Waiting list behaviour and the consequences for NHS targets

John Bowers
University of Stirling

Abstract

The United Kingdom's National Health Service (NHS) is investing considerable resources in reducing patient waiting times for elective treatment. This paper describes the development of a waiting list model and its use in a simulation to assess management options. Simulation usually assumes that waiting is adequately described by simple queuing disciplines, typically first-in-first-out. However, waiting in the United Kingdom's National Health Service is a more complex phenomenon. The waiting list behaviour is explored through an analysis of the changes in waiting time distributions for elective orthopaedics in one Scottish Health Board, NHS Fife. The evolving distributions suggest that there have been substantial changes in priorities in response to the various NHS targets. However, in the short or medium term, the form of the distribution appears reasonably stable, providing a basis for estimating future waiting times in different scenarios. A model of the waiting behaviour and prioritisation in the appointment allocations was embedded in a simulation of the complete elective orthopaedic patient journey from referral, through outpatients and diagnostics to surgery. The model has been used to explore the consequences of various management options in the context of the NHS target that no patient should wait more than 18 weeks between referral and treatment.

Keywords: health service, simulation, queuing

Introduction

Waiting is a key characteristic of the United Kingdom's NHS. Unlike many aspects of the quality of care, waiting is relatively easy to measure and has thus received much attention. Targets have been set for short timescale waits, such as the wait in Accident and Emergency departments, and also the longer waits for elective outpatient or surgical appointments. While there is much debate (Bevan and Hood, 2006; Carvel, 2007) about the value of performance management in the NHS and the effect of the emphasis on the quantifiable, simple measures, these targets are the focus of much management effort.

Hospital management require models to help forecast their waiting lists and assess proposals for achieving the desired improvements. Some proposals may involve investment in short term "non-recurrent" capacity, such as the purchase of surgical treatment from the private sector to reduce the current waiting list. Other options may enhance recurrent capacity, perhaps by increasing the efficiency of existing clinics or theatre sessions, and help ensure that demand and supply are better matched in the future. Many simulations (Jun et al, 1999) have been constructed to examine the flows of patients analysing the constraints in the system. Also, more specialised waiting list models are available to hospital management (The CheckList Partnership, 2007; Steyn, 2008) providing tools to address the current NHS target to treat patients within 18 weeks of referral for elective, or non-emergency, care. While these models may

distinguish between major patient groups, e.g. the "urgent", "soon" and "routine", they are essentially based on the assumption of a first-in-first-out (FIFO) queue discipline. The FIFO assumption is reasonable in most simulations of systems involving inert entities but it may be less appropriate for healthcare simulations in which patients have very different priorities reflecting their individual needs. Some alternative queuing models have been explored, considering the theoretical impact of adopting prioritisation schemes reflecting clinical need (Goddard and Tavakoli, 2008). Other simulations have incorporated diagnosis dependent prioritisation for specific treatments, such as lung (Ouwens et al, 2002) and liver (Ratcliffe et al, 2001) transplants. There are also examples of simulations adopting waiting list prioritisation based on a combination of the time that the "patient has already been on the list" and "urgency" (Cameron et al, 2006). These studies have explored the likely impact of adopting hypothetical prioritisation systems, helping identify possible approaches that might enhance equity, survival or cost effectiveness. This paper attempts to capture the actual practice, including prioritisation and the many factors that may affect the patient's experience, as reflected in the empirical waiting time distributions. The empirical distributions were incorporated in a simulation designed to help NHS management explore the impact of proposals to meet the 18-week target. The model was developed in NHS Fife, using predominantly local data though some parameters, notably those reflecting possible alternative future practices, were estimated using experience from a number of NHS Health Boards.

Waiting list guarantees and targets

Waiting list targets vary across the United Kingdom; this study was undertaken in a Scottish NHS Board and the targets relevant to the period studied, 2004-7, are noted in Table 1. Each of the targets was associated with a public "guarantee" of a maximum waiting time for each patient. In practice this meant that there was an expectation that considerable effort would be expended to avoid breaching the target wait, with only exceptional cases exceeding this target. During 2004-7 there were separate targets for outpatient and surgical treatment but these were subsequently consolidated in the 2011 target. At the time of the study, NHS management in Scotland were beginning their preparations to meet the target of 18 weeks for the total wait during the whole process, from referral for the initial outpatient consultation to treatment (NHS Scotland, 2008).

Patient group	Maximum wait	Target date
Inpatients and daycases	12 months	1 April 1997
Inpatients and daycases	9 months	31 December 2003
Outpatients and inpatients	26 weeks*	31 December 2005
Inpatients and daycases	18 weeks	31 December 2006
Outpatients	18 weeks	31 December 2007
Abolition of ASC's		31 December 2007
Total wait (outpatient, diagnostics and surgery)	18 weeks	31 December 2011

Table 1 NHS Scotland waiting time guarantees (*for either component)

The outpatient wait is defined as the time between referral (typically from a General Practitioner or an Accident and Emergency department) and the first outpatient

appointment. The inpatient/ daycase wait is the time between a consultant's "decision to treat" and the patient being admitted for the surgical procedure. Additional investment and management effort have been focussed on these targets across Scotland, resulting in a significant reduction in waiting times over the last 10 years (Audit Scotland, 2006). It would appear that the policy has been most successful, though some caution is necessary. At 31 December 2006, 38% of patients were excluded from the count of those with excessive waits, though this had fallen to 13% by 31 December 2007 (ISD Scotland, 2008). These excluded patients were not entitled to a waiting time guarantee and were given an "availability status code" (ASC). Any patient who failed to attend their appointment would be given an ASC reflecting the reason for the need to postpone treatment. While there can be valid reasons for excluding such patients, it has been suggested that the ASC designation has also been used to manipulate waiting list statistics (Moss, 2007). It is intended to abolition the use of ASC's to exclude patients from waiting time guarantees by 2008.

The factors that may affect patients' waiting times

Achieving the waiting list targets involves confronting several issues simultaneously:
- Providing sufficient capacity throughout the patient journey (outpatients, diagnostics and surgery) to meet demand. This may require recruiting additional staff, investing in new equipment or space and making better use of existing resources to increase the "recurrent" capacity.
- Eliminating any existing excessive waiting lists. Typically this can not be achieved using the recurrent capacity alone and some short term additional resources will be necessary. This temporary, non-recurrent capacity is usually provided by purchasing care from the private sector or by financing special "waiting list initiative" clinics and theatre sessions at weekends.
- Managing the waiting list so that the allocation of appointments balances clinical need and the objective to minimise the maximum wait.
- Managing demand and monitoring referral practice so that reductions in waiting times do not stimulate excessive additional demand.

The study involved the construction of a simulation model which was used to explore the recurrent and non-recurrent capacity requirements, though the focus of this paper is the behaviour and management of the waiting list, as exhibited in the empirical distribution of waiting time. The paper considers a single specialty, elective orthopaedics, whereas a fuller analysis would consider the interactions with other parts of the health care system, such as the competition for resources with other specialties. In this paper it is assumed that orthopaedics has a prescribed share of resources such as theatre time, beds and use of diagnostic services.

A simple FIFO waiting list model is often sufficient for such analyses and has provided the basis for many useful studies (Worthington, 1987). Modelling the flows and constraints in the health care system while treating the patients as indistinguishable entities can usefully simplify the analysis and enable a relatively rapid assessment of resource requirements (Jun, 1999; Proudlove et al, 2007). However, more substantive modelling of health services requires some consideration of the personal characteristics of the patients and the attitudes of the staff. The NHS targets place an emphasis on the tail of the waiting time distribution and there is a need to understand the range of patient

experience and not just the mean wait. Discussions with staff in NHS Fife identified numerous factors that could influence the wait of any individual patient. The most important factor is the patient's clinical need, refined by a consideration of the personal circumstances and resultant impact on quality of life. However, the waiting time is also affected by the capacity, relative to the demand, and personal choice with some patients electing to wait longer. The waiting time target and its prominence in the allocation of appointments is just a further factor influencing the waiting time distribution.

In addition to understanding the distribution of the waiting time at any one stage of the patient journey, it is also important to appreciate the inter-relationship between the waits at each stage. Any model needs to include the correlations between waits so that the distribution of the total patient wait can be estimated.

The changing nature of the waiting time distribution

The result of additional investment and attention to waiting times has been a dramatic decline in waiting lists in throughout Scotland (Audit Scotland, 2006). Figure 1 illustrates the decline in the mean wait in elective orthopaedics in NHS Fife during April 2004 – March 2007. There has also been a corresponding large reduction in the tail of the waiting list distribution: the 95% tile wait was approximately 36 weeks in April 2004 but fell to 17 weeks in March 2007. Similar reductions were experienced in other specialties in NHS Fife and other Scottish Health Boards.
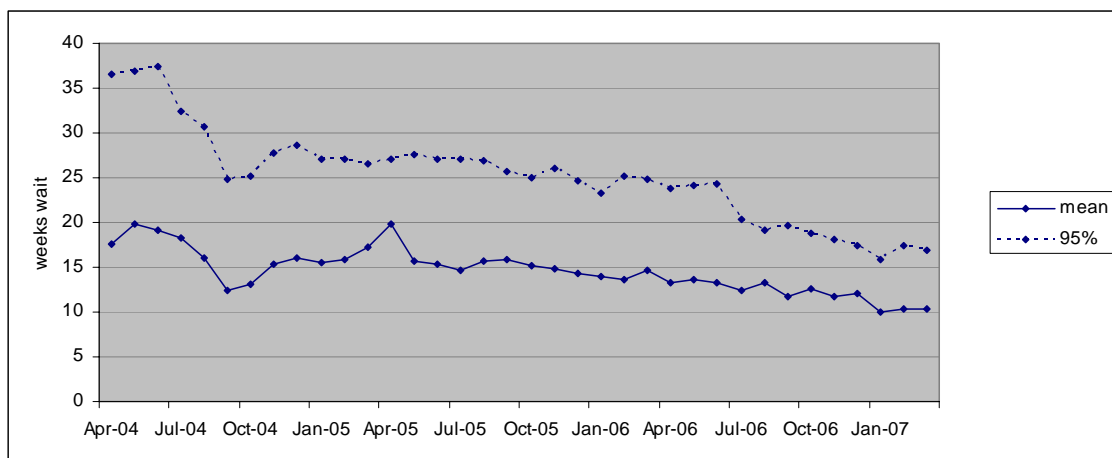

Figure 1 The declining wait for orthopaedic surgery (inpatient & daycase; exc. ASC's)

Figure 2 illustrates the changing nature of the waiting time distribution during 2004-7. The graphs record the waits experienced by patients who were added to the waiting list following a consultant's "decision to treat" in the specified period. During October 2004 - September 2005 a bimodal distribution emerged; this development corresponds to the actions taken to achieve the 31 December 2005 target of a maximum wait of 26 weeks.
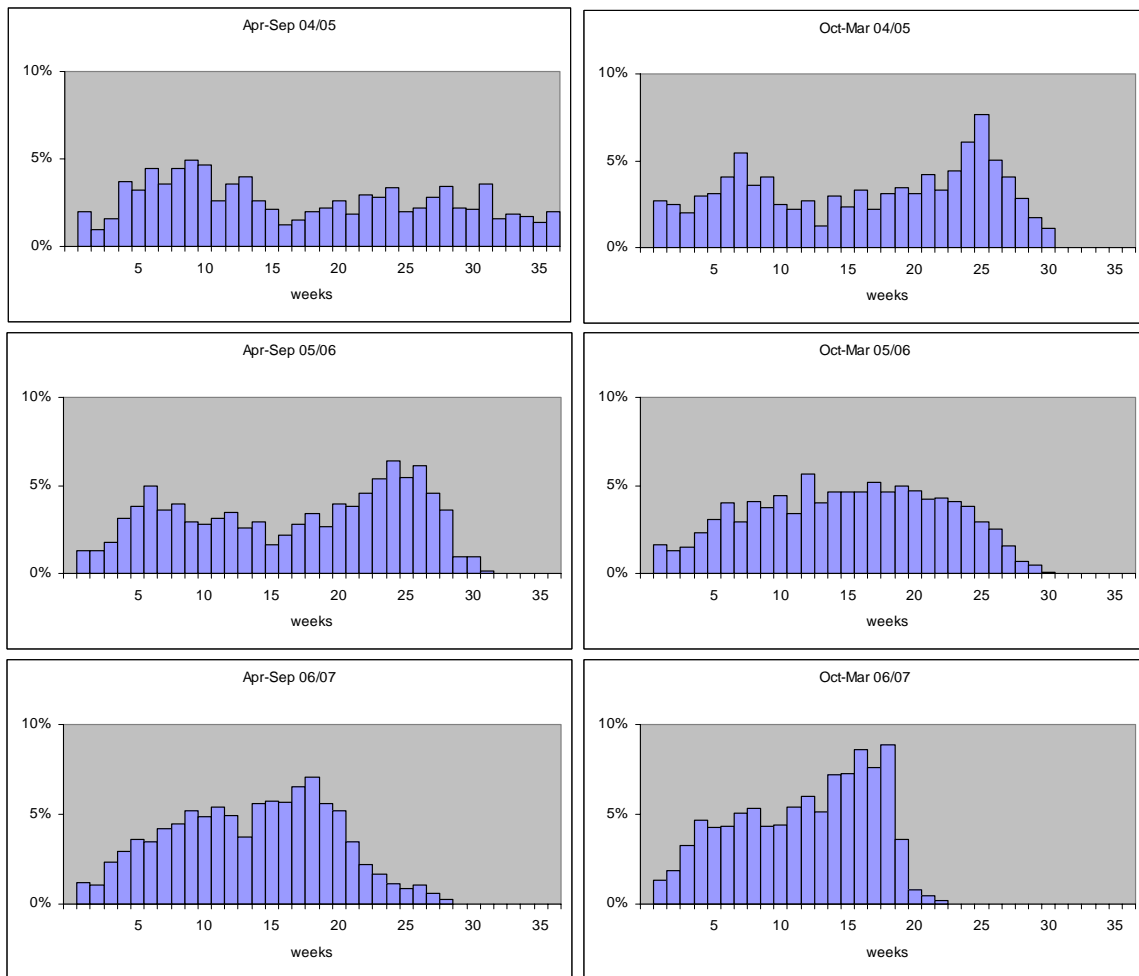
Figure 2 Waiting time for inpatient/ daycase treatment in elective orthopaedics

This bimodal pattern, with an apparent distinction between "urgent" and "routine" patients, has been observed in other specialties, for example in Ear, Nose and Throat (ENT) as illustrated in Figure 3. Discussions with NHS Fife staff suggested that this bimodal distribution could reflect a specialty operating under significant resource constraints: the target wait can only be achieved by focussing resources and strictly prioritising patients. A better resourced speciality could allow more individual discretion in determining patient appointments with less focus on waiting time. During October 2005 – March 2007 the distribution changed again, losing the bimodal feature. The form of the distribution appears to have been stable throughout 2006/7: the mean has reduced substantially from 13.2 to 10.4 weeks but otherwise the distribution has altered little implying that similar priorities have applied. During 2006/7 all patients have benefited equally from the reduction in waiting, not just those who would have been in danger of breaching the target. In a specialty with fewer discretionary resources, the patients with the longer prospective wait might benefit most from a more carefully targeted allocation.
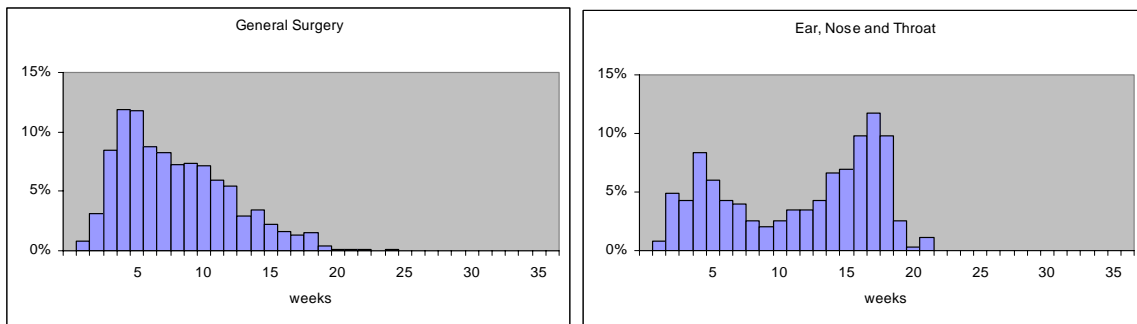
Figure 3 Examples of patients' waiting experience in other specialties (Oct 2006- Mar 2007)

While all the major specialties in NHS Fife reduced their mean waits to a similar degree, the nature of the waiting time distributions varied considerably. Figures 2 and 3 provide illustrations of the three characteristic forms that were observed in orthopaedics, ENT (Ear, Nose and Throat) and General Surgery. Further research to determine appropriate functional expressions could be valuable.

Sources of variability in waiting time

The variability in elective patients' waiting times may be the result of a number of factors. However, the stochastic effects, as in the classic queuing problem, are relatively minor. Queues for elective care in the NHS are usually substantial, compared to the arrival rate, and consequently the queue acts as a buffer absorbing much of the stochastic variation. Furthermore, the variability in arrivals is typically constrained by the capacity of preceding processes, e.g. additions to the surgical waiting list are generated at outpatient clinics. Given the large, buffering queues and constrained stochastic nature of the arrivals, more sophisticated models of congested queuing (Wall and Worthington, 2007) were not necessary in this application; simple simulation experiments provided a sufficient insight into these sources of variability. However, these models would be worth exploring more rigorously in cases where this form of variability may be more critical.

A more significant source of the variability in experience, as represented by the waiting time distribution, is associated with the date that the patient is added to the list. Changes in the long-term capacity and demand can produce trends in the waiting time while the seasonal effects of holidays or staff changes may generate cyclical variations. The time-dependent variation in capacity and demand for orthopaedic surgery during April-September 2006 is illustrated by the net cumulative increase in the waiting list in Figure 4. This suggests an approximately linear trend, with the total decline of 121 patients equivalent to approximately 2.8 weeks of waiting. While there was substantial monthly variation in supply and demand, the buffer introduced by the delay between being added to the list and treatment reduced the resultant variability in waiting time. Simple linear corrections were made to the waiting times according to the date that each patient was added to the list. This process removed the time-dependent element, producing a modified distribution of waiting times that represents the variations in patients' clinical needs and availabilities.
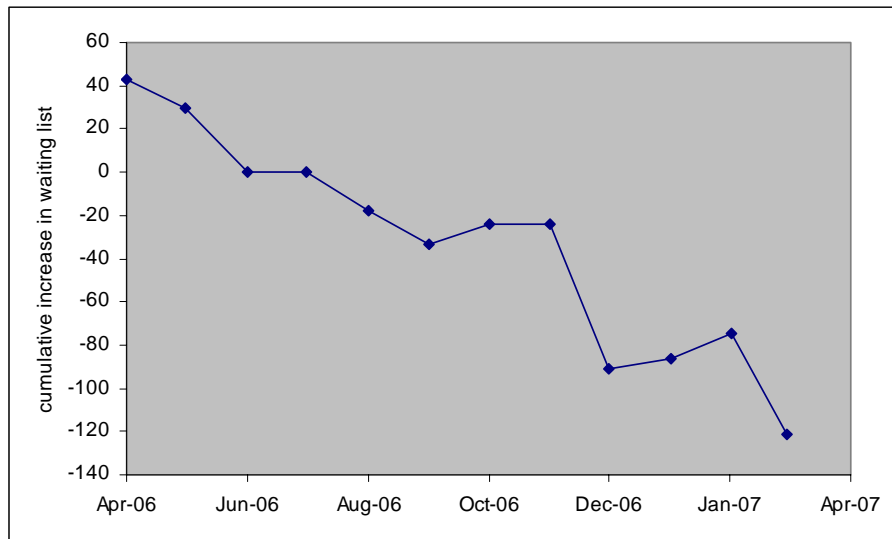
Figure 4 Consequences of time-dependent variation in supply and demand

Alternative models of waiting list behaviour

Waiting list behaviour appears to have varied over recent years. A study of the NHS national waiting time data from 2001/2 for orthopaedic admissions led to the development of a model based on an exponential distribution (Appleby, 2005). It consisted of an exponential baseline combined with a distribution of "excess admissions" associated with the response to waiting time targets. The exponential element of the model implies a world in which the probability of a patient being treated in any given period is constant, with no regard to the length of time spent waiting. This baseline exponential model provided a good fit for much of the empirical distribution; it was only when patients were experiencing waits of more than 45 weeks that their waiting time became a factor in prioritising treatment. Similar behaviour has been observed in other specialties, with the distribution of waiting time being more consistent with a random priority rather than a FIFO discipline. It appears that this behaviour is not unique, either to the United Kingdom NHS or to orthopaedics (Torkki et al, 2002).

Other simulations (The CheckList Partnership, 2007; Vasilakis et al 2007) have adapted the FIFO model distinguishing groups of patients with different priorities. The majority of patients are treated "in turn", employing a strict FIFO discipline. This "in-turn" model classifies a proportion of patients as "urgent" or "soon" who typically have waiting times of less than 8 weeks, and often shorter if the clinical need is high. The majority of the remaining "routine" patients are assumed to be treated in chronological order of their addition to the waiting list, though the model acknowledges that a proportion of the apparently designated "routine" may actually be treated sooner (The CheckList Partnership, 2007). The resulting bimodal model may provide a reasonable approximation if it is applied at a low level, modelling individual consultants' waiting lists and sub-specialties, and it incorporates the possibility of pooling lists. However, it can be difficult to obtain sufficiently detailed data. Much activity in the NHS is classified by the responsible consultant, though the actual treatment may be delivered by a number of different staff. This difficulty in classification is less pronounced with

surgical care than outpatient activity but many accessible data are aggregated by specialty, denying the possibility of readily identifying the bimodal distributions implied by this model. The waiting time distributions in Figure 2 reveal that there have been periods when the bimodal model was relevant for orthopaedics but this has not become a permanent feature, although it is a strong characteristic of ENT, see Figure 3.

The distributions observed in the present study of elective orthopaedics admissions for 2004-7, see Figure 2, did not correspond consistently to either a simple FIFO or an exponential model. The empirical distributions exhibit some characteristics of a FIFO model: distinct modes can be observed in October-March 2004/5 which appear to correspond to the two groups of patients, urgent and routine. However, the long tailed distributions with no apparent modes, as in April-September 2004/5, are more similar to the exponential baseline models observed elsewhere (Appleby, 2005). Examples from other specialties also suggest that the exponential baseline model may have been appropriate in the past but it is not usually relevant now, though the distribution for General Surgery still retains some distinct exponential characteristics, see Figure 3.

Caution is necessary in interpreting the results from a study of just a few specialties in one Health Board but the analysis of the waiting time distributions suggests a coherent narrative. It appears that the traditional approaches to waiting list management paid little regard to the time patients spent on the list, resulting in the exponential distributions observed in earlier studies. One response to waiting time targets is to increase capacity, through employing extra resources or making more efficient use of existing staff and facilities: patient priorities are not changed but everyone is treated more quickly. However, some specialties face greater challenges and have to target any additional resources more carefully resulting in some changes in patient priorities, while respecting clinical need as the most important factor. Such specialties may respond to the waiting list targets by placing a greater emphasis on treating patients "in-turn", resulting in distributions with distinct modes.

Effective management or clinical distortion?

The "in-turn" model is based on one concept of fairness. It acknowledges that there are some patients with particular clinical need resulting in them being classified as "urgent" or "soon" while the remaining patients should be treated in chronological order of the date of their addition to the waiting list. Some may view a higher proportion of patients being treated in-turn, as evidence of more effective management. Others dispute this interpretation and regard an emphasis on avoiding long waits as clinical distortion (Appleby, 2005) and challenge the morality of including any consideration of waiting time in the allocation of health resources (Culyer and Cullis, 1976): consultants should only consider the duration of the wait of a patient in the context of their condition and the reduction their quality of life caused by the delay in receiving treatment. A variety of prioritisation measures have been proposed (Mullen, 2006); some are based purely on immediate clinical need, others consider a wider range of factors influencing the benefit of the treatment and some include an explicit component reflecting the time spent waiting. This debate was reflected in interviews with staff in the specialties studied. The initial reaction of some consultants was critical of the idea of a simplistic "in-turn" model, disputing the idea that this is a fair approach to allocating care. However, it

became apparent that in practice many staff implement policies consistent with the idea of treating patients in-turn whenever possible.

Developing a predictive model of waiting list behaviour

One approach to modelling waiting list behaviour is to include more detail and model each consultant's list separately: the variation in wait might be partly explained by the differing demands for each consultant's specialist services. However, the data in this study did not provide a reliable record for individual consultants. Even if such data were available, modelling each consultant's work as completely independent activities is unrealistic. In most specialties there are some common activities which can be undertaken by a variety of staff. Although it might be important that patients with particular conditions are directed to staff with very specific skills, there is usually a substantial common element that can be allocated to a range of staff to balance the workloads across the specialty. Similarly, resources may be redirected to match demand; some clinic or theatre time may be reallocated to the staff with longer waiting lists. Exploiting such flexibility depends on the staff and local management but most specialties are reasonably flexible and take actions to avoid excessive waits where possible. Explicitly modelling this transfer of resources, or demand, is problematic and a simpler approximation is to adopt the specialty as the unit of analysis.

A model explicitly distinguishing between "urgent", "soon" and "routine" categories of patients was also considered. However, discussions with staff suggested that this distinction was subject to substantial variability, both between staff and over time. It was argued that patients' needs cannot always be categorised so simply and there is a need for a more subtle approach. A simple model was developed using a modified normalised empirical distribution, with the time-dependent variation removed. Observations of the changes in waiting time distributions, as in Figure 2, suggest that, in the short or medium term, the priorities of patients are invariant. Patients are seen in the same order but the actual waiting time is dependent on the capacity relative to the demand. This was modelled by assuming that waiting time distribution maintains a constant form but is subject to a simple linear transformation reflecting the changing mean wait. If $w(t)$ was the observed distribution of waiting times with mean μ, this model assumes an estimate of the future distribution $w'(t)$, mean μ', where:

$$w'(t) = \frac{\mu}{\mu'} w\left(\frac{\mu}{\mu'} t\right) \qquad\qquad (1)$$

This approach is an example of "no explicit rationing of access to the waiting list, but clinical priority determines the speed at which patients move through the list and are admitted to hospital" (Goddard and Tavakoli, 2008). The model appears to offer a reasonable basis for projecting the waiting experience, at least in the short or medium term. As an example, the empirical distribution of April 2006 – September 2006 was used as $w(t)$ to derive $w'(t)$ for orthopaedic surgery for October 2006 – March 2007. The P-P plot of Figure 5 compares the observed empirical cumulative probability waiting time distribution with that implied by the waiting list model, suggesting a reasonable goodness of fit. This is despite a significant change in the mean waiting time during the period, see Figure 1. Similar results were obtained when applying the

same approach in other specialties, even though there were substantial differences in the form of their distributions, see Figure 3. The model assumes that similar patient selection policies are applied; the use of such a model during periods of dramatic change would not be valid.
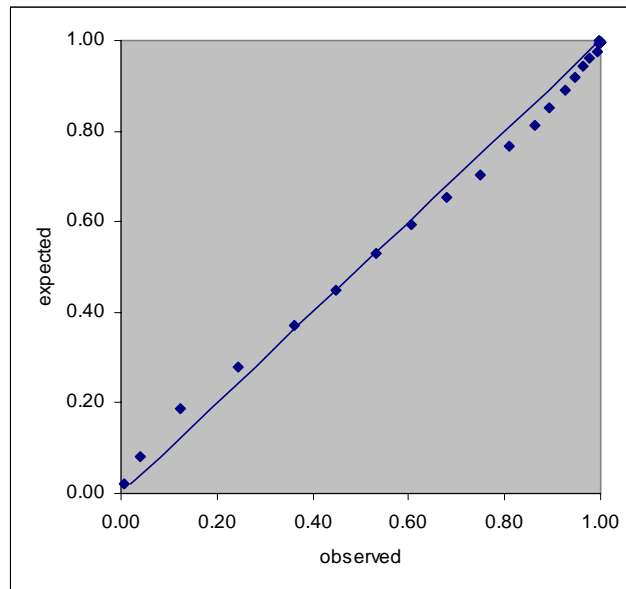

Figure 5 P-P plot comparing observed and expected distributions of waiting times for October 2006 – March 2007

Modelling the appointment process

In practice, new appointments are made based on an assessment of clinical need, taking into account the patient's personal circumstances and availability, combined with a review of the vacancies in the schedule. When the waiting list is short and capacity and demand well matched, even a patient with a low priority may be offered an early appointment. This balancing of priority and the waiting time was modelled by considering:

r_i rank of patient to be added to the queue for stage i reflecting clinical need and availability; this was specified as a percentile (0% indicating the minimum possible need, 100% the maximum need) randomly sampled from a uniform distribution; $r_{outpatient}$, $r_{diagnostics}$, $r_{surgery}$ may be correlated.

t_i waiting time for appointment at stage i.

$W_i'(t_i)$ cumulative waiting time distribution for stage i; this varies depending on the magnitude of the queue and hence the mean wait, see equation (1).

$$W_i{'}(t_i) = \int_0^{t_i} w_i{'}(t)\mathrm{dt} \quad = \quad \int_0^{t_i} \frac{\mu_i}{\mu_i{'}} w_i\!\left(\frac{\mu_i}{\mu_i{'}}t\right)\mathrm{dt} \qquad (2)$$

$W_i^{'-1}(r_i)$ inverse cumulative waiting time distribution for stage i; this is used to determine the predicted waiting time for a patient of rank r_i.

$d_{i,0}$ date/ time added to queue for stage i.

$d_{i,1}$ predicted date/ time that stage i of care will be delivered, i.e. predicted appointment time given the ranking of the patient's need, where

$$d_{i,1} = d_{i,0} + W_i^{'-1}(r_i) \qquad (3)$$

The actual appointment time in the simulation will depend on local queuing conditions caused by short term variations in supply and demand. The predicted date/ time $d_{i,1}$ was used to position the patient in the waiting list, translating the patient's rank into weeks' waiting; this provides a mechanism for relating clinical need and time spent waiting, ensuring that the patients progress at the appropriate "speed" (Goddard and Tavakoli, 2008).

Incorporating the waiting priorities in a capacity planning model

A simulation of elective orthopaedics care was developed using Simul8 (Simul8 Corporation, Glasgow, UK) to examine the problems of meeting the 18-week target for the complete patient journey, from referral to treatment. The model considered referrals and the routing of patients through various diagnostics, with some eventually requiring surgery; Figure 6 provides a summary of the process. The capacity of each component of the process may vary, resulting in different waits depending on the sequence followed by each patient. The simulation incorporated waiting time models at each stage: the ranking, r_i, of a patient was sampled randomly and interpreted in terms of the predicted appointment date $d_{i,1}$ using a separate distribution $w_i(t)$ for each component (outpatients, ultrasound, CT scan, MRI and surgery). The mean wait at each stage was continually monitored such that the distributions could be scaled to update $w'_i(t)$ throughout the simulation period.
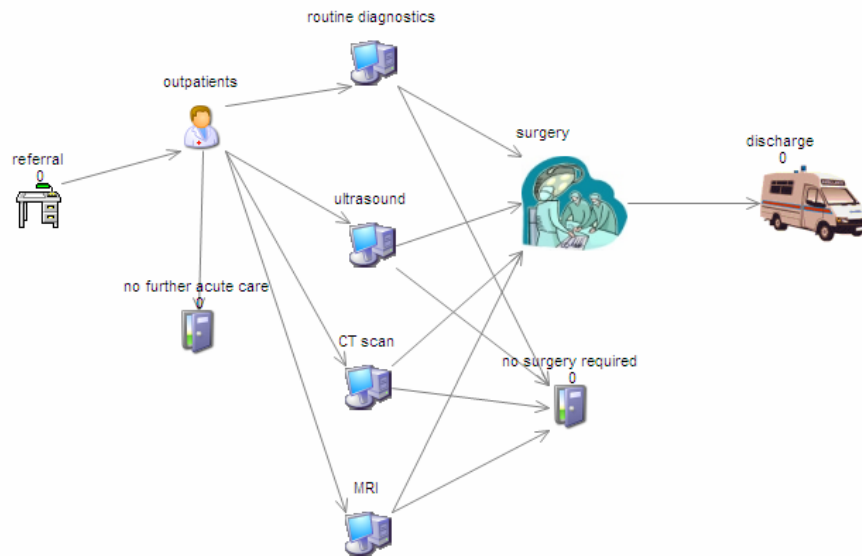
Figure 6 Simulating the complete patient journey

The waiting list model included the correlations between patients' rankings and hence waits, at each stage: a patient who has a high priority when being allocated an outpatients' appointment may well have a high priority at diagnostics and surgery. However, the relationship is far from perfect and there are many cases in which initial fears are not confirmed by the diagnostic evidence, resulting in a reduction in the priority attributed to the patient. The current information systems are disjoint and it was not possible to obtain a complete record of the transit of patients through outpatients, diagnostics and surgery. This should be resolved in time to enable effective monitoring to achieve the 18-week target but the current study had to rely on smaller samples to estimate the correlations of waiting, and hence priority, at each stage.

The model included the historic conversion rates, dictating the proportions of patients following the various possible routes through the elective orthopaedic system. It was used to explore basic requirements, such as the recurrent capacity needed to meet demand at each stage and also the effects of one-off changes; special waiting list initiatives can have substantial downstream consequences and it is important that sufficient capacity is provided to ensure the flow of additional patients through the whole system. However, having a more sophisticated model of waiting list behaviour enabled more subtle analyses, examining the extreme patient experiences rather than the mean waits.

One such analysis examined the appropriate targets for each component of the patient journey: given expected mean waits for each component, the consequences for the whole patient journey were analysed. A number of scenarios were explored using the simulation and the results summarised in Table 2. Scenario A corresponds to the current position: while the individual 2006/7 targets for outpatients and surgery have been met, see Table 1, the complete wait far exceeds the 18-week target: the mean is 23.7 weeks and only 32% of patients wait less than 18 weeks. It was suggested that the outpatient waiting list in particular might be reduced relatively easily; some reductions in surgery and MRI waiting time were also thought achievable, as in scenario B. The simulation analysis suggests that such changes would produce a mean total wait of 17.3

weeks but many patients will wait considerably longer. Scenarios C, D and E represent more challenging sets of component target waits: some would require significant investment in special waiting list initiative sessions. Only scenario E appears to provide the possibility of meeting the 18-week target and given the timescale, see Table 1, it appears an achievable target. This analysis assumed that the referrals could be managed; if shorter waiting lists result in more demand the targets would be more difficult to achieve.

scenario	mean weeks wait					complete wait (weeks)		wait < 18 weeks	
	out-patients	ultrasound	CT	MRI	surgery	mean	95%	mean	95%
A	12.3	4.0	3.1	6.1	7.9	23.7	±0.4	32%	±0.6%
B	8.0	3.0	3.0	4.0	7.0	17.3	±0.4	50%	±1.8%
C	7.0	3.0	3.0	4.0	6.0	15.4	±0.4	60%	±2.8%
D	5.0	3.0	3.0	3.0	6.0	12.8	±0.5	79%	±4.1%
E	4.0	3.0	3.0	3.0	5.0	10.8	±0.5	92%	±2.5%

Table 2 Simulation experiments exploring target mean waits

Conclusions

The challenge of meeting the NHS waiting time targets will require a variety of measures: investment in additional resources and also changing working practice. Simulation models can be useful in assessing the likely impact of proposed changes and investment. This requires an understanding of the waiting list behaviour and the prioritisation of patients: FIFO is not a dominant principle in allocating healthcare in the NHS. While balancing clinical need and a desire to meet the waiting time targets requires care, the changing distributions of waiting times in recent years suggest that in some specialties the additional resources have been directed at achieving the NHS targets. However, in the short or medium term, the balance of priorities and the resultant form of the waiting list distribution are approximately constant, providing a basis for modelling future waiting lists.

A model of waiting list behaviour was incorporated into a simulation of elective orthopaedic care. Analysis of a variety of scenarios provided insights into the effects of inter-relationships between waiting at the three key stages of care (outpatient consultation, diagnostics and surgery) and helped identify target waits for each stage. The model and the analyses of the various scenarios were presented to NHS Fife staff as part of a larger project investigating the reduction of waiting times. The study helped in the development of a better understanding of the components of the total wait and to determine information needs to manage the achievement of the new 18-week target.

Further research could be valuable, exploring possible standard functional distributions as models of the variety of waiting time experiences. Also, the proposed approach could benefit from a more rigorous treatment of the separation of variation due to patient priority and that generated by the time-dependent nature of supply and demand.

References

Appleby J (2005). *Sustaining Reductions in Waiting Times: Identifying successful strategies*, King's Fund.
http://www.kingsfund.org.uk/publications/kings_fund_publications/sustaining.html, accessed 13 August 2008.

Audit Scotland (2006). *Tackling waiting times in the NHS in Scotland*.
http://www.audit-scotland.gov.uk/docs/health/2005/nr_060216_waiting_times_km.pdf, accessed 13 August 2008.
Bevan G and Hood C (2006) Have targets improved performance in the English NHS? *BMJ* **332**: 419-422.
Cameron R, Dugdale RE and Page MJ (2006). A Capacity Planning Simulation Model and its Application to a Nuclear Medicine Service) ESM'2006, LAAS, Toulouse, France. http://lattice-net.co.uk/ESM2006_BIO-04_Cameron_etal.doc, accessed 24 July 2008.
Carvel J (2007). Public sector targets to be scrapped, *The Guardian* 18 July 2007. Available at http://www.guardian.co.uk/uk/2007/jul/18/politics.topstories3, accessed 13 August 2008.
The CheckList Partnership (2007). *Modelling the 18 week wait: a practical guide for CheckList users*. http://www.checklist.co.uk/h_18.htm, accessed 13 August 2008.
Culyer AJ and Cullis JG (1976). Some economics of hospital waiting lists in the NHS. *J Social Policy* **5**: 239–264.
Goddard J and Tavakoli M (2008). Efficiency and welfare implications of managed public sector hospital waiting lists, *Eur J Oper Res* **184**: 778-792.
ISD Scotland (2008). *Waiting Times Statistical Publication Notice*.
http://www.isdscotland.org/isd/5413.html, accessed 24 July 2008.
Jun JB, Jacobson SH and Swisher JR (1999). Application of discrete-event simulation in health care clinics: a survey. *J Oper Res Soc* **50**: 109-123.
Moss L (2007). Troubleshoot teams sent in to cut cancer waiting times, *The Scotsman*, 30 May 2007. Available at http://news.scotsman.com/healthofthenhs/Troubleshoot-teams-sent-in-to.3290322.jp, accessed 13 August 2008.

Mullen P (2003). Prioritising Waiting Lists: How and Why? *Eur J Oper Res* **150**: 32-45.

NHS Scotland (2008). *18 Weeks the Referral to Treatment Standard*.
http://www.18weeks.scot.nhs.uk/, accessed 24 July 2008.

Ouwens J, Groen H, TenVergt E, Koërter G, Boer W and van der Bij W and the National Steering Group for Lung Transplantation (2002). Simulated waiting list prioritization for equitable allocation of donor lungs. *J of Heart and Lung Transplantation*, **21**:797-803.

Proudlove NC, Black S, Fletcher A (2007) OR and the challenge to improve the NHS: modelling for insight and improvement in in-patient flows. *J Oper Res Soc* **58**: 145-158.

Ratcliffe J, Young T, Buxton M, Eldabi T, Paul R, Burroughs A, Papatheodoridis G and Rolles K (2001) A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation. *Health Care Management Science* **4**: 117-124.

Steyn R (2008) *Improving Patient Flow*. http://www.steyn.org.uk/, accessed 24 July 2008.

Torkki M, Linna M, Seitsalo S and Paavolainen P (2002). How to report and monitor the performance of waiting list management, *Int J Technology Assessment Health Care* **18**: 611-618.

Vasilakis C, Sobolev BG, Kuramoto L and Levy AR (2007) A simulation study of scheduling clinic appointments in surgical care: individual surgeon versus pooled lists, *J Oper Res Soc* **58**: 202-211.

Wall A and Worthington DJ (2007). Time-dependent analysis of virtual waiting time behaviour in discrete time queues, *Eur J Oper Res* **178**: 482-499.

Worthington DJ (1987). Queuing models for hospital waiting lists. *J Oper Res Soc* **38**: 413-422.