be patients whose conditions are less severe.

Expected waiting times distribution

Assuming that the blips at 12, 15 and 18 months are in some sense anomalous – that is, not a feature of pure clinical judgements concerning admission priorities – one way to calculate an expected waiting times distribution is to exclude all admissions around the time of the apparent excess admissions and then estimate a ‘best-fit’ line to the remaining data. Figure 11 shows the result of excluding all admissions between 40 and 80 weeks and the resultant best-fit curve.

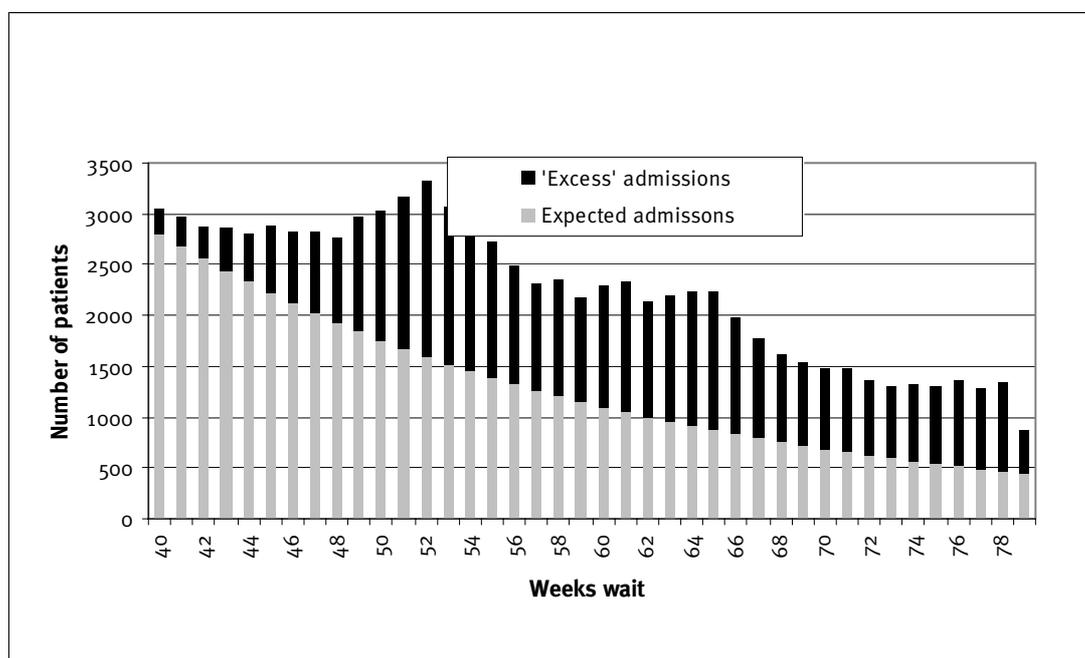
The excess can be calculated by using this estimate for the expected waiting times distribution for the period in which excess admissions occurred – that is, between 40 and 80 weeks (see Figure 12). In total, excess admissions amounted to around 38,200 patients – equivalent to 8.9 per cent of all admissions (428,457) over the whole year. Compared with the expected number of admissions for patients waiting between 40 and 80 weeks, this excess also represents an average weekly addition of around 42 per cent of total admissions.

Figure 11: Trauma and orthopaedics 2001/02: all patients admitted from waiting list, weeks waited prior to admission – actual and expected distribution



King's Fund (2003)

Figure 12: Trauma and orthopaedics 2001/02: all patients admitted from waiting list, weeks waited – ‘excess’ admissions between 40 and 79 weeks



King's Fund (2003)

Displacement of other patients by ‘additional’ admissions

There are two approaches to answering whether the apparent ‘additional’ 40–80-week wait patients admitted during 2001/02 displaced other types of patients. First, certain (minor) procedures may displace other types of (major) procedures. Secondly, longer (less urgent) wait patients may displace shorter (more urgent) wait patients. There may well be an association between minor and major, and less urgent and urgent.

First, in terms of procedures, of the top 21 highest volume trauma and orthopaedic health care resource groups, accounting for 90 per cent of all admissions during 2001/02, 12 displayed waiting times distributions with noticeably increased admissions for patients waiting 12 and/or 15 and/or 18 months. Figures 8 to 19 show waiting times distributions for patients admitted in 2001/02 for these nine procedures, together with ‘best-fit’ expected distributions (with the ‘excess’ admissions between 40 and 80 weeks highlighted).

Table 5 details the number of ‘excess’ admissions, by week, for each HRG. Overall, the 12 HRGs account for 77.4 per cent – that is, 29,583 out of 38,216 – of all excess admissions, compared with 56.7 per cent for patients who had waited between one and 39 weeks and 80 weeks or more. The table also shows that the composition of the excess admissions workload differs from that for patients who had waited between 1 to 39 weeks and 80 weeks or more. In particular, the final column of the table shows that proportionately more (for eight out of the 12 procedures) were carried out.

It should be noted that we would not expect to account for the total 'excess' admissions of 38,216 completely, except by some statistical fluke, as the estimates for this total and for the individual procedures used different best-fit curves – and, in any case, there will be a degree of uncertainty surrounding all estimates.

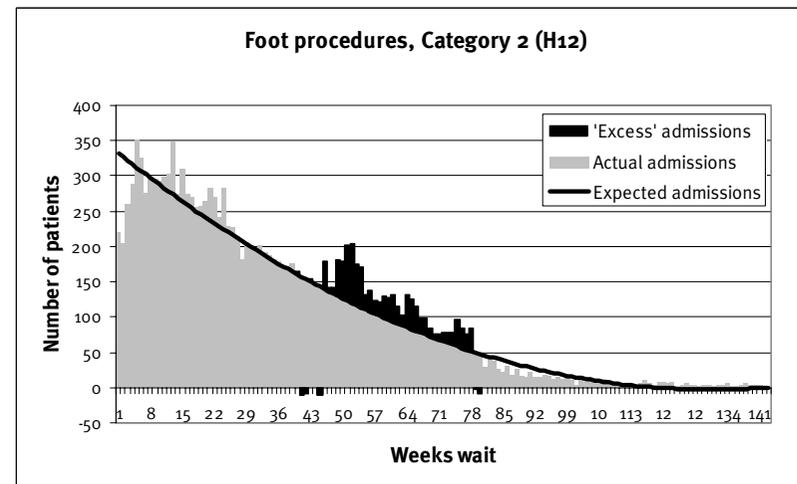
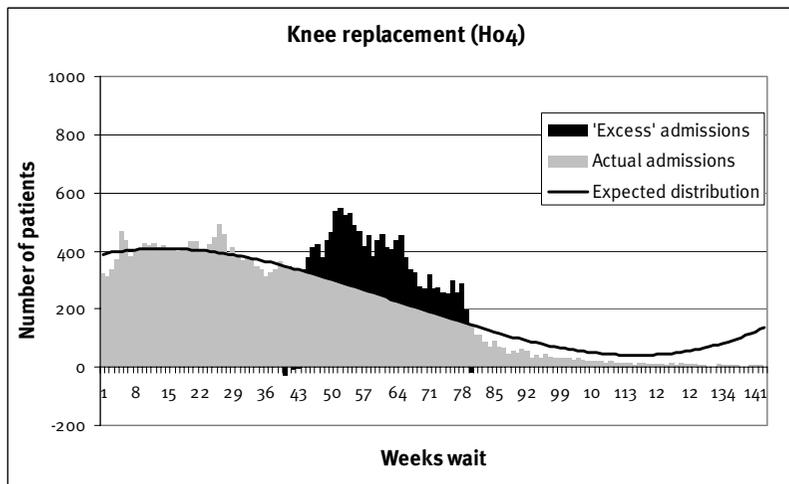
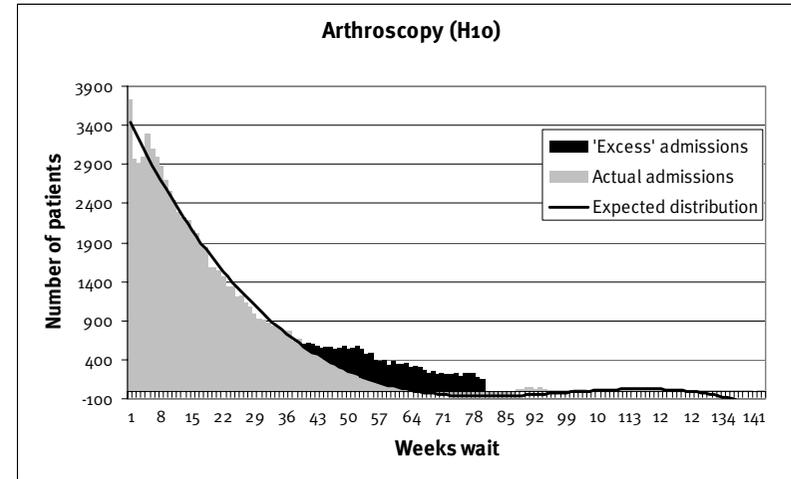
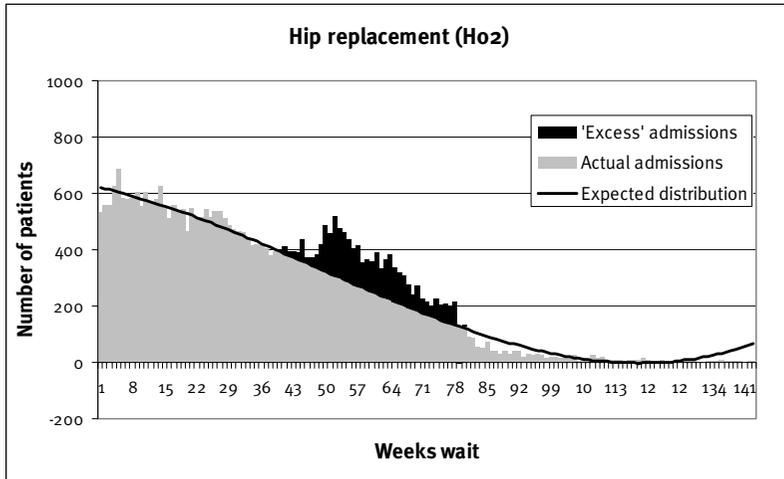
One observation to make concerning the composition of 'excess' admissions is that there is no obvious bias towards admission of less major procedures. For example, the proportion of admissions of patients undergoing primary knee replacements more than doubled, and hip replacements increased by one-third.

Table 6: Composition of 'excess' admissions

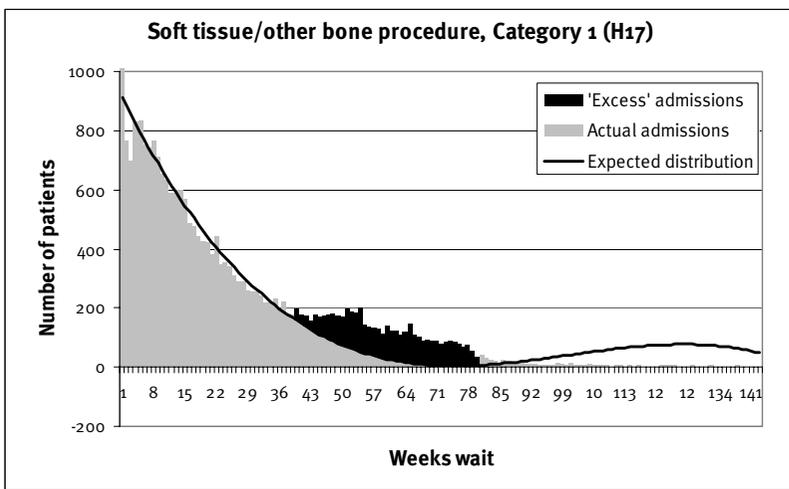
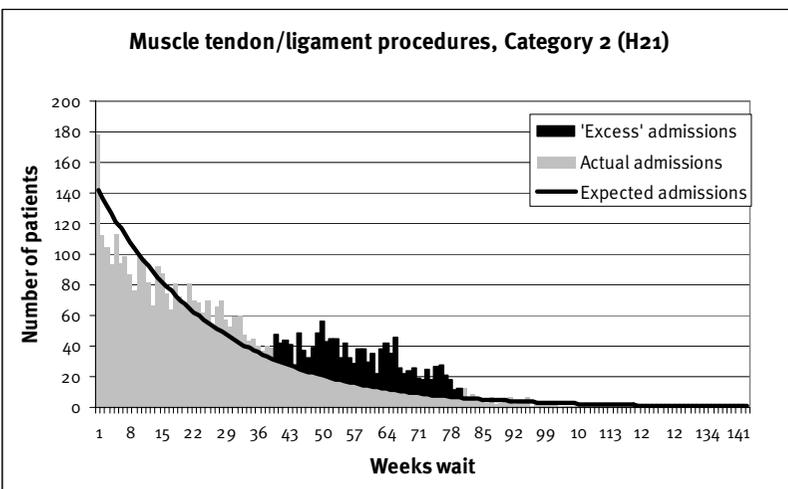
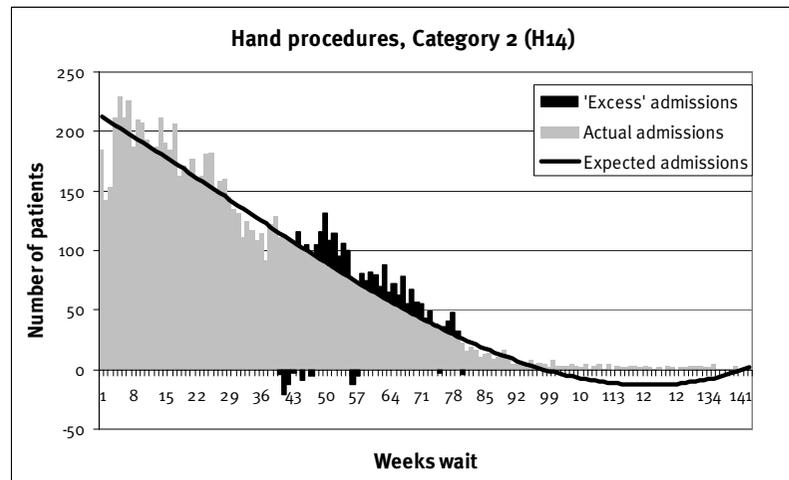
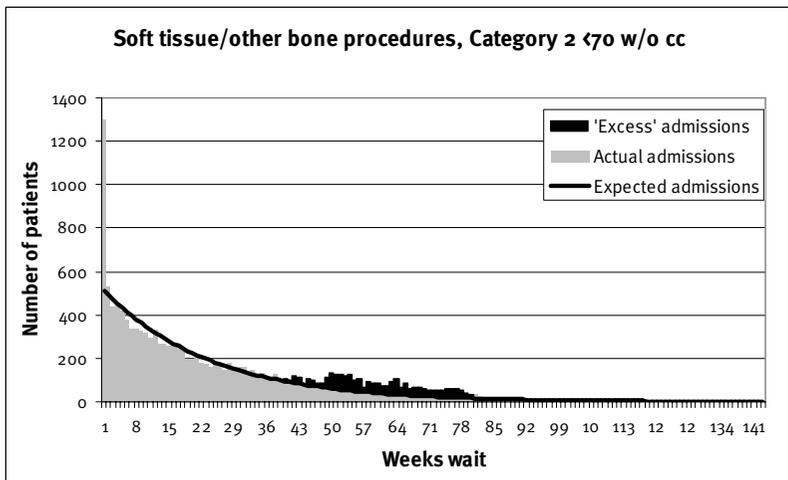
		Total admissions: 2001/02		Total admissions, excluding those waiting 40–80 weeks		'Excess' admissions – those waiting 40–80 weeks		Difference (Col 6 – Col 4)
		Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7
		no	%	no	% total admissions excluding 40–80 week waiters	no	% total excess admis- sions	
Arthroscopies	H10	85,839	20.0	72,202	21.4	9,931	26.0	4.6
All other HRGs		43,487	10.1					
Hand procedures – Category 1	H13	39,734	9.3					
Primary hip replacement	H02	35,509	8.3	23,567	7.0	3,498	9.2	2.2
Primary knee replacement	H04	32,304	7.5	19,606	5.8	5,029	13.2	7.4
Soft tissue/ other bone procedures – Category 1 less than 70 w/o cc	H17	24,244	5.7	19,651	5.8	3,444	9.0	3.2
Minor procedures to the musculo- skeletal system	H22	22,604	5.3					
Planned procedures not carried out	S22	16,732	3.9	12,433	3.7	1,955	5.1	1.4
Intermediate pain procedures	A07	16,257	3.8					
Foot procedures – Category 2	H12	15,317	3.6	10,264	3.0	992	2.6	-0.4

Removal of fixation device less than 70 w/o cc	H52	15,158	3.5					
Soft tissue/ other bone procedures – Category 2 less than 70 w/o cc	H19	14,035	3.3	10,564	3.1	1,605	4.2	1.1
Hand procedures – Category 2	H14	9,908	2.3	6,762	2.0	356	0.9	-1.1
Invalid primary diagnosis	U01	9,274	2.2	6,997	2.1	1,171	3.1	1.1
Minor skin procedures – Category 1 w/o cc	J37	9,093	2.1					
Muscle, tendon/ligament procedures – Category 1	H20	8,454	2.0					
Revision procedures to hips/knees	Ho6	7,458	1.7	5,882	1.7	571	1.5	-0.2
Foot procedures – Category 1	H11	6,892	1.6					
Muscle, tendon, ligament procedures – Category 2	H21	4,383	1.0	3,030	0.9	716	1.9	1.0
Surgery for degenerative spinal disorder	Ro2	4,096	1.0					
Spinal fusion/ decompression excluding trauma	Ro3	3,953	0.9	3,043	0.9	315	0.8	-0.1
Soft tissue/ other bone procedures – Category 1 more than 69 or wcc	H16	3,726	0.9					
			Sub-totals	194,001	56.7	29,583	77.4	20.1
Grand totals		428,457	100.0	337,952	100.0	38,216	100.0	

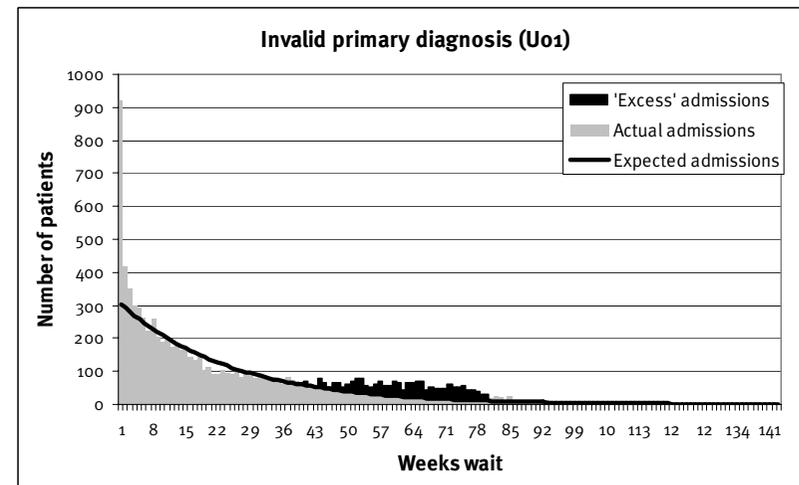
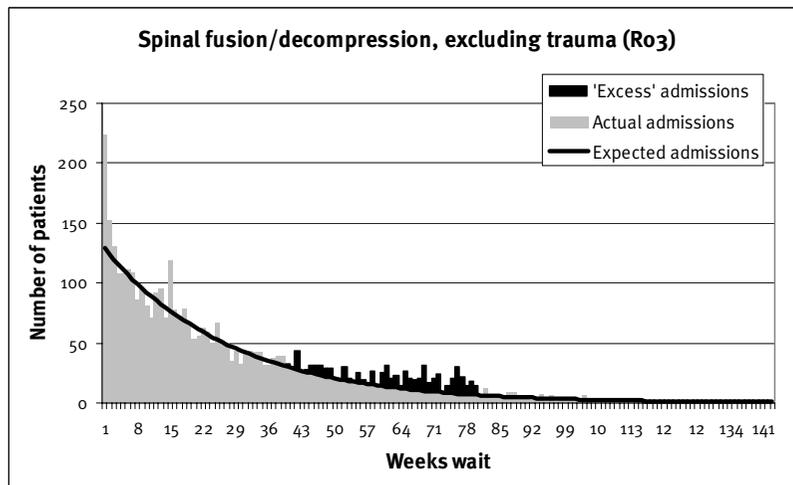
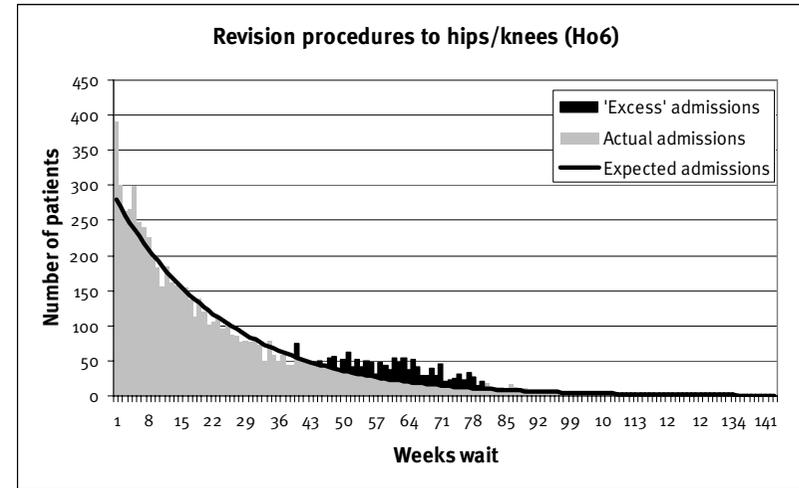
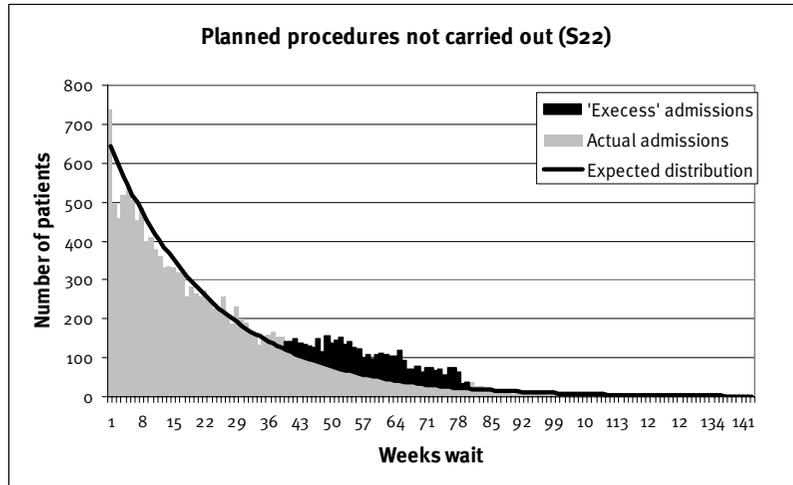
Figures 13–16: Expected and actual waiting times distributions: trauma and orthopaedic patients admitted during 2001/02



Figures 17–20: Expected and actual waiting times distributions: trauma and orthopaedic patients admitted during 2001/2 (continued)



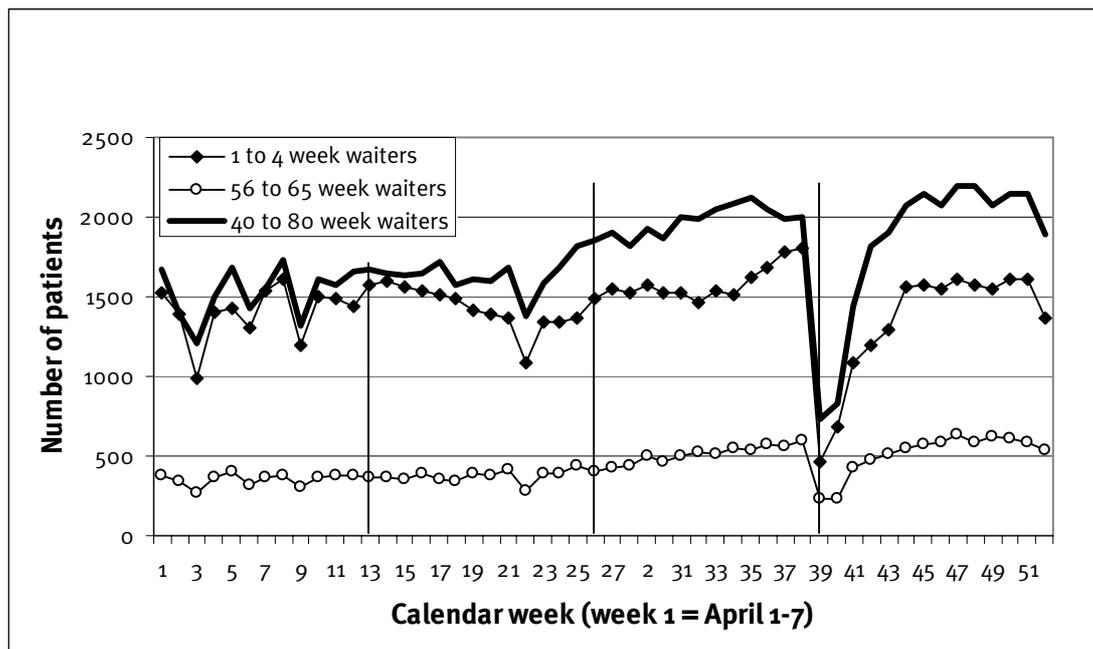
Figures 21–24: Expected and actual waiting times distributions: Trauma and orthopaedic patients admitted during 2001/2 (continued)



An alternative type of displacement could be long-wait patients usurping shorter-wait patients. If this were the case, and assuming that length of wait reflects clinical urgency, then this would lend support to the contention that the 15-month maximum waiting times target adversely affected the clinical priority with which patients were admitted from waiting lists.

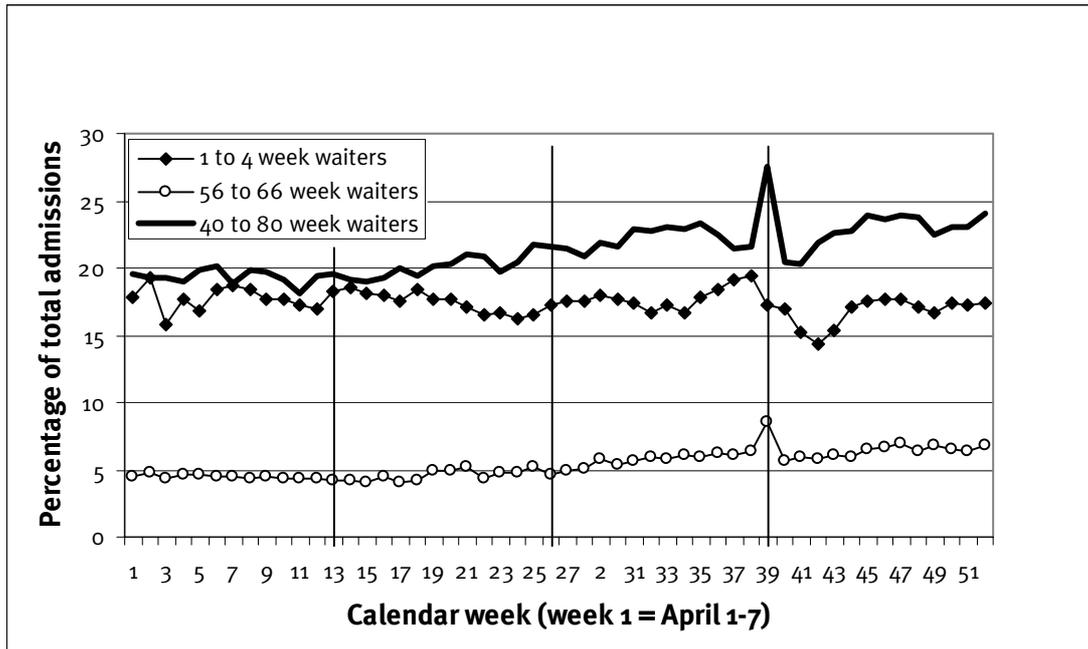
Figures 25 and 26 show the actual number and percentage of all admissions of short- and long-wait patients, and the timing of their admission during 2001/02.

Figure 25: Trauma and orthopaedics 2001/02: admissions of short- and long-wait patients by calendar week – actual numbers



King's Fund (2003)

Figure 26: Trauma and orthopaedics 2001/02: admissions of short- and long-wait patients, by calendar week – percentage of total admissions



King's Fund (2003)

Apart from the striking dip in admissions of both long- and short-wait patients in week 39 (Christmas week), there appears to be no obvious trade-off between the admissions of short-wait patients (that is, those who waited one-to-four weeks) with either measure of long-wait patients (that is, those who waited between 56 and 65 weeks – around 15 months – or those who waited between 40 and 80 weeks prior to admission).

Admissions in Christmas week and the first few weeks of the new year (2002) are not surprising, but the magnitude of the fall – particularly for those patients who waited between one and four weeks – is large. Indeed, admissions for short- and long-wait patients alike only fully recover five weeks into the new year. Although there appears to be some compensatory increases in short-wait admissions in the weeks prior to Christmas, this raises an issue of the flexibility of the clinical criteria used to admit patients from the waiting list. Three other dips in admissions occur during the year: Easter (week 3), May Bank Holiday (week 9), and the week of the August Bank Holiday (week 22).

Although from the beginning of October 2001 there is a slight upward trend in the proportion of admissions of patients who waited around 15 months (and those who waited between 40 and 80 weeks), there is no concomitant decrease in the proportion of short-wait patients admitted. A caveat to bear in mind is that the data may be too coarse to reveal displacements – particularly if the number of displacements is relatively small and the period of displacement shorter than one week.

'Deadline effect'

Given the method used to monitor compliance with the 15-month maximum-wait target (a census of patients remaining on waiting lists at March 31), we might expect to observe a

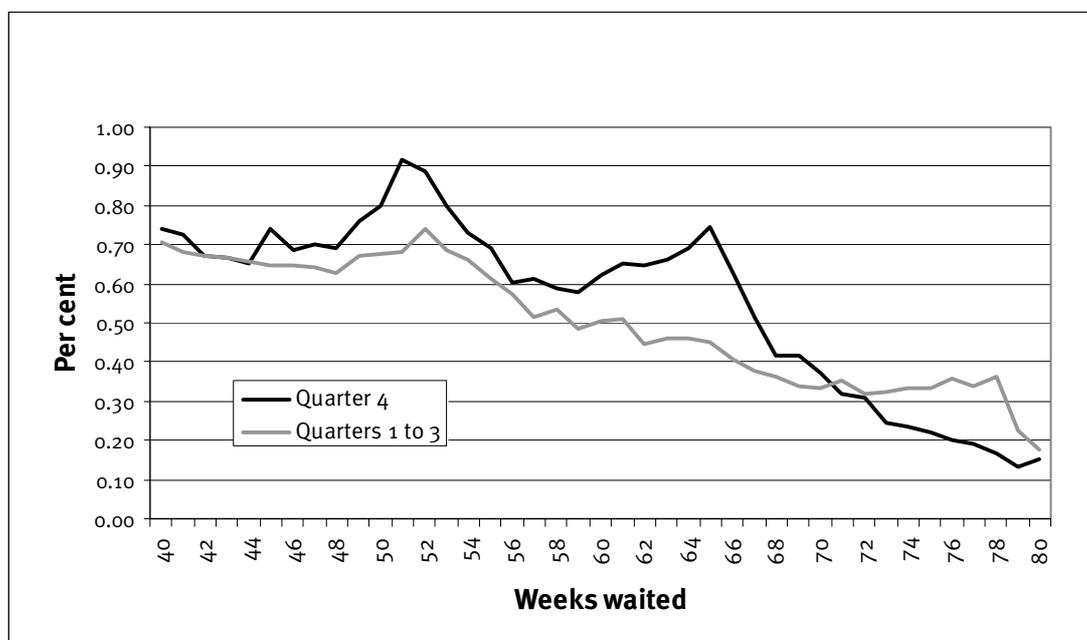
'deadline effect', with potential breachers of the target (those waiting more than 15 months) tending to be admitted in increasing numbers as the target deadline approached. If this were the case, then it might increase the likelihood of clinical admission priorities being distorted. Figures 25 and 26 suggest that there may indeed be a 'deadline effect', with increasing proportions of long-wait patients being admitted towards the end of the year (that is, in the weeks and months up to March 31). Figure 27 shows more clearly that admissions of patients waiting around 15 months peak more strongly in the fourth quarter than in the previous three. However, there is also a peak for patients who had been waiting around 12 months, and, in quarters 1 to 3, for patients waiting 12 months and 18 months.

Moreover, the mean and median waiting times for all patients admitted in the final quarter were 175 days and 115.5 days respectively, while equivalent times for quarters 1 to 3 were around five days less, at 169.4 days and 109.2 days respectively.

Using the same methodology for estimating the number of 'excess' admissions for patients who had waited between 40 and 80 weeks, as above, excess admissions in quarters 1 to 3 for patients waiting between 40 and 80 weeks represented an average increase in workload of 40 per cent for that period. For the final quarter, however, the equivalent increase was 48 per cent. Figure 28 details additional admissions as a proportion of total actual admissions, on a weekly basis, for quarters 1 to 3 and for quarter 4.

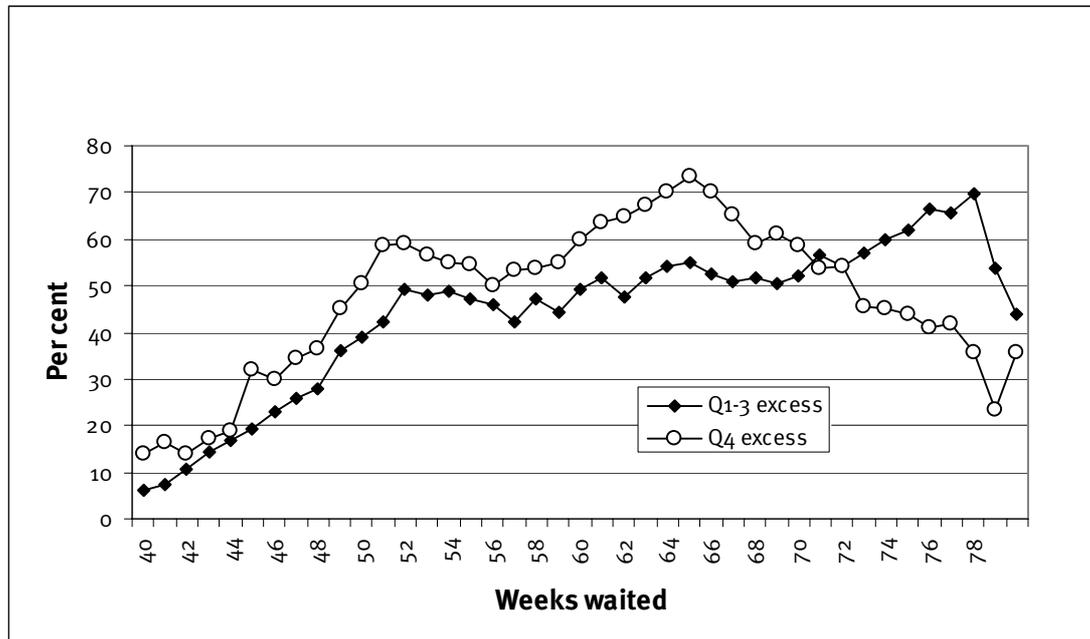
Given that the monitoring of waiting times targets has since moved on to a continuous basis (that is, with no one waiting more than the target time at any time during the year), the deadline effects noted here may no longer be a problem, and possible distortions to clinical admission priorities reduced.

Figure 27: Trauma and orthopaedics: quarter 4 and quarters 1–3 admissions from waiting list by time waited



King's Fund (2003)

Figure 28: 'Excess' admissions, quarters 1–3 and quarter 4 as a percentage of actual weekly admissions



King's Fund (2003)

'Excess' admissions

In both the total admissions and the quarterly breakdown admissions distributions, there are noticeable peaks not only at 15 months, but also at 12 and 18 months. The quarterly breakdown reveals a slightly different picture, with peaks at 12 and 15 months for quarter 4, and peaks at 12 and 18 months for quarters 1 to 3.

While an increase in admissions of patients waiting around 15 months could be linked to the 15-month waiting times target, it is more difficult to explain the peaks at 12 and 18 months. However, as is clear from a comparison between 2001/02 and 1997/98 admissions distributions (*see below*), the peaks in the 2001/02 distributions at 12 and 18 months almost exactly match similar peaks in the 1997/98 data. This might suggest that these peaks in admissions reflect a more longstanding aspect or factor in clinicians' admission criteria, related not to need but to the time patients have waited – one year and 18 months in some sense representing limits beyond which certain patients should not wait.

Comparison of 2001/02 and 1997/98 distributions

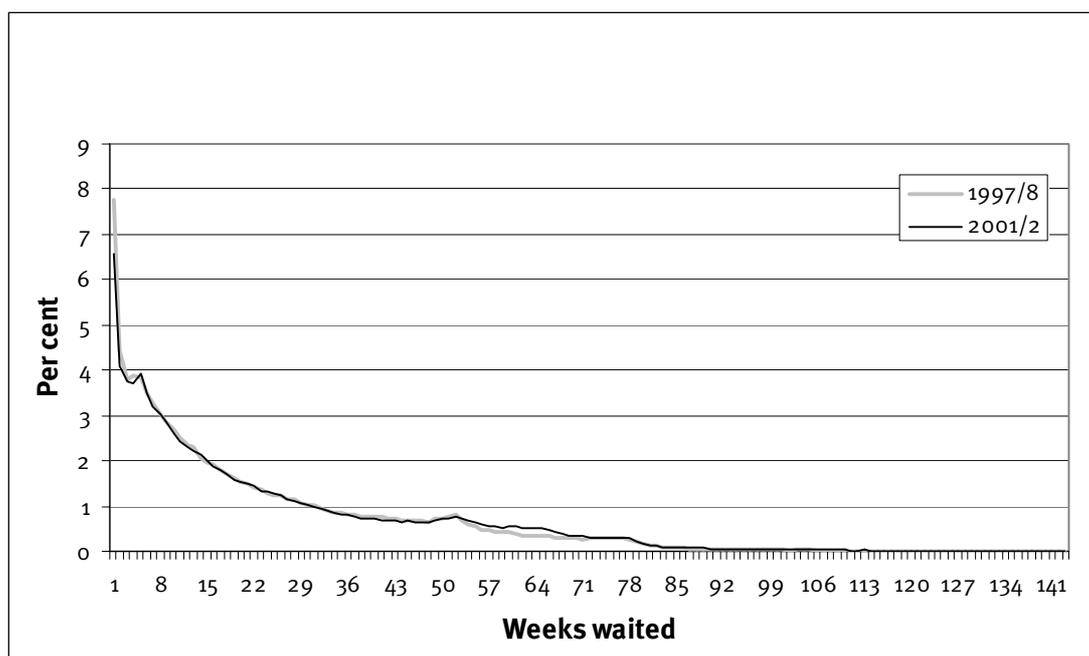
Support for the case emerging from the foregoing analysis that the waiting times target in 2001/02 changed admission behaviour is suggested by a comparison between waiting times distributions for 2001/02 and a previous period – 1997/98 – before the implementation of the target.

Figure 29 shows the waiting times distributions (as a proportion of total admissions) for English trauma and orthopaedics patients for 2001/02 and 1997/98. The distributions

are very similar, except for a divergence for patients waiting between 53 and 78 weeks. Figure 30 (which shows the weekly difference between 2001/02 and 1997/98 in the proportion of admissions) highlights this divergence, and very clearly shows the difference peaking at around 65 weeks, or 15 months. There is also a notable drop in the proportion of admissions of those waiting one week (which is discussed below).

This comparative analysis suggests an alternative way of quantifying the apparent 'excess' admissions for 2001/02: using evidence of the percentage difference in distributions as a basis for recalculation.

Figure 29: Waiting times distributions: trauma and orthopaedics, all patients, England – 2001/02 and 1997/98



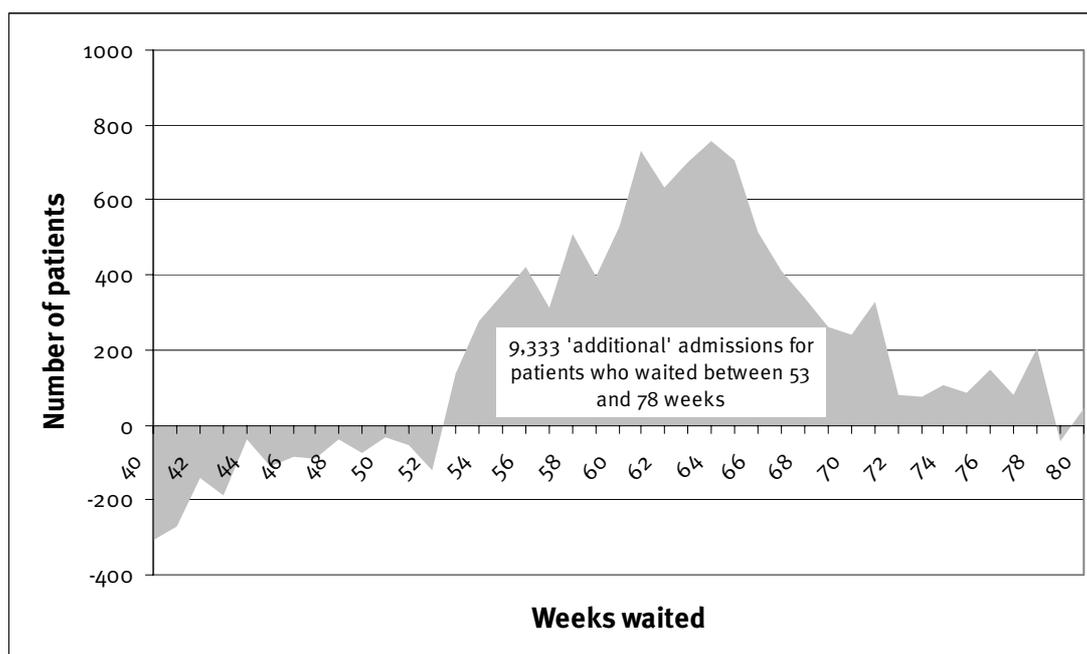
King's Fund (2003)

Figure 30: Difference in proportion of patients admitted – 2001/02



King's Fund (2003)

Taking the period for patients waiting between 40 and 80 weeks, Figure 31 shows the excess admissions in 2001/02 over those in 1997/98, taking into account the overall increase in admissions between 1997/98 and 2001/02.

Figure 31: Additional admissions in 2001/02 over and above 1997/98

King's Fund (2003)

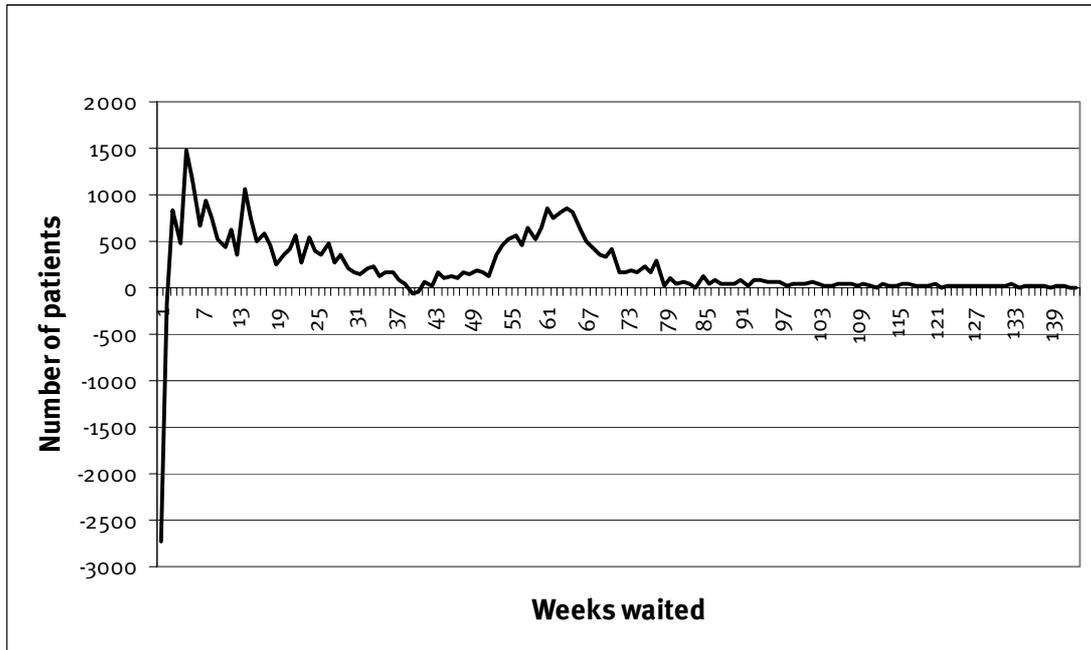
On this basis, the number of 'excess' admissions in 2001/02 for patients who waited between 53 and 78 weeks is estimated to be 9,333 – significantly less than the 38,216 previously estimated using 'best-fit' curves to derive 'expected' numbers of admissions – and representing just 2.2 per cent of the total number of admissions for the whole year. This difference in estimates is largely due to the apparent 'excess' admissions of patients who had waited around 12 months dropping out of the figures.

Comparisons between 1997/98 and 2001/02 waiting times distributions may also provide some indication of whether the additional admissions identified in Figure 18 potentially displaced or delayed other admissions, either in terms of the types of procedures carried out or in terms of the time other patients waited.

All other things being equal, the total increase in trauma and orthopaedic admissions between 1997/98 and 2001/02 of 7.6 per cent (398,198 to 428,457) could be reflected in an equivalent proportional increase in admissions for all patients, no matter how long they waited before being admitted. However, as Figure 30 shows, the overall increase in admissions was not spread equally across all waiting time categories. In particular, there were around 2 per cent (7,625) fewer patients who had waited one-to-four weeks admitted in 2001/02 than would be expected if the overall increase in admissions had been equally distributed. In other words, the fall in admissions of patients waiting between one and four weeks accounts for a large proportion of the additional patients who waited between 53 and 78 weeks (in other words, 9,333).

Moreover, as Figure 32 shows, taking the difference between actual admissions in 1997/98 and 2001/02 (rather than differences in proportions) again reveals the increase in admissions of patients who had waited around 15 months. It also shows that those who waited between one and two weeks fell (by 2,901) in absolute terms, and that those who waited between two and four weeks rose (by 2,779) – perhaps indicating a delay in admission.

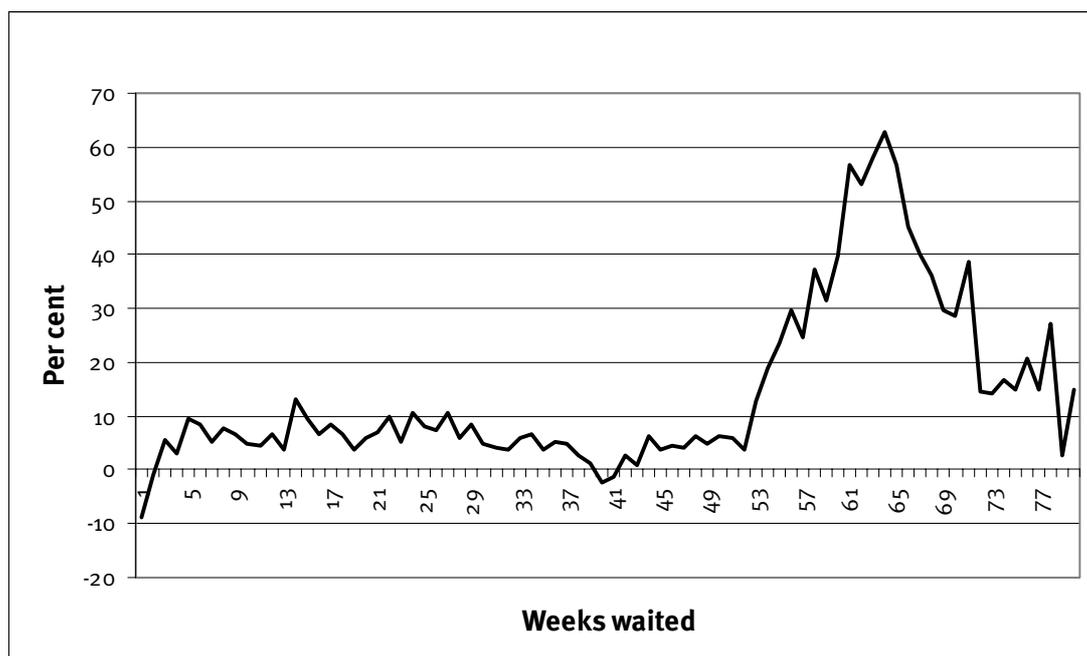
Figure 32: Difference in actual admissions: trauma and orthopaedics – 1997/98 and 2001/02



King's Fund (2003)

As Figure 33 shows, from the point of view of the percentage change in admissions by waiting time, admissions in all categories of waiting time increased – most markedly for patients waiting around 15 months – except for those who waited one week and those who waited 40 weeks. (The graph is truncated at 80 weeks as inpatient numbers are small and give rise to large percentage changes.)

Figure 33: Percentage change in all trauma and orthopaedics admissions – 1997/98 to 2001/02 by time waited



King's Fund (2003)

The fall in the numbers of admissions of patients who waited one week cannot be taken as prima facie evidence of a direct substitution of long-wait for short-wait patients, as other factors may be responsible for the change in the distribution of the waiting lists between 1997/98 and 2001/02. For example, it is not clear why very short-wait patients (presumably urgent cases) would take the brunt of any substitution for long-wait patients.

It might be supposed that if any substitution took place, it would be those patients who had already waited some time but whose delay in admission would not impinge on the waiting times target who would be most at risk. This would include those admitted between 32 and 52 weeks – a group that in Figure 32 shows a decrease of around 2,000 compared with 1997/98, as opposed to an increase of 9,333 in those admitted who had been waiting around 15 months. However, this result is not easy to resolve with the fact that the trend in admissions of short- and long-wait patients over the calendar year did not show the sort of inverse relationship that might be expected if substitution had occurred (see Figures 25 and 26).

Apart from potential displacements or substitutions with regard to the time patients waited before admission, it is also possible that the nature of the trauma and orthopaedic workload, in terms of types of procedures carried out, also changed between 1997/98 and 2001/02.

In total, the ten health care resource groups in Table 5 accounted for around 96 per cent of the net 'excess' admissions of patients who waited between 53 and 78 weeks prior to admission in 2001/02. Hip- and knee-replacement cases accounted for over 56 per cent of all 'excess' admissions.

The (percentage) waiting times distributions for hip and knee replacements both show the most marked shifts between 1997/98 and 2001/02, as can be seen in Figures 34 and 35. While in general, more patients of every category of time waited were admitted in 2001/02 compared with 1997/98, proportionally more long-wait (and particularly those waiting around 15 months) were admitted. As a consequence, average waiting times increased (by nearly five weeks for hip replacements and by four weeks for knee replacements), and the spread of the distribution reduced (standard deviations fell).

Overall, therefore, not only did the composition (in terms of procedures) of the admissions of patients who had waited between 53 and 78 weeks change, but for the two procedures accounting for the majority of the 'excess' admissions of those waiting 53–78 weeks, proportionally fewer patients waiting between one and 44 weeks were admitted (although the actual numbers of nearly all categories of patients increased).

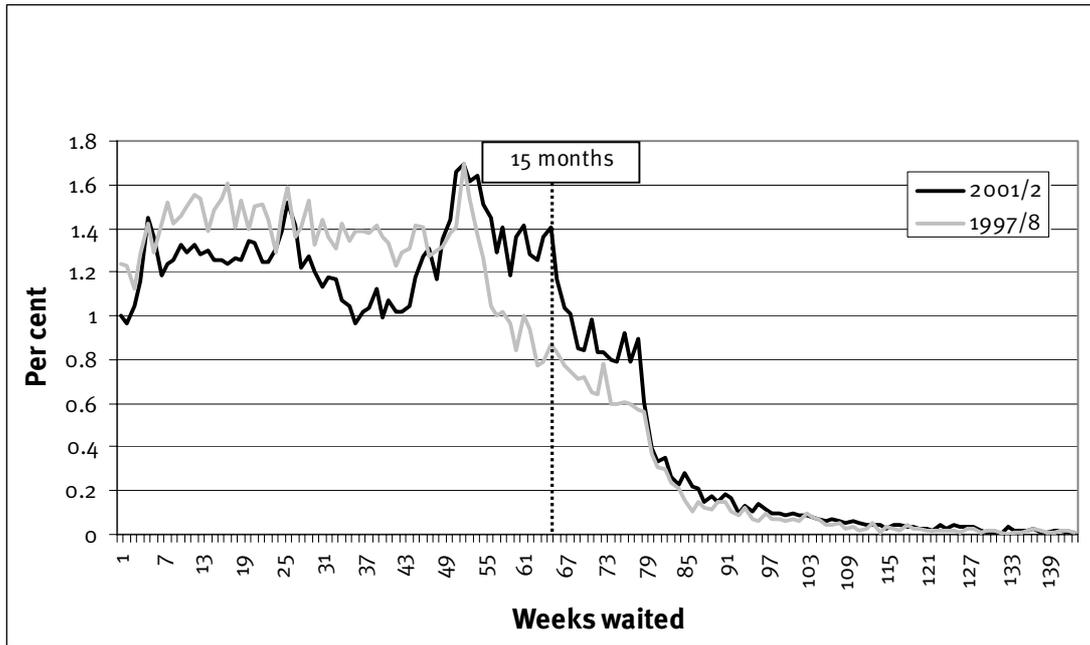
Table 7: Changes in composition of trauma and orthopaedics admissions – 1997/98 to 2001/02

Health care resource groups	Total admissions	Total admissions	Admissions of patients who waited	Admissions of patients who waited	Estimated 'excess' admissions of patients who waited	'Excess' admissions as % of all 'excess' admissions
	2001/02 Col 1	1997/98 Col 2	53–78 weeks 2001/02 Col 3	53–78 weeks 1997/98 Col 4	53–78 weeks Col 5	Col 6
	No	No	No	No	No	%
Primary hip replacement	35,509	28,858	8,196	4,367	2,823	30.2
Primary knee replacement	32,304	22,195	9,670	4,937	2,484	26.6
Soft tissue/Other bone procedures – Category 1 less than 70 w/o cc	24,244	23,354	2,939	2,298	553	5.9
Planned procedures not carried out	16,732	15,102	2,417	1,832	387	4.1
Foot procedures – Category 2	15,317	13,570	2,864	2,251	323	3.5
Soft tissue/other bone procedures – Category 2 less than 70 w/o cc	14,035	14,631	1,993	1,658	403	4.3
Invalid primary diagnosis	9,274	5,916	1,477	518	665	7.1
Revision procedures to hips/knees	7,458	6,837	987	658	269	2.9
Muscle tendon/ligament procedures – Category 2	4,383	3,918	776	505	211	2.3
Spinal fusion/decompression excluding trauma	3,953	4,177	544	421	146	1.6
Subtotal	163,209	138,558	31,863	19,445	8,952	
% of all trauma and orthopaedics	38.1	34.8	63.0	50.8	96.0	
All trauma and orthopaedics	428,457	398,198	50,556	38,312	9333	2.2

1 'Excess' admissions = additional admissions over and above increase in total admissions between 1997/98 and 2001/02 for the HRG – that is, excess = (Col 3 – (Col1/Col2) x Col 4).

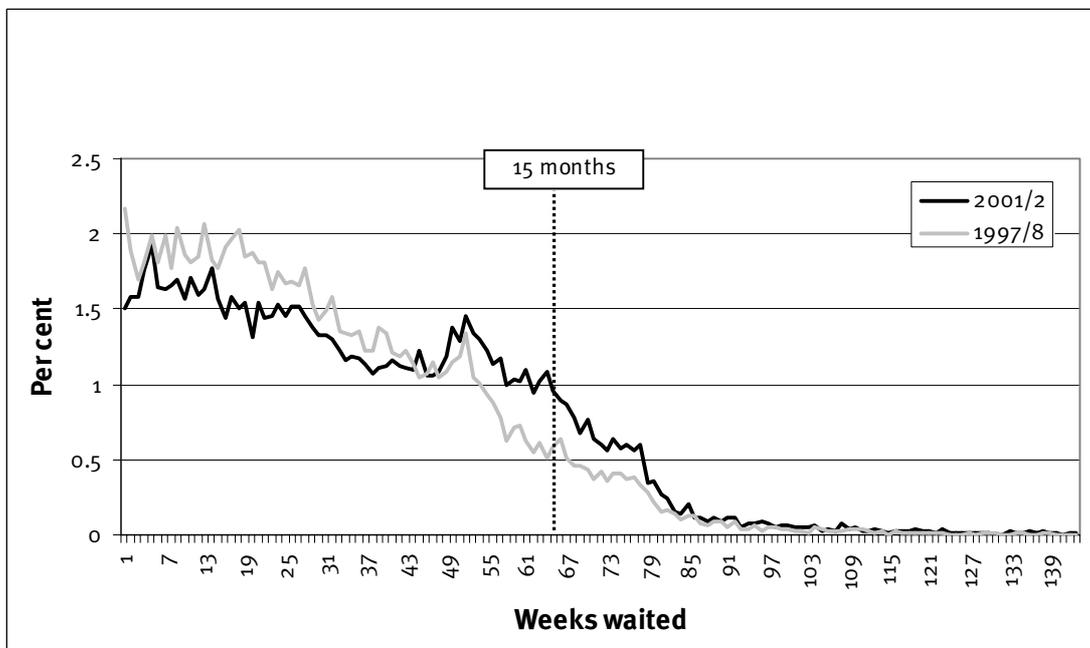
2 Subtotal excess also based on calculation in first footnote. It is not the sum of the ten HRGs' excess admissions.

Figure 34: Knee replacement: 1997/98 to 2001/02 waiting times distributions, admissions as a percentage of all admissions



King's Fund (2003)

Figure 35: Hip replacement: 1997/98 and 2001/02 waiting times distributions, admissions as a percentage of all admissions



King's Fund (2003)

Summary

The foregoing analysis suggests that trusts and their orthopaedic consultants responded to the 2001/02 15-month maximum waiting time target in part by altering the proportion of their waiting list admissions coming from different wait-time categories. Compared with an estimated expected distribution of admissions of patients waiting between 40 and 80 weeks, admissions in 2001/02 were around 38,000 higher than expected – an average addition in admissions of 42 per cent over that period.

These additional admissions peaked not only for patients waiting 15 months, but also for those waiting 12 and 18 months. Similar peaks at 12 and 18 months were found in an earlier (1997/98) admissions distribution, which perhaps suggest that a longstanding factor in clinicians' admission criteria is the time for which patients have waited, with 12 and 18 months representing important milestones or limits for some patients.

The composition of these extra admissions was found to differ compared with those who had not waited between 40 and 80 weeks prior to admission. Twelve procedures accounted for 77 per cent of 'excess' admissions, with arthroscopies and knee and hip replacements accounting for nearly half of these. These same procedures, plus soft tissue and other bone procedures (Category 1, for those aged under 70 without complications) alone increased their share of admissions in patients waiting 40–80 weeks. It does not seem to be the case, therefore, that the apparent extra admissions of patients waiting between 40 and 80 weeks were long-wait minor cases.

Examination of the timing during the year when short- and long-wait patients were admitted revealed that the proportion of weekly admissions of long-wait patients (both 40–80 weeks and 56–65 weeks – in other words, around 15 months) rose from around October towards the end of the financial year, but that the proportions of patients waiting between one and four weeks did not change in any responsive way. This analysis also revealed that holidays – particularly Christmas week – produced very great distortions in admissions affecting all types patients, no matter how long they eventually waited for admission.

Further to this part of the analysis, a 'deadline effect' was notable, with 'excess' admissions of patients waiting around 15 months peaking in the final quarter of the year as the target deadline loomed. While this may have contributed to an increased likelihood, and feelings on the part of consultants, that their clinical priorities were being distorted in the weeks before March 31, subsequent changes to the monitoring of the target may have dealt with this problem.

While peaks in admissions for 2001/02 were found at 15 months, there were also peaks at 12 and, to a lesser extent, at 18 months. Comparison of waiting times distributions for 1997/98 revealed very similar peaks at these latter two dates, but not at 15 months – confirming the likelihood that it was the 15-month target in 2001/02 that was responsible for this change in admissions.

This 'before-and-after' comparison of waiting times distributions also provided an estimate of the number of excess admissions of patients waiting around 15 months prior to admission of 9,333 (2.2 per cent of all admissions) – probably a more accurate estimate of the impact of the 2001/02 target than previously calculated, using statistical estimates of 2001/02 waiting times distributions, as the 12-month peak in admissions cancelled out in the comparison.

Whether these apparent 'excess' admissions were at the expense of other patients in particular (more urgent) cases is very difficult to answer with aggregate national data. However, examination of the composition of these admissions did not immediately suggest either that more minor cases were substituted for more major cases (in fact, it appears the opposite was true, with hip and knee replacements dominating these admissions) or that very short-wait (and hence most urgent) cases gave way to longer-wait (presumably less urgent) ones. However, for hip and knee replacements – the two procedures accounting for a majority of 'excess' admissions of patients waiting around 15 months – while admissions increased for virtually all categories of patients, proportionally fewer patients waiting between 1 and 44 weeks were admitted in 2001/02 than 1997/98.

However, whether this can be taken as proof of a substitution between long- and short-wait patients is difficult to answer. By definition, if the proportion of admissions of one group of patients increases, the proportion of another group must fall. But, as already noted, actual numbers of admissions of virtually all groups of patients rose between 1997/98 and 2001/02, which complicates any interpretation of the change in proportions of patients admitted.

While it was found that the number of patients who had waited one week prior to admission fell when 2001/02 is compared with 1997/98, it seems very unlikely that this difference was due to the 'excess' 15-month wait patients usurping their position on the waiting list. On balance, there are likely to be other explanations for this difference.

The box below summarises the answers to the five questions posed earlier (see p 8).

Summary of answers to research questions

1. Can the apparent additional admissions for those waiting around 12 and 15 months be quantified?

A best estimate, based on the difference in proportions of patients waiting prior to admission in 1997/98 and 2001/02, suggests that there were around 9,333 'excess' admissions of patients waiting around 15 months. This is equivalent to 2.2 per cent of all orthopaedic admissions in 2001/02.

2. Is it possible to determine whether additional admissions displaced other types of patients, either in terms of the types of procedures carried out or in terms of the time patients waited?

The estimated excess admissions were different in terms of the proportions of the types of procedures carried out compared with patients who had not waited between 53 and 78 weeks, with more major cases being admitted (over 56 per cent of the additional admissions consisted of hip and knee replacement procedures).

While we could not unambiguously establish whether or not additional admissions had lead to delayed treatment for other patients, there was no evidence that very short-wait patients suffered.

3. Given that the 2001/02 maximum waiting times target deadline was March 2002, is there a 'deadline effect', with additional patients tending to be admitted in the fourth quarter of the year?

There is a 'deadline effect', with proportionately more long wait patients being admitted towards the second half of the year. There was no discernible trade off however with the proportions of short wait patients admitted.

Holidays had a large effect on admissions, however, with Christmas week and the subsequent four weeks affecting the proportion of admissions of people who waited 1–4 weeks most.

4. As the 2001/02 target was for no patients to be waiting more than 15 months, why is there also an apparent increase in admissions for patients waiting around 12 months (and, to a lesser degree, 18 months)?

The peaks at 12 and 18 months were also present in the 1997/98 waiting times distribution, indicating, perhaps, that consultants' admission criteria are in part driven by particular waiting time milestones.

5. How does the 2001/02 waiting times distribution compare with a period before the implementation of the 15-month target?

The two distributions were very similar, except for greater admissions in 2001/02 of patients waiting around 15 months. There was also a noticeable reduction in those waiting one week. However, it is not believed that the former increase is related to the latter fall.

Conclusions

As the survey carried out for this research indicated, there is no doubt that there is a genuine feeling among a significant minority of clinicians, particularly from trusts with relatively poor records of achievement on reducing waiting times, that attempts to meet maximum waiting times targets can clash with their own clinical judgements concerning when to admit patients from waiting lists.

However, as we stated in the background to this research, even though for an individual clinician a particular change in the order in which patients are admitted from a waiting list (arising from attempts to meet waiting times targets) may run counter to their own clinical judgement, variations in clinical judgement means that the true comparison at the heart of this issue should not be between on the one hand, a clinical 'gold standard' for admission and on the other, a version of this that is amended by managerial/political priorities, but between the latter and the actual waiting times distribution arising from variations in clinical opinion concerning admission priorities.

However, clinical opinion has historically given rise to considerable variations in waiting times experienced by patients from one part of the country to another (indeed, within the same specialty within the same hospital). There are, of course, other factors that can explain such variations, but only in part. As our small survey confirms, even in a hypothetical situation clinical opinion varies considerably and, it could be argued, gives rise to a greater differences in the order in which patients are treated than we could attribute to attempts to meet recent waiting times targets.

However, quantifying the impact of the 15-month waiting times target in the absence of an admission-criteria 'gold standard' is difficult. Of two approaches used, a 'before-and-after' comparison of the waiting times distributions for trauma and orthopaedics for 2001/02 and 1997/98 probably provided a more accurate picture of the impact on admissions from the waiting list.

From this analysis, it is clear that there was a definite response to the 15-month maximum waiting time target for 2001/02, manifesting itself in the distribution of admissions (by waiting time), with an increase in admissions of around 9,333 (2.2 per cent of all orthopaedic admissions in England in 2001/02) for patients waiting around 15 months, an increase in average waiting times overall of 1.5 weeks, and a decrease in the spread of the distribution of admissions. As the survey showed, in respect of poorer trust performers in particular, the increased activity involved a pattern and intensity of working with which many may have felt uncomfortable. However, now that waiting times targets are being assessed on a continuous basis, this should be less of an issue.

It has not been possible using national data to show unambiguously that the admission peak at around 15 months does represent clinically relevant distortions. However, they do provide grounds for suggesting that these were not of major importance as follows:

- Urgent cases do not appear to have been displaced – holidays had a greater impact. If cases have been deferred to meet targets, they are likely, therefore, to have been less urgent ones (for example, from our analysis, those waiting between 33 and 52 weeks) and the scale of their deferral, relative to their average wait, modest.
- The extra cases largely consisted of significant (hip and knee replacements, for example) rather than minor operations, and hence no evidence was found of substitution of lesser for more serious cases.
- The form of target prevailing during 2001/02 meant that for most of the year the risk of distortion was absent.
- Moreover, although the 15-month peak in admissions appears to be a new phenomenon, in other respects the distribution of waiting times does not appear to have changed greatly since the introduction of the 15-month target.
- Our overall conclusion, therefore, is that serious and extensive clinical distortions are likely to have been fairly limited. However, we recognise that the use of national aggregate statistics may understate the problems at local level.

3. A framework for system-based information requirements for managing the supply of elective care

In the first stage of this research (see Section 1 and Understanding whole systems, p 31), we suggested that the effective management of the provision of elective care required an understanding of the whole system of elective care as part of the whole hospital system, and the hospital system as part of the local health economy. This requires an analytic approach (and implies the collection and use of information) that was not apparent in its entirety in any of the hospitals we have looked at – even those whose performance was relatively good. In this section, we report on our research into a system-based framework of the information required to manage the supply of elective care and, *inter alia*, waiting lists.

Background

The section Key elements for managing the elective care system (p 42) suggests key tasks for managing the elective care system. Most of these related to the collection, analysis and use of information in broad terms. Thus we suggested that the hospital – in planning services and monitoring performance – should carry out the following activities, outlined in that section, all of which relate in one way or another to the availability and use of information.

In our study, we did not find any hospital that had developed its information systems in a way that allowed all of these tasks to be accomplished, although there were hospitals where some of these tasks had become part of routine hospital management. In this paper, we propose a system-based view of the provision of elective care to identify what information a hospital requires in order to manage its delivery of elective care, and hence waiting lists, effectively.

Figure 36 illustrates the relationship between information and the management of services in the elective care system. The central tenet of our approach involves the identification of the nature of the system within which the hospital's elective care services operate. This then indicates the information that will be required in order to manage elective care within the broader system. This information will also feedback into the analysis of how the system works.

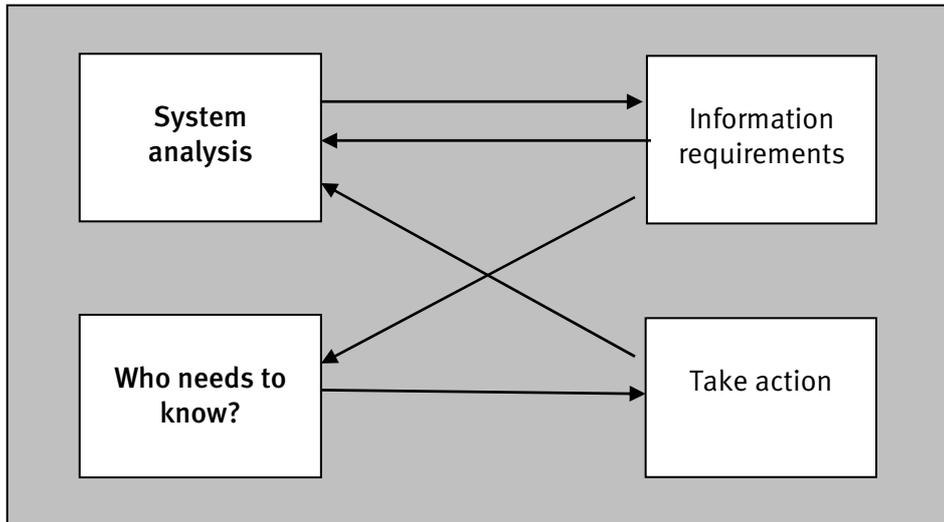
The information will be used by key people in the hospital – both clinicians and managers – for two purposes:

- to manage the delivery of services effectively
- to plan the pattern of services that will meet the expected structure of demand over the effective planning period (which might be a year or more).

Clinicians and managers have an array of actions available to them in order to maintain the effective delivery of elective care services in the hospital. The impact of these actions will feed back into the analysis of the system, and hence may result in changes to the information requirements of the system, and so on.

In a paper of this nature, however, it is not feasible to attempt to set out all the information that anyone attempting to understand and manage the elective care system could potentially require. We are aware that, for example, the Modernisation Agency (2002c) has published detailed ‘how to do it’ manuals, which contain very detailed specific recommendations.

Figure 36: Framework for system-based information requirements



Our aim, therefore, is limited to defining a framework within which any systematic data collection and analysis should be carried out. The systems viewpoint adopted here emphasises linkages – in other words, various forms of connection between the elements which comprise it. This viewpoint leads in the direction of attempting to take all possible influences on board when trying to improve the performance of any one element of the system, such as outpatient clinics for one specialty.

But, as is shown below, understanding even one such clinic is highly demanding in information terms. In practical terms, therefore, there is a need to break the whole system down into manageable parts while at the same time acknowledging the connections between them.

Our starting point is the definition of the elective care system. ‘The policy environment’ (p 82) considers how the elective care system relates to the other systems which influence it, and of which it forms part.

The sections ‘Understanding demand’ (p 85), ‘The production of care: outpatient services’ (88), ‘The demand for elective operations’ (p 89), and ‘The production of care: elective operations’ (p 91) look at the elective care system in more detail using a demand-and-supply framework.

Key elements for managing the elective care system

- Produce a menu of services for each specialty, with estimates of time taken per service, and the level of expertise required. Using this menu, the hospital should produce a detailed profile of demand for services (operations and consultations), on a daily basis, showing levels of variability in numbers referred, in the difficulty of cases, and in the time taken per service.
- Produce detailed output schedules for the current configuration of services indicating how many slots are available, how much time is allocated per slot, and the nature of the output produced.
- Produce an analysis of the potential constraints on output other than availability of direct-contact staff, such as operating theatre staff, and theatre slots. This would include:
 - all testing and investigation issues, both pre- and post-operation or consultation
 - care availability within the hospital setting, which would be primarily staffed beds and requires an understanding of the needs of individual patients in terms of the number of bed-days required
 - care availability outside of the hospital
 - availability of other professionals for activities such as rehabilitation, such as physiotherapists, occupational therapists and dieticians.
- Produce an analysis of patients who do not attend for one reason or another, or are excluded from treatment at points in time (suspensions or cancellations), so that the impact on 'real' waits and on the efficient use of resources can be derived.
- Produce a transparent costing of all of the various options proposed for producing services, based on detailed investigation of the variable inputs, plus an appropriate allocation of fixed costs.
- On an annual basis, the hospital should produce a profile of demand and supply for each individual service, and how this would be met throughout the year, together with a range of variability and how this would be dealt with.
- Negotiate the option to use physical capacity, such as theatres out-of-hours, so that when there are unpredicted surges in demand that would cause patients to wait longer than planned, these options can be taken up. Options such as the use of private facilities could be considered in the same way.
- Introduce a system of 'notional' booking for all patients in queues, so that when a patient joins a queue there is a notional slot allocated (but not necessarily given to that patient). This should be applied to all outpatient slots, inpatient and day case treatments, and tests and investigations.
- Ensure the progress of patients through the elective system is traced by using a unique patient identifier. This would make it possible to produce estimates of total average waits in different parts of the elective system within the hospital.
- Manage the interface between emergency and elective care by forecasting the likely profile of emergency demand, assessing the scope for reducing bed and nursing requirements for elective care by, for example, scheduling operations involving short hospital stays in times of likely peak demand, and working out the most effective and efficient balance between ringfenced and general pool beds.
- Improve the referral process with the aim of reducing unnecessary referrals and clinic visits through, for instance, the use of protocols, and other forms of closer working between primary and secondary care.

- Assess, by reference to suitable comparators or benchmarking, the scope for improving productivity of staff and facilities – for example, the number of operations per full time consultant or per theatre session.

The system perspective

We begin by stating why we think a system perspective is required for planning elective care activity. We are aware that much advice has been issued on elective care (Audit Commission 2003, Department of Health 2002). The vast majority of this has focussed on a particular pathway (for example, access to cancer services) or on a facility such as an operating theatre. On a day-to-day basis this is entirely appropriate, and even at this level – as we will argue later on – a large amount of detailed information is required to plan and operate effectively.

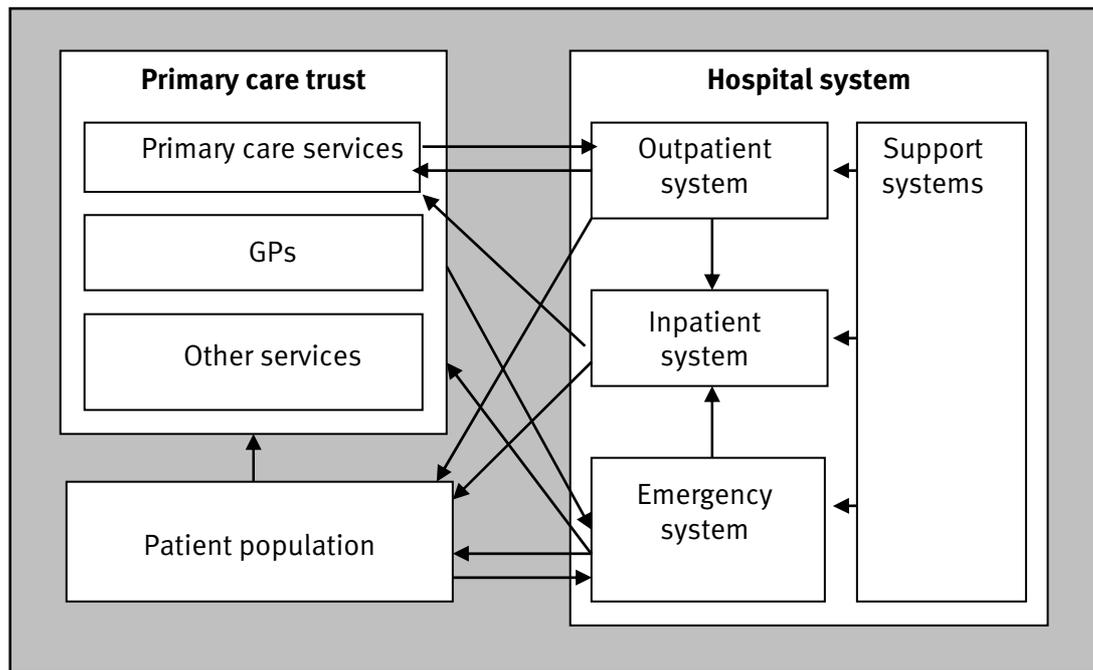
However, to plan purely in this way risks missing out some of the critical factors bearing on the way that the elective care system works. Put more generally, it is widely acknowledged that within the total health care system, including tertiary, secondary and primary care, everything connects with everything else, but that it is a practical impossibility to allow for all possible interconnections, all the time. For practical purposes, therefore, the overall system has to be broken down into parts but the interconnections remain.

For the purposes of this paper, we define a system as a set of inter-related actors (individuals such as consultants or GPs, organisations such as PCTs and so on) who influence each other in direct and easily recognisable ways, and also in indirect ways that may not be readily apparent. In some cases, there is feedback, where the actions of one actor influence others, which ‘rebound’ and influence those actions in a later time period (see Figure 37).

Setting the boundaries of a system is a matter of judgement, and depends to some degree on the question being studied. In what follows, we use three interrelated systems:

- the whole local system (or local health economy), which includes other health care providers, as well as the source of demand for the hospital’s services
- the hospital system as a whole, and all the activities that this involves
- the various specialties or even individual consultants.

Figure 37: A local health care system



The elective care system forms part of all three systems. It is treated as a system in its own right because, although it is often set out as a simple sequence (see Figure 38), in fact there can be significant interactions between its various elements. For example, the number of referrals may be influenced by the number of people waiting further down the sequence and/or the time they have waited. In Figure 39, these interactions are shown in the feedback loops on the right and left sides of the diagram. We will consider them in more specific terms later on.

Figure 38: Elective care pathway

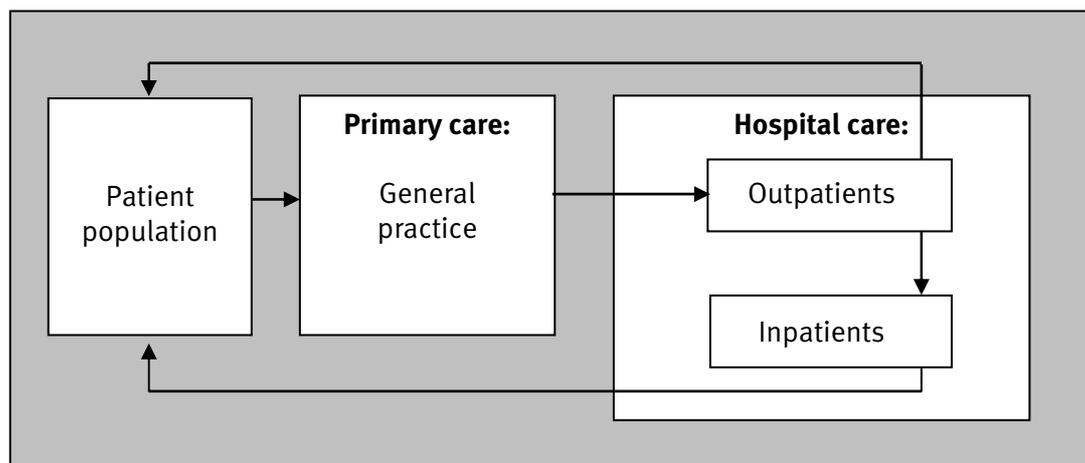
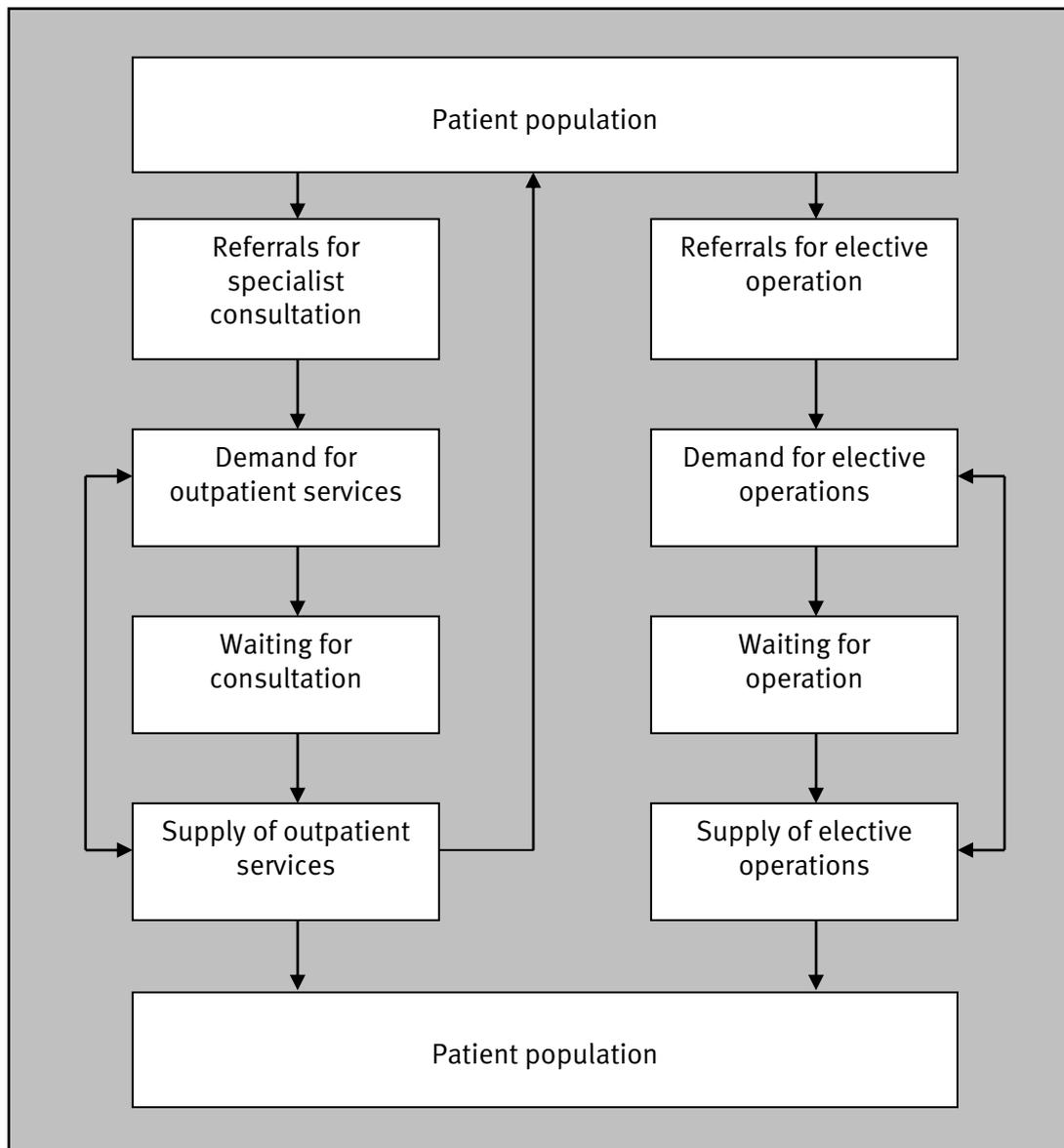


Figure 39: The elective care system

The three systems considered here are embedded within the national policy environment. While in principle everything may influence everything else, in practice it makes sense to assume that the individual hospital is affected by, but does not affect, the policy environment in which it operates. We briefly consider this environment first, and then go on to take each of the three levels in turn.

Policy environment

The policy environment in which the elective care system operates is undergoing rapid change. The Government's long-term goal of improving access to elective care remains, but the introduction of payment by results, patient choice and patient empowerment through better information, the allocation of the bulk of the health budget to PCTs, the implementation of the working time directive and many other policies influence how

the three systems we go on to consider operate. Perhaps most significant of all is the creation of a market in elective care, which means that the individual hospital's elective care system is part of a wider system.

These developments make the hospital's planning task more complex than it has been in the past, and increase the need for it to understand more fully the demand it faces and the costs of meeting that demand.

Level 1: The health economy

The hospital is part of the local health economy, which consists primarily of other hospital providers, community health service providers, GPs and other non-NHS provision. Figure 39 (p 82) shows demand for the hospital's outpatient services arising from referrals from the pool of potential patients. By simple analysis of its own historic referrals and changes in the local demographic structure, the hospital can gain some understanding of the demand it is likely to face if existing trends continue.

The system perspective implies the need to look beyond existing trends, however, to the factors determining those trends, and the extent to which they are affected by what the hospital itself does (or needs to do in reaction to other exogenous constraints – for example, meeting waiting time targets). The former will include the health status of the population, which can be captured to some extent by the local demography. But other important factors include changes in medical technology, changes in patient expectations, and the various policy changes – some of which are directly concerned with waiting times, such as the Patient Choice project, and some of which are not, such as developments in out-of-hours care.

Moreover, the hospital should be aware of the position of other (public and private sector) acute providers within its local economy who may be direct competitors, as well as the position of local GPs whose actions may increase or decrease the level of demand the hospital faces for a given potential pool of referrals – for example, if GPs opt not to provide out-of-hours services, or through the activities of GPs with a special interest (GPSIs). Moreover, there may also be the possibility that the hospital can redirect demand from its services, particularly if it is able to have good collaborative relationships with other local providers, or that other providers – including GPs – will take over part of the workload.

Thus the key system issues at this level are:

- the likely developments in the supply of elective care services outside the hospital
- how these developments may be influenced by the hospital's own decisions to expand or reduce capacity, in the light of its financial position
- how the hospital's own decisions affect demand for its services
- how GP referral decisions affect the demand for elective surgery.

Each of these issues entails more than just the collection of simple information. The hospital will need to carry out some form of scenario planning, or may even need to produce a formal market-demand analysis. Retail businesses in the private sector spend considerable sums on analysing likely demand for existing or new products. As an example, while the fall in demand of heart operations that has occurred in recent years due to improved preventative measures came as a surprise to many health service managers, such 'market' analysis would have provided information and insights into this change in demand.

Level 2: The hospital

A hospital can be seen as a cluster of some 40 or more specialities and a large range of support services most of which (such as diagnostics) they share. Within each is the major division between emergency or 'hot' work and planned or 'cold' work. Even at this very broad level, there are many interconnections between the two broad functions, on the demand and the supply side alike.

The hospital must aim to understand the position that the provision of elective care occupies within its whole business (unless it just provides elective care). To do this, it must map out its technical production processes, both for outpatient consultations and for elective operations, the nature of the flow of demand that results within its own production system (outpatient consultations lead to elective operations, for example), and the interaction of these with other functions of the hospital – primarily emergency care, but also planned operations.

Recently, with the introduction of treatment centres and day case units, the general trend has been to simplify the hospital system by dividing it into two – in some cases, by ringfencing resources within a hospital, while in others, by dividing these functions between two hospitals.

Nevertheless, at the planning level, the hospital makes decisions that determine the use of staff, theatre capacity and care facilities (beds, nursing staff and so on) between elective and emergency care. This may also involve decisions as to their allocation over the course of the year (for example, less bed and theatre use for elective care in winter). Moreover, decisions are also made at this level regarding the allocation of elective resources between different specialties.

The key system issues at this level bear on the links between the various activities of the hospital, including each specialty as a separate system. They include:

- the extent to which it is feasible and economic to separate the elective care system from the rest of hospital activity
- the best means of providing for variations in demand
- in areas where elective work is not isolated, the nature and scale of linkages, such as joint staff, joint facilities and cross referrals
- how out-of-hours arrangements affect demand for emergency.

The first and second of these are essentially planning issues. As we have noted, the trend is for separation, but we are not aware of any study demonstrating the overall impact of doing so (for example, the extent to which research, training and emergency functions are hindered or made more expensive by separation although concern has been expressed by the medical profession particularly about the training and associated quality issues when some of the more straightforward elective procedures are removed entirely from the hospital).

Where elective care remains closely linked to emergency care, the emergency and elective systems need to be planned jointly, for example (as we found in one of our study sites) by scheduling elective work in anticipation of variations in emergency demand for theatre and other resources.

The third of these is in part a planning but also a monitoring issue. For the purposes of planning, assumptions have to be made about all the areas where the individual speciality is linked to others, and to the hospital as a whole. Information flows are

required to check that these assumptions hold good on a continuing basis. In our earlier work, we found that hospital management had no systematic knowledge of these linkages and was, in general, unaware in a systematic way of the changes taking place – for example, cross-referrals within the hospital, which can mean that demand for consultations rises faster than earlier trends suggest (because, in effect, one patient has two initial consultations).

Level 3: The specialty

Each specialty is embedded in the hospital system but also interacts with the wider health economy. The analyses that the hospital uses for overall planning purposes, (for example, to plan the division between emergency and elective care) should be based on these individual specialty-level analyses. However, we would expect monitoring and management to take place at the specialty level – or even at a more disaggregated level if consultants do not pool lists (although it will have to be consistent with the overall plans for the use of hospital resources).

On the basis of a number of studies in the UK and abroad, previous King's Fund reports (Hamblin, Boyle and Harrison 1998, Harrison and New 2000) have argued that changes in the length of time people wait will influence the numbers of people being referred for treatment – that is, that there is a feedback effect. This may arise for a number of reasons: where waits are long, GPs and patients may prefer to try self-management options, they might use the private sector (including complementary practitioners), or they may simply decide to live with their condition. As waits fall, the balance of advantage shifts, and they may decide to join the (shorter) queue. Similarly, consultants may modify their decisions in the light of the performance of their own specialty – that is, their decisions to admit, or deferring a decision to admit, may be influenced by the performance of the specialty in meeting whatever targets it is attempting to meet.

In the new policy environment, the scale of this effect is likely to grow. For example, in our first report we noted that one hospital had experienced sudden increases in outpatient referrals, which appeared to be a response to its improved performance. But also, as patient choice becomes a reality, such responses are likely to become more important as PCTs purchase on the basis of 'quality' (a key, if not overriding, aspect of which will be waiting times).

The system issues that arise, therefore, include:

- the way in which referrals will respond to reductions in waiting times for outpatient consultations and for treatment
- how decisions to treat (that is, the conversion ratio) respond to changes in waiting times.

Once a view is taken of the likely level and trend in demand for the current planning period, then detailed planning can begin. In the following sections, we consider what information a hospital might require to plan its elective care. This discussion is intended to be illustrative of the level of detail required to plan, manage and monitor the processes making up the elective care system. It is far from being exhaustive but is sufficient to demonstrate that, even at this level, information demands can be extensive, and still more so once variation and uncertainty are allowed for.

Understanding demand

There are two broad categories of elective care services provided by the hospital: outpatient consultations and elective operations. Demand for the former arises from referrals to a specialist – which gives rise to an outpatient appointment. Demand for the latter results from decisions by specialists, in consultation with the patient, usually during one or more outpatient consultations. In what follows we ignore planned procedures (although would note that in proportionate terms, these have been rising rapidly in recent years and hence must be taken into account when estimating future capacity requirements).

Total demand for these services can be broken down further: by specialty, by consultant, by type of operative procedure, or by whether an operation is performed as a day case or as an inpatient episode. Ultimately, the hospital should be able to monitor demand at each of these levels – although whether by groupings of specialties, operations or consultant teams is not obvious.

It is only by having a clear view of the nature of the demand for its services that the hospital will be able to manage and adjust its capacity successfully to meet the variations in this demand that will occur. Moreover, the hospital will also be in a better position to take measures that will influence the level and nature of the demand for its services (for example, through the use of referral guidelines or online/telephone consultation procedures) and also, crucially, the meeting of desirable goals, such as reducing and sustaining reductions in its waiting times.

The system issues at this stage focus on:

- the impact on the demand a hospital experiences of its own performance in dealing with that demand
- interactions along the care pathway – for example, how changes in performance at one stage impact on the workload at other stages.

Outpatient services

In this section, we focus on demand for outpatient services. Demand for elective operations will be discussed in ‘The demand for elective operations’ (p 89) as an output from the outpatient consultation.

We distinguish two types of demand for outpatient services: the new consultation and the follow-up consultation. We assume demand for a new outpatient appointment comes from two main sources:

- GP referral (it is only this that is included in the Government’s target. This is clearly inappropriate and will give rise to distortionary effects)
- another consultant (who could be internal or external to the hospital).

Demand for a follow-up appointment results from a decision arising in a previous consultation between the clinician and the patient. Elective and emergency surgery also often leads to a demand for a follow-up appointment.

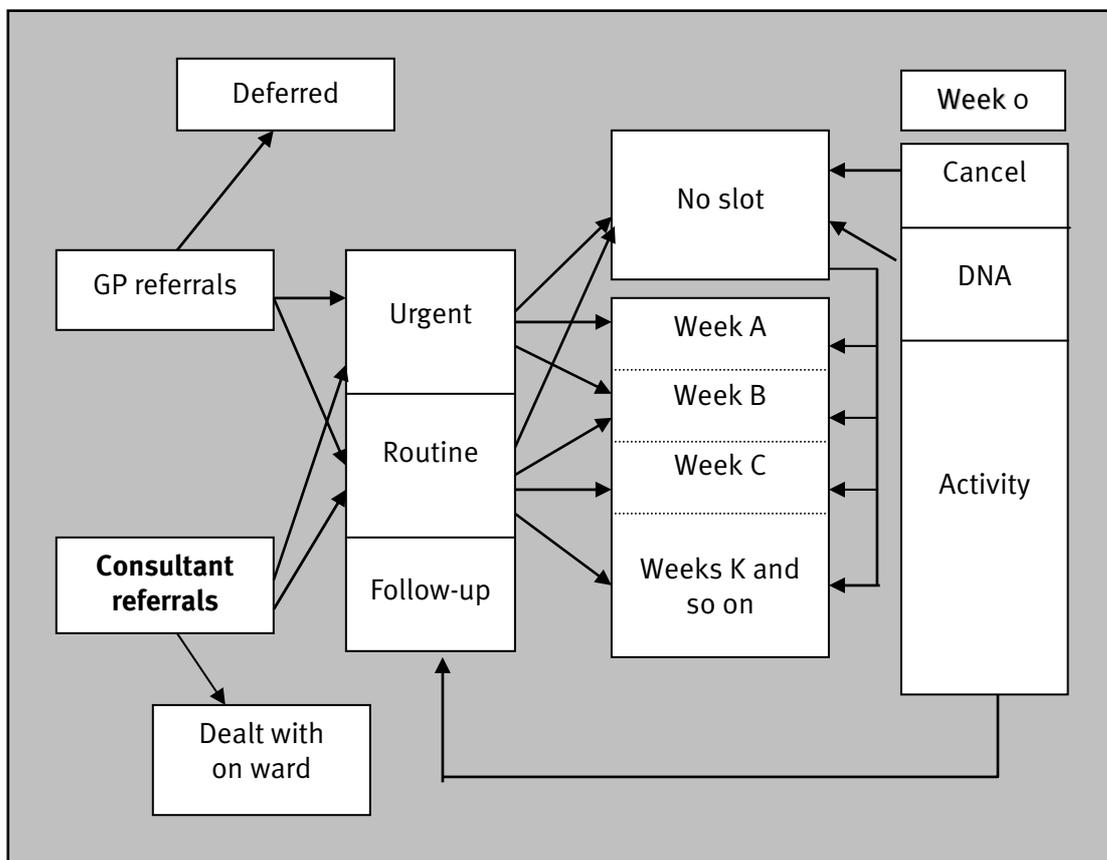
New weekly demand

GPs can refer to a specialty in general or to a particular consultant (specialist). It is possible that a proportion of referrals are rejected (for various reasons). Figure 40

illustrates how demand for outpatient services arises. Each week, consultants receive requests from GPs for new outpatient appointments, and these are allocated new appointment slots according to some priority system (which includes ‘first come, first served’). For example, referrals may be categorised as urgent and routine, and the hospital attempts to give urgent referrals an appointment within some target time (for example, referrals of patients with suspected cancer). On the production side, some slots may be reserved for urgent referrals.

Referrals may be allocated according to a scale of the amount of time they are likely to take. This could be very detailed or simply one figure for all. Consultation times vary considerably. Other referrals are usually from other consultants, either within the same hospital or from another hospital. In some cases, internal referrals will be dealt with on the wards rather than in outpatient clinics, and hence do not impinge on clinic capacity.

Figure 40: The demand for outpatient care



Follow-up appointments are usually decided by the clinician in discussion with the patient at the time of a consultation. The referral will tend to be allocated a slot at some future time, according to the agreement between the clinician and the patient. A proportion of follow-up appointments will result from referrals following elective or emergency surgery.

Existing demand

At any point in time, there will be an existing set of patients waiting to see a specialist, with slots already allocated. Some hospitals operate a system where there is a delay before slots are allocated to patients. The hospital can identify how long this delay is.

The existing queue will be allocated to particular days and clinicians, and will have an expected consultation time associated with each prospective consultation (where this is not explicitly done we can assume that some average consultation time is associated with each booked slot).

Each week there are changes to existing queues as:

- some request a change in appointment
- a slot is lost for some reason and therefore a new slot has to be offered
- a consultation takes place
- the patient does not attend.

It is important to be aware of the flow of people into outpatient services on a weekly basis. The information that this requires is crucial to managing the corresponding supply of services. At its simplest, the information required is:

- the number of:
 - GP referrals
 - consultant referrals
 - GP referrals accepted
 - consultant referrals dealt with on the ward (or in any alternative manner)
 - follow-up appointments
 - consultations that take place;
 - DNAs
- the number waiting:
 - for an outpatient appointment with a slot
 - for an outpatient appointment without a slot
 - who cancel an appointment
 - who request different appointment
 - who lose a slot (where the hospital cancels)
- the proportion of urgent referrals from GPs
- the average consultation time required (by type of consultation)
- the average time taken by each consultation (by type of consultation).

The production of care: outpatient services

We now define the potential productive capacity of the hospital in terms of key types of service provided, and explain how demand is met by these services. This, in turn, suggests the nature of the information that must be collected about the production of care.

First, we define two types of output – an outpatient consultation and an elective operation. Each of these can be sub-divided further, as we shall see. In this section, we look at outpatient services. These consist of consultations between a patient and a member of the specialist team. This does not have to be the consultant leading the team. However, in defining capacity there is a need to be clear about who provides the service as this may impact on a number of variables, including the type of patient seen, time

taken, or outcome. In this report, we refer to this input as a ‘clinician’ without identifying what type of clinician is involved.

A fundamental requirement is to be clear about the level of capacity of service provision, and also what possibilities there are in the short and medium term for changing this capacity – hence introducing the flexibility to deal with unanticipated events.

Determining capacity

The input requirements for the provision of outpatient services consist of:

- the clinical input (labour)
- the consulting room (a physical facility where the consultation takes place)
- (possibly) equipment for the immediate provision of diagnostic testing such as x-rays – over and above the usual equipment that a clinician routinely requires.

Appendix 6 describes in more detail the outpatient production process, and, illustratively, sets out the input resources – such as clinicians, clinics, and slots – which need to be considered in order to determine the level of outpatient capacity.

However, as the discussion in Appendix 6 shows, once the process is looked at in more detail, the amount of information required is extensive. In particular, the degree of variation in all of the factors listed needs to be monitored. At its simplest, information requirements for this part of the system will be:

- the number of clinics, consultants, consulting rooms, and slots per clinic
- the average time per slot
- any additional information on distinctions between clinics.

The demand for elective operations

The demand for elective operations arises from decisions made by the clinician during a consultation with the patient. One outcome of the outpatient attendance is a decision by the clinician that the patient requires an elective procedure. If the patient agrees to this, then the consultant will ensure that the patient is placed in a queue for elective care.

In some cases, this means that the patient is immediately allocated a slot for an operation. If this slot suits a patient, then they are placed on the waiting list as a ‘booked’ waiting list patient. If no slot is allocated immediately, then the patient is waiting but with no date for an operation. Eventually, the patient will be offered a date for an operation, and if this is suitable, will be booked in for an operation.

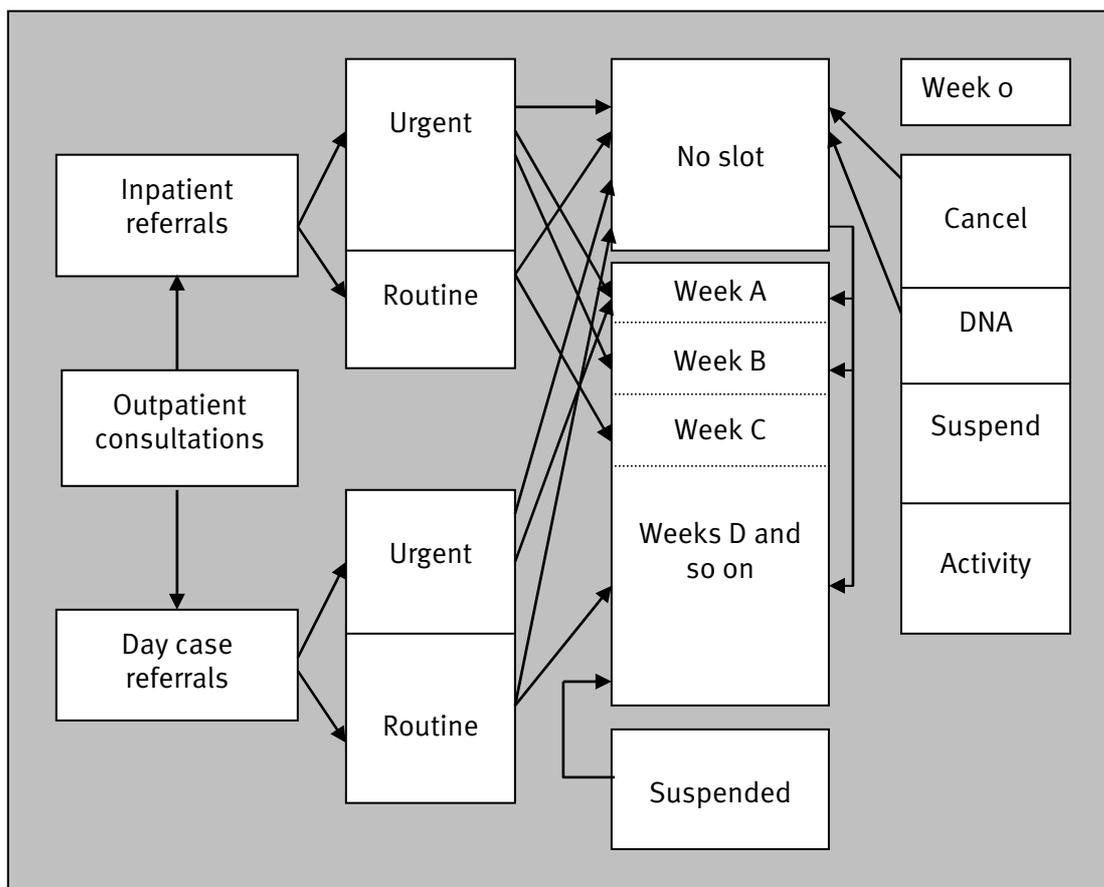
New weekly demand

The clinician will place the patient on a waiting list for a particular procedure, or procedures. Almost always, the operation will be performed by that clinician or if not, by another member of the team.

Each week, patients will be added to the queue for the clinician’s services. These patients will be allocated a place in the queue according to some priority system. In some cases, they will be given a date immediately, while in others they will simply be waiting to receive the offer of a date. Figure 41 illustrates the relationship between the demand for and the supply of elective operations.

There is a key distinction between elective care where the patient is seen as a day case and as care provided on an inpatient basis. The former does not require an overnight stay and hence does not require the use of a hospital bed for the night (and the other resources that this entails). The patient may use a bed for recovery before going home. Inpatient care requires an overnight stay. The average length of this stay will vary according to the operation carried out and other factors – particularly the overall condition of the patient. This is important, as availability of beds will be a factor in determining the availability of inpatient elective services.

Figure 41: The demand for elective care



In passing, we note that once patients are in the queue, they may still be offered outpatient consultations while waiting. This allows assessment of their condition, and may form an important part of the judgement concerning when they should be dealt with – although this will become less important as the overall time waiting for an operation decreases.

Existing demand

At any point in time, there will be an existing set of patients waiting for an operation. Some of these will have slots (dates) allocated, while others will not. In most cases, there will be some priority associated with each patient.

Clinicians will have their own lists of patients corresponding to their own theatre slots over the planning period (the next year, for example). There will be an expected length of

operation for each patient. This enables the clinician to plan services effectively. It also enables the slot to be filled with a suitable replacement if an existing patient is removed from the list.

Each week there may be changes to existing queues. These can result from:

- people deciding not to wait any longer and informing the hospital. (This might include the patient dies, or gets better, or leaves the area or country)
- people requesting a change in slot (where this has been given)
- an operating slot is lost for some reason
- the patient being suspended from the waiting list, usually because they are no longer available for an operation, or because it is decided that it would not be appropriate to operate at that time
- an operation takes place
- an operation does not occur, or is abandoned for some reason.

To monitor demands placed upon the elective care system requires at a minimum the following data, on a weekly basis:

- the number of consultant referrals for an operation:
 - as an inpatient
 - as a day case
- the proportion of urgent referrals as:
 - inpatients
 - day cases
 - inpatients that are given a date
 - day cases that are given a date
- the proportion of non-urgent referrals as:
 - inpatients that are given a date
 - day cases that are given a date
- the average time required for an operation as an inpatient, by category of operation
- the average number of days in hospital required for an inpatient operation
- the number waiting for an operation who:
 - cancel
 - request different appointment
 - lose a slot (where the hospital cancels)
- the number of operations that:
 - take place
 - fail to occur for some reason on the day
- the number of people waiting for an operation who are suspended
- the number of DNAs that occur
- the average time of an operation.

The production of care: elective operations

Elective operations include those delivered as day cases and those where an overnight stay is required. However, these are probably best discussed separately as there are significant differences in the way in which they are delivered.

The service provided is an operation on a patient which usually comprises a single operative procedure but may include more. Often the operation is performed by a consultant, but it can be by a member of the consultant's team, with or without the consultant in attendance.

A fundamental requirement is to be clear about the level of capacity of service provision that is potentially available. Where there are so many different operations possible, taking various lengths of time, and with varying outcomes, how should capacity be measured?

Determining capacity

The capacity of an elective care service depends on the availability of operating theatres and of clinical and other staff necessary for the performance of the operation.

Either staff or theatre space could be a constraint on delivery. However, assuming that the total number of theatres available provides a maximum on the number of theatre hours available, it is only then that the availability of clinical staff becomes a constraint on the number of theatre slots used.

In this case, it is not just the consultant team that is a constraint: the availability of anaesthetists, appropriate nursing staff, porters or specialist equipment (or, indeed, the patient themselves) may prevent an operation proceeding.

For day cases, these are the only inputs that need to be considered. However, for inpatient care, a bed for an overnight stay is also required (with appropriate other resources such as nursing staff). Lack of bed resource may be an effective constraint on the provision of elective services. This can be of particular importance if (or when) beds are increasingly occupied by patients admitted as emergencies whose flow into the hospital and length of stay is less certain than that of elective patients.

Decisions will have been made – at the planning level – about the number of theatre sessions that a particular team has in a week, and what each session will be used for. For the moment, we take those decisions as given, although it is possible that the allocation of resources could change during the planning period in response to unanticipated events.

Appendix 8 describes in more detail the production process for elective procedures, and, illustratively, sets out the input resources that need to be considered in order to determine the level of elective procedure capacity.

Again, as with the outpatient production process, what emerges from the discussion in Appendix 8 is that the determination of actual operating capacity is no easy matter. The information that this requires, however, is crucial to managing the supply of services. At its simplest, this includes:

- the number of:
 - operating theatres
 - theatre sessions
 - consultants
 - anaesthetists
 - nurses (or nurse teams)
- the length of theatre sessions
- the average length of an operation.

Discussion

Earlier, we identified some of the key issues raised by taking a systems view of elective care: that is, the relationships between different systems and between parts of the whole system which need to be understood. We then looked in more detail at the demand for and supply of elective care and the information a hospital requires to plan its activity, and subsequently to understand what is happening to the relevant flows and how it is performing in dealing with them.

We have identified two kinds of requirement: an understanding of key relationships (including the factors making for change in the situation a hospital faces), and an ability to describe numerically the main features of the demand for, and supply of, care at a detailed level.

Although some hospitals have much of the information required for improved planning and operational management, in our original survey of nine trusts with good, indifferent and poor performance on waiting times, we found none where this was being used in a systematic way, in the context of an understanding of how elective care relates to the various systems within which it is located. Instead, while those performing well appeared to have a better grasp of these issues, we found that the analysis underlying planning treats parts of the overall system as separate entities with little or no recognition of the various interconnections, knock-on effects, and different system levels that we described earlier.

What a hospital requires is a model of the whole system within which elective care is but one part. Implicitly or explicitly, any hospital devising a 'production plan' must make assumptions about the relationships determining how its elective care system works, within whatever means it uses, such as computer-based models, spreadsheets, or simple arithmetic formula. The systems questions we have identified focus on areas where assumption may be proved wrong by system behaviour (either within a system or between systems).

Some of the relationships or systems issues will be hard for any one hospital to address fundamentally. For example, the nature and scale of the feedbacks within the elective care system are more appropriate for centrally commissioned research. Detailed information is set out earlier in this paper (see 'The policy environment', p 82, 'Understanding demand', p 85, 'The production of care: outpatient services', p 88, 'The demand for elective operations', p 89). This type of information is primarily required for management purposes, but it can also be used to test whether or not some of the key relationships (such as cross-referrals between consultants and specialties and conversion rates) are changing. In other words, it has analytic as well as managerial value.

A critical issue that we have not addressed here is the relative merits of investing in more detailed information of the kind set out in sections 1–4, as against more general, sometimes qualitative, information that bears on the systems questions set out above. A particular example of this is in determining where the major sources of variation and uncertainty may lie.

Thus, as we noted above, the starting point must be to ask the question 'How does the elective care system work?'. The prospect is that the answer to this question – known only imperfectly now – is about to change as the new arrangements start to take effect.

Further research

This research started with the known observation that waiting times varied considerably across trusts, specialties and clinicians. Given this, the obvious research question is why this is the case, and further, how have some (albeit relatively few) trusts managed to reduce waiting times and subsequently sustain relatively short waits for their patients?

The first stage of our research confirmed a suspicion that there was no single answer to this. Rather, success in reducing waiting times involved a number of elements, which we summarised as four key factors:

- a sustained focus on the task
- an understanding of the nature of waiting lists
- detailed information, analysis, forecasting, monitoring and planning
- development of appropriate capacity.

None of these (nor the further issues we identified within each factor) are necessarily surprising and, it could be argued, all represent key aspects of good management per se.

However, from this first stage of the research, a number of further issues emerged that merited further investigation. Two have been reported here: an attempt to quantify possible clinical distortions arising from attempts to meet maximum waiting times targets, and a whole-systems approach to defining a framework of the sort of information trusts required in order to tackle waiting times reductions.

Although the NHS has made great strides in reducing maximum waiting times over the course of this research, going further to meet the NHS Plan three-month waiting time target, and then sustaining such a reduction, will require continued and ongoing effort. Furthermore, while the first stage of our research sets out good management practice with respect to reducing (and sustaining reductions in) waiting times, additional (policy-orientated) research could focus on gaining a better understanding of the phenomenon of waiting lists – particularly from the point of view of a key group: clinicians.

Moreover, possibilities for extending the nature of the waiting times targets to include targets for reducing the actual waiting time experience of patients (that is, the time taken from GP referral to admission to hospital) raise additional demands on the NHS from the point of view of the four factors identified in the first stage of our research. Given all this, further research issues could include:

- evaluation of workload planning tools, such as demand-forecasting models, production models and ‘market analysis’ techniques
- investigation of efficiency of operating theatre production processes
- descriptive analysis of variations in clinicians attitudes to waiting, including approaches to setting priorities for operation lists
- prospective, clinician-level research into the issue of possible distortions to clinical priorities arising from setting waiting times targets
- investigation of clinicians’ relationships with their management colleagues as a possible explanation for variations in complaints of clinical distortion (arising from attempts to meet targets)
- further analysis of clinical distortions issue and impact of waiting times targets on waiting times distributions through extension of ‘before-and-after’ analysis of waiting times distributions for 2002/03 and for specialties other than trauma and orthopaedics

- re-analysis of ‘before-and-after’ waiting times distributions by trusts grouped by waiting time performance, to investigate possible relationships between waiting times and impact of waiting times targets
- analysis of the ‘super-short’ waiting phenomenon: what is the explanation for the very large numbers of patients admitted having spent no days, or only one day waiting?
- whole-systems analysis of the implications of extending waiting times targets and goals on reducing the actual waiting time experience (in other words, from GP referral to admission).

References

- Appleby J, Boyle S, Devlin N, Harley M, Harrison A, Locock L (2003). *Sustaining Reductions in Waiting Times: Identifying successful strategies*. Interim report to the Department of Health. London: King's Fund.
- Audit Commission (2002). *Access to Care: Ear, nose and throat and audiology services*. London: Audit Commission.
- Audit Commission (2003). *Operating Theatres*. London: Audit Commission.
- Auditor General for Scotland (2002). *Review of the Management of Waiting Lists in Scotland*. Audit Scotland.
- Department of Health (2002). *Reforming the NHS: Financial flows: introducing payment by results*. London: Department of Health
- Dixon J, Damiani M (2002). *Managing the Pressure: Emergency hospital admissions in London, 1997–2001*. London: King's Fund
- European Observatory (2000). *Health Care Systems in Transition: Norway*. London: European Observatory on Health Care Systems.
- European Observatory (2001). *Health Care Systems in Transition: Denmark*. London: European Observatory on Health Care Systems.
- Gravelle H, Smith P, Xavier A (2003). 'Performance signals in the public sector: the case of health care'. *Oxford Economic Papers*, vol 55, pp 81–103.
- Hamblin R, Harrison A, Boyle S (1998). *Access to Elective Care: Why waiting lists grow*. London: King's Fund.
- Hanning M, Spangberg U (2000). 'Maximum waiting time – a threat to clinical freedom? Implementation of a policy to reduce waiting times'. *Health Policy*, vol 52, pp 15–32.
- Harley M, Jayes R, Yates J (1999). *Long and Short Waiting Times in ENT*. Report for the Department of Health. Birmingham: Inter Authority Comparisons and Consultancy.
- Harrison A, New B (2000). *Access to Elective Care: What should really be done about waiting lists*. London: King's Fund.
- Hurst J, Siciliani L (2003). *Tackling Excessive Waiting Times for Elective Surgery: A comparison of policies in 12 OECD countries*. Paris: OECD.
- King's Fund (2002). 'Waiting', in Appleby J, Coote A (eds) *5-year Health Check: A review of Government health policy 1997–2002*. London: King's Fund.
- Lian O, Kristiansen I (1998). 'Waiting list guarantee: between medicine and bureaucracy'. *Tidsskr Nor Laegeforen*, vol 118:25, pp 392–106.
- Locock L (2001). *Factors Contributing to the Achievement of Low Waiting Times in Dorset*. Birmingham: Health Services Management Centre, University of Birmingham.
- Martin RM, Sterne JA, Gunnell D, Ebrahim S, Davey Smith G, Frankel S (2003). 'NHS waiting lists and evidence of national or local failure: analysis of health service data'. *British Medical Journal*, vol 326, p 188.
- Met Office (2001). *Forecasting the Nation's Health: An evaluation by the Health Forecasting Unit*. Bracknell: Met Office.
- Ministry of Health and Social Affairs (1999). *Act of 2 July 1999 no 63 relating to Patients' Rights*. Oslo: Ministry of Health and Social Affairs.

Modernisation Agency (2002a). *Sustainability and Spread in the National Booking Programme*. Research into practice programme. Summary report no 2. London: Modernisation Agency.

Modernisation Agency (2002b). *Spreading and Sustaining New Practices: Sharing and learning from the Cancer Services Collaborative (CSC)*. Research into practice programme. Summary report no 3. London: Modernisation Agency.

Modernisation Agency (2002c). *The Big Wizard*. London: Department of Health.

National Audit Office (2001). *Inpatient and Outpatient Waiting in the NHS. Report by the Comptroller and Auditor General*. HC 221 Session 2001/02. London: The Stationery Office.

National Audit Office (2002). *Inappropriate Adjustments to NHS Waiting Lists. Report by the Comptroller and Auditor General*. HC 452, Session 2001/02. London: The Stationery Office.

Pawson R, Tilley N (1997). *Realistic Evaluation*. London: Sage Publications.

Pettigrew A, Ferlie E, McKee L (1992). *Shaping Strategic Change: Making change in large organisations: the case of the National Health Service*. London: Sage Publications.

Ramchandani M, Mirza S, Kirkby G (2002). 'Pooled cataract waiting lists: views of hospital consultants, general practitioners and patients'. *Journal of the Royal Society of Medicine*, vol 95, pp 598–600.

Secretary of State for Health (2000). *The NHS Plan*. Cmd 4818–I. London: The Stationery Office.

Appendix 1: Terms of reference, original research proposal

The aim of this study is to identify factors linked to effective waiting times performance (maintaining low/no proportion of inpatients waiting > six months) by comparing and contrasting three types of NHS trust which have been:

- (a) successful
- (b) unsuccessful
- (c) temporarily successful.

The first part of the study will establish a historical statistical profile of activity patterns (including use of private sector), resource use, referrals, workloads, and productivity for all trusts and waiting list specialties. Interviews with key stakeholders in each health economy will aim to establish management actions on waiting times, consultant management of referrals, the use of financial and other incentives, and how those involved explained waiting times performance.

The outputs from Part One will be a *prima facie* list of relevant factors influencing good and bad performance and possible relationships between them. The second, prospective, stage of the study will draw on these findings and test hypotheses for good performance by instituting appropriate changes in some of the (b) and (c) group trusts and tracking changes in waiting times and other factors for a further period of time.

Appendix 2: Selection criteria for trusts included in the study

On the basis of their historic trends in the proportion of their total inpatient and day case lists waiting over six months, three groups of trusts were identified for study.

For all trusts in England, the all-specialty inpatient (including day cases) quarterly waiting lists from June 1998 to March 2002 were obtained from the Department of Health's waiting times website. In addition, the numbers of patients who had been on the lists for over six months were obtained, and expressed as a percentage of the total lists.

The March 2002 figures were used as the basis for selecting the six trusts with very low or very high percentages waiting more than six months. In order to exclude trusts with little inpatient activity (such as community trusts), all trusts with a total waiting list of less than 1,000 patients were eliminated from further consideration, as were specialist trusts (for example, those providing orthopaedics only).

The remaining trusts were sorted in ascending order of the percentage waiting over six months, since one of the key government targets for 2005 is that no one should wait more than six months for admission.

Successful trusts

We selected three trusts with very low percentages waiting over six months (trusts A, B and C). All three of the selected trusts had had relatively low percentages waiting over six months each quarter since June 1998.

Temporarily successful trusts

We selected three trusts that showed significant change over time in percentages waiting over six months (trusts G, H and I).

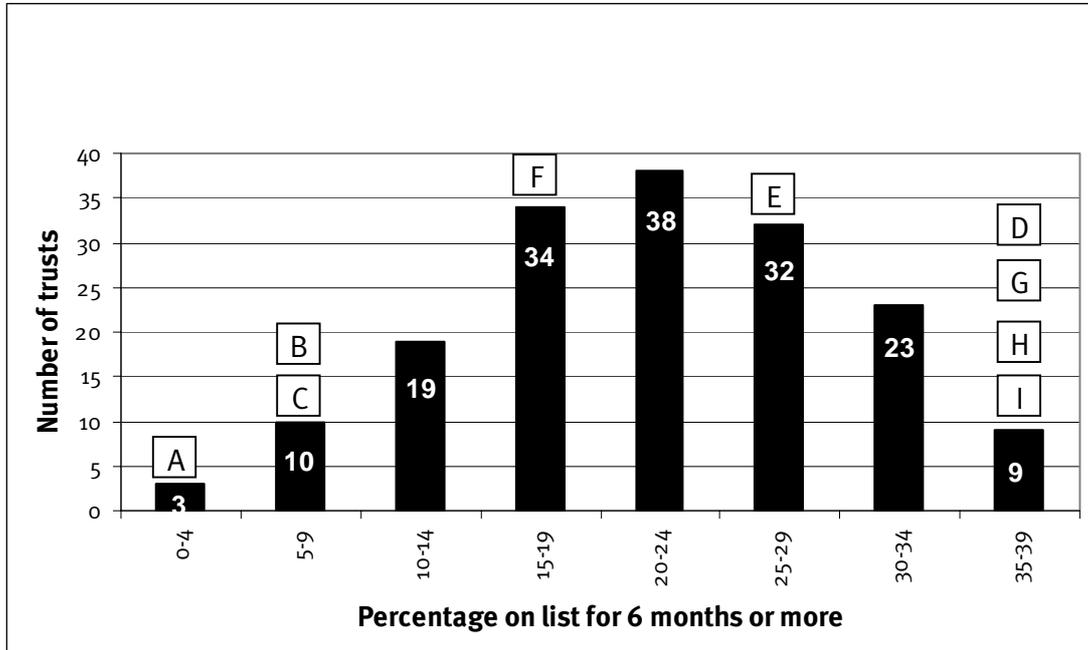
The first of these had shown a fairly steady and significant reduction in the percentage waiting over six months from a high value in June 1998 to a relatively low value in June 2001, but had not improved further since then. The second had shown a reduction in the percentage waiting over six months from June 1998 to March 2000, but the percentage had then deteriorated again to its original value until September 2001, after which the position improved again. The third had shown a large reduction in the percentage waiting over six months from June 1998 to September 2000, but the situation had then deteriorated rapidly until it was worse than the starting position.

Unsuccessful trusts

Finally, we selected three trusts with very high percentages waiting over six months (trusts D, E and F). Although there were other trusts with similar or worse performance, one was excluded because it involved a recent merger, another because it was already the subject of intensive scrutiny, and a third because we were including it in our temporarily successful group. All three of the selected trusts had had relatively high percentages waiting over six months each quarter since June 1998.

Figure 42 shows the trusts selected in relation to the total number of English trusts (following exclusions noted above). Figure 2 in the main report shows waiting time trends for all nine trusts.

Figure 42: Acute trusts, all speciality waiting lists – March 2002



King's Fund (2003)

Appendix 3: Interview schedule

Introduction

Explain background and purpose:

The Department of Health commissioned the research:

- to understand the range of factors affecting waiting times in different trusts
- to try to explain different experiences in managing to keep waiting times low or no
- trying to identify generalisable factors which could apply to other sites
- trying to identify how far a trust's waiting times performance is a product of unique local context
- and we are looking specifically at inpatient and day case waiting times, but interested to know about possible inter-relationships with outpatient waits.

Happy to be taped? No individual names will be identified, and no sites will be named unless the trust chooses to be identified.

We will circulate our draft interim report (around end of December/early January) for information/comment.

A: Background/warm-up questions

1. How long have you worked in this trust (and other parts of the local health economy)? In what roles?
2. What has happened to waiting times in your trust?
3. Do the formal performance figures feel like a realistic picture, or is the reality better/worse?
4. Is this a recent situation, or has it always been like this?
5. Is your trust's performance similar to neighbouring trusts? In what ways different?

B: Factors affecting waiting times

(Show graph of % of people waiting over six months)

1. Can you briefly talk me through the trends and changes in trends on this graph, summarising some of the factors which you feel explain this picture?
2. In brief, have there been particular obstacles in achieving reductions in waiting times?
3. How have these been overcome (or not)?

C: Hospital culture

1. Can you describe how you see the culture of this hospital? (Such as?)
2. What is morale like locally?
3. Recruitment and retention?
4. Who are the key managerial and clinical leaders?
5. How long have they been in post?

6. What are relationships like? (between managers, between clinicians, between managers and clinicians)
7. Are there any very difficult relationships?
8. What impact has the hospital's culture had on the management of waiting times? (Such as?)
9. Is the general attitude positive or negative in relation to waiting time reduction?

D: Hospital organisation

1. Does the trust/each specialty have a waiting list manager? Is this a full-time job?
2. Did the local health authority have a waiting list manager?
3. Does the trust have a discharge co-ordinator?
4. Is elective and emergency work split across sites?
5. Is elective work/facilities 'ringfenced' – protected from emergency work to any degree?
6. Are there any particular problems with emergency work encroaching on elective work in any way?
7. Is there one or more dedicated day case units?
8. Does the trust have an admissions ward?
9. Does the trust have a high dependency unit? (Note: all of our trusts have intensive care beds.)
10. Does the trust have access to intermediate care beds? (Where, numbers?)
11. What percentage of day cases is booked (all specialty totals)?
12. What percentage of inpatients is booked (all specialty totals)?
13. Does the trust make use of one-stop clinics?
14. Is pre-operative assessment combined with the out-patient appointment?
15. Is booking/allocation of outpatient appointments centralised?
16. Do consultants pool waiting lists?

E: Hospital management:

1. Can you summarise for me the trust's financial position over recent years?
2. What is the financial situation this year?
3. If heading for a deficit, what are the reasons for this?

4. Have waiting times been a constant priority locally, or intermittent, driven by specific initiatives?
5. How important have national waiting list initiatives been in the past?
6. Has the trust had access to reliable information to help manage waiting times (what sort, who sees it, how often etc)?
7. How far has there been clinical ownership and support for waiting times initiatives?
8. Has the trust asked consultants to do extra sessions at weekends?
9. Has the trust asked consultants to do extra sessions in the evening?
10. Has the trust asked consultants to do extra sessions in private hospitals?
11. Are consultants paid for these extra sessions?
12. Are these extra sessions continuing?
13. Have other measures been taken to increase capacity, such as new posts or theatres?
14. Have there/are there any particular staffing problems?
15. Has the trust carried out any studies of its own or introduced new management measures to improve performance (in other words, are any reports available?)
16. Has/does the trust used/use any models to plan capacity or manage waiting lists (such as a checklist)? Is this useful?
17. Has the trust/health authority/primary care trust (PCT) used any direct financial or other incentives to reduce waiting times?
18. In your opinion, of all the initiatives and changes you have mentioned, which have contributed most to reducing waiting time?
19. Is there any hard evidence to support your opinion?

F: Effect on clinical practice

1. Has the drive to reduce waiting times affected the way clinicians practice?
2. How have clinicians reacted?
3. Has it distorted priorities?
4. Have there been any knock-on consequences on other services or on quality of care?

G: Local health economy

1. What are relationships like between local health organisations?
2. What are relationships like between primary and secondary care?

3. Do local trusts get on well together or compete? Did they get on with the health authority?
4. How important has the role of the health authority and its chief executive been? How has that role been managed during reorganisations from hierarchy to quasi-market and now to 'partnership'?
5. Have your commissioners' decisions or actions made much difference to the management of waiting times?
6. What incentives or sanctions have they used?
7. Do referring GPs or PCTs have quotas (slots) for numbers of referrals?
8. Are formal protocols in place for managing the primary/secondary interface and for which specialties?
9. Does the trust allow GPs to book day cases direct on to the list?
10. Are GPs given feedback on appropriateness of referrals?
11. Does the trust run outpatient clinics in primary care premises?
12. Are there any GP specialists providing initial consultations in the area?
13. What are the demographics of the local population here? Has that made any difference to demand?
14. What are the referring patterns of local GPs?

H: Private sector

1. Has the trust used the private sector to reduce waiting times?
2. What has been the attitude of senior staff and local purchasers to using the private sector?
3. Has there been spare capacity locally?
4. Is the trust still using the private sector?
5. Has the health authority/PCT used the private sector?

I: Wrap-up

1. Reflecting on all the helpful and unhelpful factors we have been discussing, which do you think have been the most crucial?
2. What lessons would you identify from your experience that others could learn from (good and bad)?

Appendix 4: Survey questionnaire



The impact of guaranteed maximum waiting times on clinical priorities: A questionnaire

Sustainable reductions in waiting times: Identifying successful strategies

A study funded by the Department of Health and managed by the King's Fund

August 2003

CONTACT: John Appleby, Chief Economist, King's Fund, 0207 307 2540,
jappleby@kehf.org.uk,
11-13, Cavendish Square, London W1G 0AN

These codes identify the trust and specialty and are for analytic and administrative purposes only. Neither trusts nor consultants will be identified.



Trust code	Spec. code	Form code

Background information to this survey

A research group from the King's Fund, the London School of Economics, the University of Birmingham and City University is undertaking a study on behalf of the Department of Health to identify successful strategies for achieving sustained reductions in waiting times. A report on the first stage of the work is available on the home page of the King's Fund website (www.kingsfund.co.uk).

As part of the second stage to the study we are now investigating the impact which waiting times targets have had on the order in which patients have been admitted from waiting lists.

To help us with our work we would be very grateful if you could complete this short questionnaire and return it in the attached prepaid envelope.

All results from this survey will be anonymous. Responses will be coded to identify trusts, but individual responses are not relevant to this study's aims.

If you have any questions or would like to discuss any issues raised in the questionnaire, please contact:

John Appleby
Ruth Thorlby

Tel: 0207 307 2540 or email jappleby@kehf.org.uk
Tel: 0207 307 2646 or email rthorlby@kehf.org.uk

PLEASE RETURN ALL SIX PAGES OF THIS SURVEY IN THE ENVELOPE PROVIDED BY:

SEPTEMBER 9, 2003

ALTERNATIVELY, YOU CAN FAX THE QUESTIONNAIRE TO: JOHN APPLEBY ON 0207 307 2807

1. In general, given existing resources, how long do you think it is reasonable for patients to have to wait to see you for treatment?

	Urgent case	Routine case
For a first outpatient appointment	_____ months	_____ months
For admission from the inpatient waiting list	_____ months	_____ months

2. What actions did you take in order to meet the March 2003 maximum inpatient waiting time target of 12 months?

(Tick as appropriate)

None, no need as no patients waiting over 12 months	_____
Held additional theatre sessions during working week	_____
Held additional theatre sessions outside normal working week	_____
Some patients treated privately – paid for by the trust	_____
Some patients transferred to another consultant	_____
Some patients deferred after clinical review	_____
Some patients reclassified as outpatients	_____
Potential 'breachers' accommodated within existing theatre sessions	_____
Other (detail)	

3. In the year leading up to the March 2003 target of 12 months maximum inpatient waiting time, did you treat patients in a different order to that suggested by their clinical priority as assessed by you?

	(tick one) ↓	If YES, how many patients treated sooner than clinical need suggested?	And how many patients had their treatment delayed as a result?
Yes	_____	_____ (number)	_____ (number)
No	_____		

4. If you ticked the ‘Yes’ box in Q3, can you say, on average, what the impact this changed priority had on the waiting times of the affected patients?

(number)

Patients treated sooner _____ fewer days on waiting list

Patients delayed _____ extra days on waiting list

5. Again, if you ticked the ‘Yes’ box in Q3, can you say what the impact this changed priority had on the clinical condition of the patients involved?

	Patients treated sooner (Tick one)	Patients delayed (Tick one)
■ A major negative impact	_____	_____
■ A minor negative impact	_____	_____
■ No impact	_____	_____
■ A minor positive impact	_____	_____
■ A major positive impact	_____	_____

6. If you ticked the ‘Yes’ box in Q3, can you provide examples of patients treated in a different order to that suggested by their clinical priority: (eg the procedures involved).

Date	Example

7. Please add any other comments in relation to the issues raised in this questionnaire.

Please add your name and trust (in capitals) here. This is only to help us in following up non-responders to the survey. This page will be detached and destroyed on receipt of the questionnaire.

Name:

Trust:

Appendix 5: Examples and comments from the consultants' questionnaire

Examples of distortion

- 'Patients in soon or urgent categories for joint replacement waiting for much longer. Causes 'waiting list hip' – a hip in which a simple joint replacement is possible when put on waiting list but by the time they come to be done a major joint reconstruction is required. This takes twice as long to do, probably has a less good outcome and patient stay is much extended.'
- Patient waiting for total hip replacement who was unable to get out of the house because of OA hip who waited as long as anyone else, ie one year!'
- 'Smaller cases in order to fit a larger number of patients on an operating list.'
- 'Foot deformity put ahead of older patient in pain from knee arthritis for nearly 12 months. Planned removal of 'metalwork' brought forward as needed to meet time target when clinically better left three months longer.'
- 'Most "soon" procedures expected within six months regarded as routine – 12 months, eg knee arthroscopy, total hip replacement, total knee replacement.'
- 'Patients rushed through to meet target when they could have waited without detriment. Many orthopaedic conditions are long standing and can wait. Movement from original consultant to one without a specialist interest may lead to a less satisfactory operation. Change of consultant to meet target means that they lose continuity of care and have a different opinion.'
- 'Routine cases, ie repair eardrums, otosclerosis brought in at expense of patients unwell with eg recurrent tonsillitis, sinusitis and nasal blockage. No cancer patients were delayed to my knowledge.'
- 'Lists being filled with hernias and gallbladders resulting in longer waits for patients awaiting cancer surgery (while still remaining within the 30-day target).'
- 'Long-wait cholecystectomy patient with few symptoms treated before cholecystectomy patient with possible ductal stones.'
- 'Hernias (asymptomatic) treated sooner than more symptomatic ones. Routine cholecystectomy performed prior to soon/urgent.'

General comments

- 'I work in a trust with very aggressive approach to long waiters. By using all possible alternatives the waiting times are significantly reduced. It has been at the expense of maintaining proper clinical governance processes, medical staff education and development.'

- ‘Waits for surgery and outpatients are distorted by many factors. For example, we operate on patients from Wales sooner than those from England because we are paid to do so. Currently the Government is paying to reduce cataract waiting time but not those for squint surgery or any other surgery.’
- ‘Only been consultant since January 2003, but my own inpatient waiting list was building up while I did long waiters from other lists.’
- ‘My problem is with outpatient waiting times. My clinics (and those of my colleagues) are under continuous pressure exacerbated by the need to meet artificial targets. The partial booking system, and the hospital call system, have improved efficiency and driven down the DNA rate, but the politically imposed obsession with GP referral access times means that a patient with a trivial complaint referred by a GP (target driven) takes priority over a patient referred internally from a colleague (doctor, nurse speech therapist – therefore not target driven) who believes the patient may have a serious disease such as cancer.’
- ‘Creates serious friction between consultants caring and talking to patients and managers trying to meet targets.’
- ‘Knock-on effect of extra cataract throughput results in follow-up patients with blinding conditions such as diabetic retinopathy or glaucoma having appointments postponed on multiple occasions – decisions taken by appointment clerks.’
- ‘Pressure to treat non urgent/potential breachers has been resisted as most of my workload in cancer related. However, I am aware of cases being displaced by a week or two to accommodate targets. I have no evidence that this has compromised treatment but long waiters do compete with the cancer two-week wait [target] patients for the limited resources.’
- ‘My trust encouraged evening and weekend operating lists. There were financial incentives for some staff but not others. Clerical staff in particular felt that their extra work was neither recognised nor remunerated. I believe extra payments to medical staff also caused resentment within the medical workforce, eg paediatricians do long hours compared to me yet my pay was greatly enhanced. These measures will only work in the short term.’
- ‘The process adds extra pressure on everyone. Physicians normally prioritise based on clinical need and this is likely to remain so. However, the added pressure of targets means that patients with less clinical need are being pushed forward. This has a concertina effect on clinics and operating lists. This impacts on everyone in the NHS – doctors, nurses, admin staff and secretaries. It is essentially a subtle form of coercion or bullying.’
- ‘Consultants are no longer in control of waiting lists and are unable to keep things under close review. Result in delay of urgent patients in order to get long waiters done on time.’
- ‘It is impossible to guarantee waiting times unless there is some kind of cap on referrals. We eliminated out the varicose vein waiting list by negotiating agreed priorities and a cap on the total number of GP referrals a year. This allowed us to eliminate the waiting list with extra clinics and lists and to maintain control.’

'I did not have any patients waiting for more than 12 months but I did operate on some other consultants patients to meet the target. I saw my colleagues operate on minor cases while major more deserving cases were waiting. This was done with the management's encouragement so the numbers operated on will be high (you can operate on six minor cases on one list or just one major case). Some minor cases who could have waited without any harm for many months were operated on within days of being seen to meet the initiative lists.'

Appendix 6: The outpatient production process

This appendix is illustrative of the factors and analysis (and, by implication, the sort of information required) that, ideally, hospitals need in order to carry out in order to plan the productive capacity of their outpatient departments. What follows is not intended as a detailed blueprint for running and organising a typical hospital's outpatient department.

Determining capacity

Clinics

The capacity of the outpatient service over a typical week depends primarily on the availability of consulting rooms and clinical staff. Either of these could be a constraint on delivery. However, we assume that the total number of consulting rooms available provides a maximum number of clinics at any one time. (This rule could be broken if other rooms could be brought into service, or in the long term new facilities could be provided. However for planning purposes it seems a reasonable assumption.) At this point, the availability of clinicians becomes the constraint on the number of clinics.

Of course, there will have been decisions made about the allocation of rooms between different uses and specialties, as there will also have been about the different use of clinician time. However, for the moment we take these decisions as given.

Slots

To determine how many patients can be seen in these clinics, a hospital needs to determine an average consultation time, as this will then be used to provide the number of slots per clinic. Some patients will take longer, and others less than the average, by definition. Given that the average consultation time will be based on historic data, the average time is likely to change, however. If it increases, then the unfortunate consequence is that the clinic time overruns. (We are aware that many outpatient clinics allocate more than one patient to a single slot so as to ensure that clinician time is not wasted.)

Testing and diagnostics

So far, we have treated an outpatient consultation as if it were a discrete service occurring at one point in time. Frequently, however, the consultation will result in further tests being required. When these take place 'at once', then the patient is often required to resume the consultation with the clinician. In these cases, the allocation of slots must be able to take this into account.

Clinics will tend to be either ones where such splits in time of the consultation occur practically all the time, or else where this is not the case. Clinics therefore need to be organised in advance to take this into account.

However, another unknown is that the patient will generally enter a queue for one or more tests and will arrive back for the continuation of the consultation at an uncertain time. Again, there will be an average time between going for tests and returning; this needs to be factored into the organisation of the clinic.

Other factors

A cause of concern to the service is the number of patients who fail to attend on the day without giving any prior warning ('did not attend', or DNA). The result is that there are fewer consultations than the maximum, and often the patient has to be given a slot at a later date.

Patients may give warning that they cannot attend, but it is impossible to fill the slot with another patient in which case, again, there is unused capacity. The high proportion of DNAs has led some services to overbook patients for clinics in order not to waste slots. This can be quite effective from the viewpoint of ensuring capacity is used but it can also result in inconvenience to patients who are double-booked in this way, and hence wait longer and/or have a less satisfactory consultation through shortage of time.

Types of clinic

All clinics could be treated as essentially the same. However, hospitals have several possible ways of distinguishing clinics. These include by:

- consultant team
- sub-specialty
- complex or simple
- new or follow-up
- priority or routine.

Any of these may be an effective way of differentiating outpatient services. Problems occur, however, if the result is that demand for some clinic slots outstrips supply, while for others there is spare capacity. In such circumstances, it is important to consider whether the way in which clinics are categorised is making a useful distinction between patients.

Thus, if two consultant teams provide the same service with the same capability, it would seem wasteful of resources if it were not possible to switch patients between clinics to use up all available slots. On the other hand, if one clinic is more expert at dealing with hands and another is more expert at dealing with legs, this may be a sensible division of clinic space. Ultimately, such divisions will depend on the structure of demand for services.

It may also be useful to be able to distinguish patients according to the likely time a consultation will take. For example, follow-ups may on average take less time than first appointments. Hence the allocation of slots should take account of this. This can be achieved within the same clinic, but it requires effective use of the available information. Similarly, some consultations may require less time than others. If this distinction can be made, this will help in the allocation of slots to patients.

There is yet another distinction: that between patients who have an urgent need to be seen, and others who might be classified as routine. This distinction does not always correspond with the likely time that a consultation will take.

Drawing on additional capacity

There may be circumstances when it will be necessary to run additional clinics or to purchase these in some other way – for example, if it becomes clear that waiting time targets will be breached.

There are various possibilities, including:

- outpatient clinics outside normal hours – which thus makes more intensive use of existing capital stock
- making available more clinics within existing hours, by switching room use and increasing the load on clinicians
- using facilities elsewhere – for example, in the private sector.

A hospital needs to be aware of these possibilities in planning its service provision. There may be times when this is the most cost-effective way to deal with a temporary increase in demand rather than having excess capacity throughout the year.

Appendix 7: The elective inpatient production process

This appendix is illustrative of the factors and analysis (and, by implication, the sort of information hospitals require) that, ideally, hospitals need to carry out in order to plan the productive capacity of their elective inpatient work. What follows is not intended as a detailed blueprint for running and organising a typical hospital's elective inpatient care work.

Determining capacity

The capacity of an elective care service depends primarily on the availability of operating theatres, and of clinical and other staff necessary for the performance of the operation. A key aspect of capacity is the availability of operating slots.

Number of operating slots

It is difficult to be definitive about the number of slots that are available in any one theatre session. Operations will clearly vary in length. It is possible to provide an estimate of how long on average different types of operation will take. However, there are various compositions of types of operation to fill the time available that might be chosen.

So, if an operating session lasts four hours, the operating team might choose to do eight half-hour operations or three one-hour operations and two half-hour operations, or some other combination. Thus for any given session, we cannot say precisely what the capacity is in terms of number of operations performed (though we can in time available). However, the average operation length (which can be calculated from historic data) can be used to give an overall impression of the total capacity. This could be considered as an absolute production maximum. If sessions are badly planned or managed, then this capacity is unlikely to be attained. But there are a number of other factors to take into account that may result in this maximum not being attained.

Other factors

Although a certain number of operations are planned to take place within a session, it is possible that some may over-run, resulting in the cancellation of an operation, and hence less being produced than was intended. This is a natural consequence of having used the average to determine what is possible. Where the operation takes longer than expected there could be a displacement of an operation. However, where it takes less time than expected, it is very unlikely that another operation can be fitted in (not least because the service does not operate in this 'just in time' way).

Operations could also be cancelled if any of the resources needed to make it happen are not available – at the last moment. This includes the various clinicians involved. It is also possible that a shortage of available beds will result in the cancellation of operations. This illustrates the advantage of operations which can be carried out as day cases, and hence are not affected by bed availability.

Drawing on additional capacity

There may be circumstances when it will be necessary to run additional theatre sessions or to purchase these in some other way – for example, if it becomes clear that waiting time targets will be breached.

There are various possibilities, including:

- theatre sessions outside normal hours, which thus makes more intensive use of existing capital stock
- making available more theatre sessions within existing hours – by switching theatre use between specialties, clinicians, or ‘hot’ and ‘cold’ work
- using facilities elsewhere – for example, in the private sector.

A hospital needs to be aware of these possibilities in planning its elective surgery provision. There may be times when this is the most cost-effective way to deal with a temporary increase in demand rather than having excess capacity throughout the year.