Looking in the wrong place for healthcare improvements:

A system dynamics study of an accident and emergency department

DC Lane, C Monefeldt and JV Rosenhead - *The London School of Economics and Political Science*

Accident and Emergency units provide a route for patients seeking urgent admission to acute hospitals. Public concern over long waiting times for admissions motivated this study, whose aim was to explore the factors which contribute to such delays. In collaboration with a major London teaching hospital, a system dynamics model of the interaction of demand pattern, A&E resource deployment, other hospital processes and bed numbers was developed. The paper discusses the formulation of this model; the calibration of a Base Case simulation; and the outputs of policy analysis runs of the model which vary a number of the key parameters. Two significant findings appear to have policy implications. One is that while some delays to patients are unavoidable, reductions can be achieved by selective augmentation of resources within, and relating to, the A&E unit. The second is that reductions in bed numbers do not increase waiting times for emergency admissions, their effect instead being to increase sharply the number of cancellations of admissions for elective surgery. This suggests that basing A&E policy solely on any single criterion will succeed in transferring the effects of a resource deficit to a different patient group.

Introduction

Accident and Emergency and the crisis in British health care

The National Health Service (NHS), a treasured British institution, has been in semi-permanent crisis during two decades of government restraints on public sector expenditure. Despite the election of a new government in May 1997, problems endure (*Independent*, 16.8.97). Among the indicators of crisis have been repeated reorganisations, closure of facilities, lengthy waiting lists, cancellation of scheduled admissions to hospital, and depletion of budgets before the year end leading to curtailment of activity. In recent years public and political concern has focused in particular on the performance of Accident and Emergency departments at acute hospitals.

Accident and Emergency (A&E) provides access to hospital services for urgent cases. The normal process for admission to hospital is via referral to a hospital consultant in an appropriate specialty, with the general practitioner (GP) playing a ‘gatekeeper’ role. However a parallel route is necessary for emergencies. Consequently, the A&E (or casualty) department is used by individuals brought by ambulance and or presenting themselves for treatment. The latter, in particular, include people whose medical conditions vary widely in severity. A&E therefore performs a sorting function, deals itself with a range of less seriously ill patients and assesses more serious cases for admission as hospital inpatients. Only some 15 to 20% of patients arriving at A&E are eventually admitted to beds on hospital wards.[1]

Public concern has focused on what are widely felt to have been excessively long waiting periods at A&E. Official guidelines specify that patients requiring treatment as inpatients should be allocated to a bed within two hours.[2] Although these standards apply not to the duration of wait from the individual’s arrival at A&E but only to the time elapsed after a decision to admit, they are nevertheless routinely broken in many hospitals. The spectacle of sick people whose condition merited admission as inpatients having to wait overnight or longer on trolleys, or even being treated in ambulances parked outside the unit (*Guardian*, 12.1.96), has provoked widespread criticism. At times A&E units have become so congested that they have been 'closed to blue lights' - even emergency ambulances are barred and are diverted to other hospitals (*Evening Standard*, 17.12.96).

A variety of explanations has been offered for the crisis affecting A&E, a recent report providing a comprehensive account of possible factors. [3] A commonly held view attributes it to the closure of hospital beds and the consequent reduction in hospital bed capacity. The British Medical Association estimated that 9000 acute beds were closed in England 1991-5; and the Labour Party (when in opposition) produced figures of 13,000 acute bed closures between 1989/90 and 1995/6 - some 11% of the total (*Guardian* 12.1.96, 14.1.97). Shortages of beds, in this view, have multiple effects on A&E. Directly, patients arriving via A&E find that there is no bed available. Indirectly, bed shortages cause cancellations of scheduled non-emergency admissions; as a result of delay, some of these become emergency cases, necessitating entry through A&E. Perhaps, it has been suggested, the surge of attendances at A&E (up from 3,260,000 in London financial year 1992-93 to over 3,825,000 in 1996-97[4]) has been a behavioural response by patients and their GPs to the difficulty of gaining admission through the referral procedure. This rise in the number of emergency admissions has occurred at the same time as bed numbers have been falling, and as difficulties of A&E staff in admitting emergency patients to hospital beds has been growing.[1]

Other factors possibly implicated in the A&E crisis are beds ‘blocked’ by patients fit for discharge because local government lacks community care resources to receive them; and government waiting list reduction initiatives. The latter are targeted at the tail of those who have been waiting longest - the effect can be to give precedence to less serious cases, dislodging more urgent ones which may then require emergency admission (*Daily Telegraph* 12.1.96). Another plausible mechanism is that competition on price promoted since 1990 by the internal market in health services has increased bed occupancy rates to a level where there is inadequate slack left to cope with demand variations. All of these explanations implicate one or more aspects of the then government's policies on the NHS, or on public expenditure more generally, in the deterioration of A&E performance. The governmental response was, typically, to urge hospitals to manage their beds more effectively, and to permit hospitals to employ doctors in A&E department beyond agreed quotas (*ibid.*).

Origins of the present study

'Casualty Watch' is a project established in 1992 as a response to public concern that cuts in the NHS, especially bed closures, were producing an inadequate casualty service and harming patients. It was set up and is still run by Southwark Community Health Council (in South London). Initially sponsored by the Greater London Association of Community Health Councils, it was subsequently funded by Lambeth, Southwark and Lewisham Health Authority and now by the South Thames Association of CHCs. (Community Health Councils, or CHCs, are statutory bodies set up to represent the community view on local health service provision.) Casualty Watch monitors the performance of hospital casualty departments throughout London, as well as some further afield. It works by volunteers making simultaneous co-ordinated monthly visits to A&E units and collecting data, both by observation and from the hospital information systems, on the numbers and waiting times to date of patients currently in A&E. These surveys enable the trend in performance to be tracked, and publicised if appropriate.

The Casualty Watch survey, however, did not attempt to analyse the causes of the delays and congestion in the system which it recorded. This was perceived as a limitation, and collaboration between the project and the Department of Operational Research at the London School of Economics in 1995 resulted in a joint effort to build a computer model to explore how factors internal and external to A&E contribute to delays in casualty. The modelling expertise was provided by LSE, while orientation to the problem area, judgements on design choices, and introductions to stakeholders were supplied by Casualty Watch.

System dynamics was selected as the appropriate modelling medium, and the model was developed through two MSc student projects in 1995 and 1996. In the first of these a prototype model was produced, which in the second project was piloted in an outer London A&E department. From October 1996 a nine month project was established, funded partly by Casualty Watch and partly by LSE, to turn the prototype into a fully operational model. During this period a close working relationship was established with the A&E department of an inner London teaching hospital - which for reasons of confidentiality we will refer to as 'St. Dane's'. With access to information held by staff and the hospital database, the model has been calibrated to fit St. Dane's, the relationships within it have been confirmed and model parameters have been estimated. This model has been used to investigate the sensitivity of waits at A&E to changes in parameter values, resource levels and allocation policies.

Contents of this paper

This paper reports on the structure of the A&E system dynamics model and on its use to explore the dynamics of the system of which A&E is a part. The second section outlines the principal features of an A&E department, justifies the choice of system dynamics, proposes a focus for the model and then describes its structure. The following section presents an analysis of the outputs of a 'Base Case' simulation - including elaboration of some noteworthy aspects of the model formulation - and addresses the question of model validity. In the penultimate section this Base Case is compared with a number of interesting scenarios, notably involving changes in bed capacity but also incorporating changes in the pattern of demand. A final section derives lessons from these simulation experiments, draws some general policy conclusions, and identifies further work to be performed.

Model Conceptualisation and Formulation

In this section we first describe the activities in and around an A&E department in a way suitable for modelling. We then justify the use of system dynamics as an appropriate approach, present the focus of our modelling study and give an overview of the model formulation.

Initial Conceptualisation of the A&E System

It is eminently possible to describe the activities of an A&E department, together with its interaction with the environment, in considerable detail. However, we have the specific aim of exploring the issue raised in the introduction: the response of A&E waiting times to reductions in bed capacity. For the purposes of modelling that issue we can focus our description. We can therefore conceptualise the A&E system in terms of two areas: the community and the hospital, with the relevant functions of the hospital sub-divided into three; the A&E department, the management of elective patients, and the wards. These can be briefly characterised as follows:

The Community. Patients flow into the hospital from the surrounding community and are subsequently discharged back into the community. There are two main patient groups: emergency patients and elective patients. Emergency patients are individuals who have been brought by ambulance, as well as 'walk in' patients who have presented themselves at the A&E department. The latter group includes 'self-referrals'; individuals whose GP has not secured a direct referral to a consultant and who have therefore chosen to enter A&E without prior arrangement. The rate at which emergency patients arrive in the A&E department depends on many characteristics of the community: density and average age of the population, accessibility to other hospitals etc. Elective patients are individuals requiring non-urgent medical treatments who stay on a waiting list until scheduled for treatment. The rate of scheduled elective admissions depends on the capability of the hospital (scale of facilities, range of services offered etc.) and the characteristics of the catchment area. The reasons for variations in these two arrival patterns is outside the scope of this study; arrival rates of patients to casualty are best considered as exogenous to any model. For both of these rates we therefore aimed to use historical data when modelling.

The rate at which patients are discharged back into the community influences hospital occupancy and is itself related to length of stay. Length of stay is influenced by a variety of factors: individual characteristics of the patient, the range and quality of hospital services, the availability of care in the community[5-8] etc. The various factors affecting the length of time patients stay in hospital are not distinguished in the model but rather aggregated into a single value for this parameter.

The Hospital - A&E Department. Many elements make up the time which elapses between a patient arriving in A&E and, if in-patient care is decided upon, subsequently leaving the department to go to a bed on one of the wards (Figure 1). Patients follow different pathways through A&E before they are eventually admitted (if indeed they are). When patients arrive in A&E they are triaged and registered and then wait for an initial consultation with an 'A&E doctor' (we use this term to designate collectively Senior House Officers - the lowest post-registration medical training grade - and Registrars in A&E). The number of A&E doctors on duty varies during a day. Patients may be treated and discharged, or might be the subject of further clinical appraisal and testing prior to treatment and discharge. More severe cases are referred to a 'Specialty Doctor', a member of a specialty support team called to A&E from elsewhere in the hospital. Such patients may undergo further test procedures and then be treated and discharged. Alternatively, they may be admitted to a ward and subsequently leave the hospital after completing treatment as inpatients. The main procedures necessary to determine whether hospital admission is required for emergency patients, and the resources required to perform these procedures, are shown in Figure 1.

The Hospital - Management of Elective Patients. Treatments for which elective patients are waiting can be 'surgical', e.g. hip replacement, or 'medical', e.g. chemotherapy. To ensure that operating theatres, surgical teams or appropriate equipments are available, such treatments are scheduled well in advance. Patients are scheduled for admission to wards the day before their treatment and 'prepped' (e.g. blood tests, ECG readings, 'starving' in preparating for the reception of drugs, etc.) by nurses.

On their scheduled day of admission patients are asked to contact the hospital by telephone to confirm their appointment. If confirmation is received then they are asked to travel to the hospital. Alternatively, they may be told that no bed is available. In this case their admission may be cancelled or they may be asked to wait at home in the hope that ward rounds later in the day may unexpectedly free-up a bed. A small proportion of patients are unable to keep their appointment, or cancel it themselves, and they lose their scheduled slot.

Elective patients who are allocated a bed are admitted onto a ward, perhaps after delay in the 'day room'. This is true for those whose admission is confirmed early and for those found a bed only later in the day. Consequently, these elective patients arrive on the wards throughout the day, though rarely reaching them later than 18:00. They are prepped and made ready for treatment the next day.

Elective patients who fail to have a bed allocated to them may have to be cancelled (and rescheduled for another date). However, St Dane's has a 'hotel unit' in which some may be asked to stay overnight. On the following day these patients must rapidly be found a bed and then prepped in good time for treatment that day. This option is labour intensive and only a small proportion of patients can be handled in this.

The Hospital - Wards. The rate at which a hospital can admit new patients - whether emergencies or electives - is influenced by the management and staffing of beds. Because lack of available beds on wards may contribute to delays in an A&E department, a model must incorporate information about bed occupancy - the status of beds on the wards. Occupancy is determined by the ratio of the accumulated number of emergency or elective patients on the wards to the total number of hospital beds. As occupancy rises, increasing priority is given to admitting onto the wards patients from A&E whilst scheduled elective treatments increasingly need to be cancelled.

Patients are discharged from wards during normal opening hours. In principal, this frees a bed. However, both patient and bed must first be identified, cleaned and got ready. This all takes nursing time and nurses have other duties; during daylight hours current patients are more active and need attending to, meals and drugs must be distributed and visitors assisted. In consequence, emptied beds do not immediately become available. This 'turnover interval' introduces a delay between patient discharge and new patient admission.

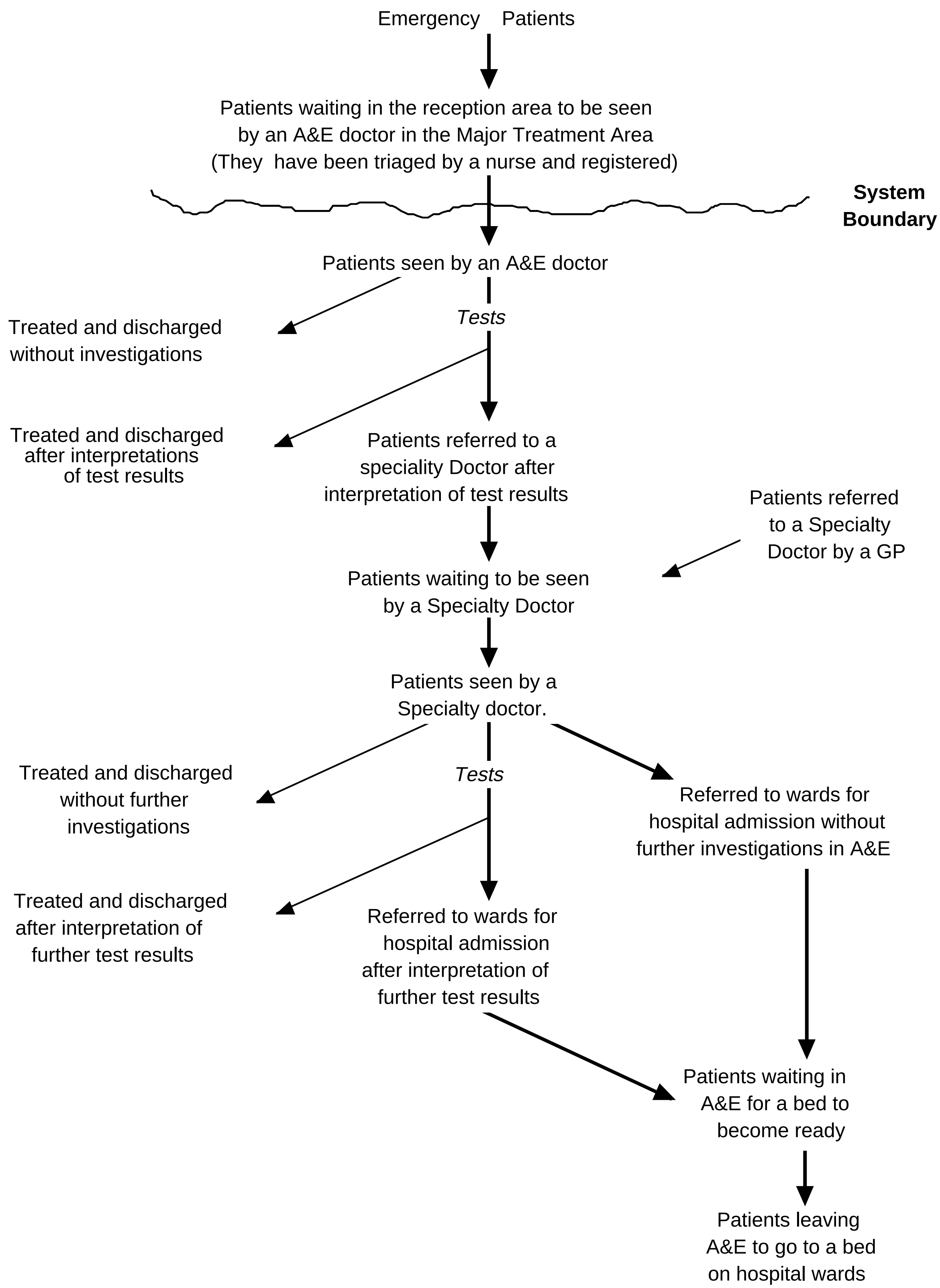

Figure 1 Schematic representation of A&E elements, processes and pathways included in the system dynamics model.

N.B. Edited in this shortened paper.*

Various healthcare systems have been the object of previous simulation studies. Here we briefly review the relevant work and the show the relevance of a study employing the system dynamics approach.

The culture of the NHS leads understandably to a focus on the handling of individual patients and this generally predominates in studies of A&E,[9-11] including those touching on waiting times.[12,13] It is therefore not surprising that operational researchers have frequently advocated the use of a discrete event simulation (DES) approach, which can generate detailed results concerning the handling of different types of patients.[14-16] However, a consequence of considering situations at a micro-level has been studies whose focus has been on isolated areas within a hospital[17-20] or on the treatment histories of specific types of patients.[21-24] System dynamics modelling[25] can play a different role in probing the operation of healthcare systems. A general discussion of this role may be found elsewhere.[26]

Model Focus: The Dynamic Hypothesis

In this section the main feedback processes that influence an A&E department are presented. An hypothesis concerning the resulting dynamic behaviour is also discussed. Together these two elements provide a focus for the study.

A system dynamics model is a causal theory of how an observed or desired behaviour is generated by a social system.[25,39-41] Such models need a clear focus and the starting point for this is a 'dynamic hypothesis'. This is a verbal and/or diagrammatic description of a feedback structure, combined with a 'reference mode', or description of the observed or desired system behaviour that the structure is thought to generate.[42] A formulated and calibrated model can test this dynamic hypothesis. The dynamic hypothesis described below served as a guide for model construction and for the policy experiments described later.

The structural causes of waiting times are complex and no two A&E, or Casualty departments will have exactly the same set of problems.[1] However, the nature of the main factors is clear. There are two main patient groups (emergency admissions and elective admissions) which interact to influence the availability of beds on the wards. The principal interactions are shown in Figure 2 as a causal loop diagram.[43]

The main feedback loops are all balancing loops, that is, they are explicitly or implicitly goal-seeking, so that a change in the value of one of the variables in a loops tends to be counter-acted by the operation of that loop. (The model does also contain two reinforcing loops, representing crowding effects but these are of less significance and are not shown here)

Loop B1 acts to drain the number of emergency patients in A&E by admitting them to hospital wards. As patients awaiting admission increase so the rate of admissions increases, assuming there is free bed capacity on wards. The total waiting time for a patient actually results from the separate delays involved in the various activities within A&E. Most of the detail of the model therefore resides in the dissaggregation of loop B1 into the separate activities described in the previous section (Figure 1).

Loop B2 acts to limit the occupancy level on wards by the bed capacity. By restricting - and even shutting down - emergency admissions, loop B2 ensures that patients are not admitted unless there is free capacity.

Loop B3 has the same goal of controlling bed occupancy but acts by influencing the rate of elective admissions. Again, bed capacity is a limiting variable.

Loop B3 also controls loops B4 whose two components together share the goal of reducing the backlog of scheduled elective patients. This is achieved either by admission to wards (loop B4a) or by cancellation (loop B4b). If there is room to accommodate all of the elective scheduled admissions then the elective admission rate will equal the desired admission rate (B4a), otherwise, a greater burden falls on B4b. The balance between B4a and B4b is itself controlled by the balance between B2 and B3 and it is important to note that the operation of B2 has priority over the operation of B3.

The previous government's claim that the quality of healthcare would not be affected by acute hospital bed closures indicates a desired behaviour of the above system.[44-46] The case might run as follows: decreasing the bed capacity increases the occupancy level of beds on the wards. This increased utilisation of beds makes up for the beds lost so that the hospital can accommodate the same number of patients as before. Consequently the emergency admission rate - and hence patients' waiting times in casualty - remains unchanged. Or, in reference mode terms: a downward step in bed capacity yields upward adjustment in occupancy and no change in waiting times.

We therefore have a system structure (Figure 2) and the desired behaviour described above. Together, these form the dynamic hypothesis that is to be tested.

---

* This plenary paper is a shortened version of an LSE Working Paper. This section has therefore been reduced in length. For simplicity and consistency, the references relating to it (references 27 to 38) have been retained. A later section has been removed, though its location is still indicated.

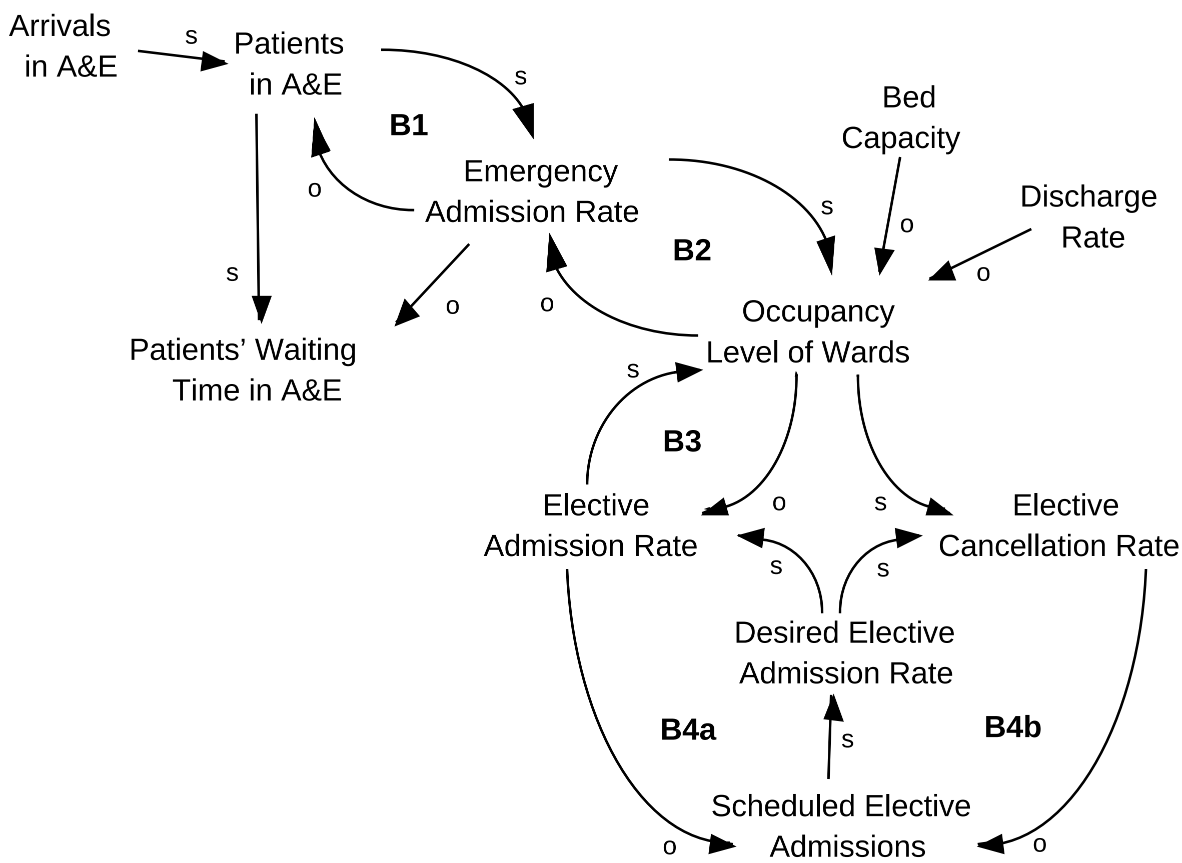

Figure 2 Causal loop diagram of the main effects determining waiting times in an A&E department.

Formulation of Model Structure and Equations

The system structure described above has been formulated as a quantitative model, the purpose being to simulate the dynamic behaviour implied by the model's assumptions.[25,39,47] The key elements of formulation are outlined here.

The most detailed segment of the model concerns the A&E department itself. An exhaustive account of these elements of the model is not appropriate here. However, all of the processes described above are represented, so that the features of this sub-section of the model are as shown schematically in Figure 1.

The model also incorporates the handling of elective patients. The detail of this sub-system is as conceptualised in the section above on the management of elective patients, the map in Figure 3 being only a simplified representation of the detailed formulation in the model. As patients are scheduled, they are modelled as flowing into a stock ('Scheduled Elective Admissions'). Individuals waiting at home, travelling to the hospital and those who have stayed overnight in the hotel unit are all included in this stock. Some patients are unable to keep their appointment ('Drop Out Rate'). Others have their treatment cancelled because the hospital does not have beds available to admit them ('Elective Cancellation Rate'). The remainder are admitted onto a ward for treatment ('Actual Elective Admission Rate'). Admissions and cancellations both on the day before treatment and on the actual day of treatment are brought together in these two flows.

The model was constructed using the iThink software,[28,29] on the Macintosh platform. With this package a multi-order, non-linear ordinary differential equation model may be built which presents its main assumptions to the user via a graphical interface, or diagram. The model was large; the core had nine stocks and 160 other variables whilst nine stocks and 16 other variables were used to calculate various performance measures. To handle the complexity of the resulting diagram, the detailed representations of the activities in A&E and of those concerned with elective admissions were built as sub-models.[48] Although the software allows all of this structure to be displayed graphically, it was overlaid with a simplified, high level map (Figure 3). This format allows the systemic interactions of the model to be effectively presented. For example, demand for A&E service can be seen to depend on both emergency patients and on patients referred by their GP. Some patients are discharged from A&E but those requiring admission go on to interact with scheduled elective admissions, the outcome influencing both cancellations and patients on wards.

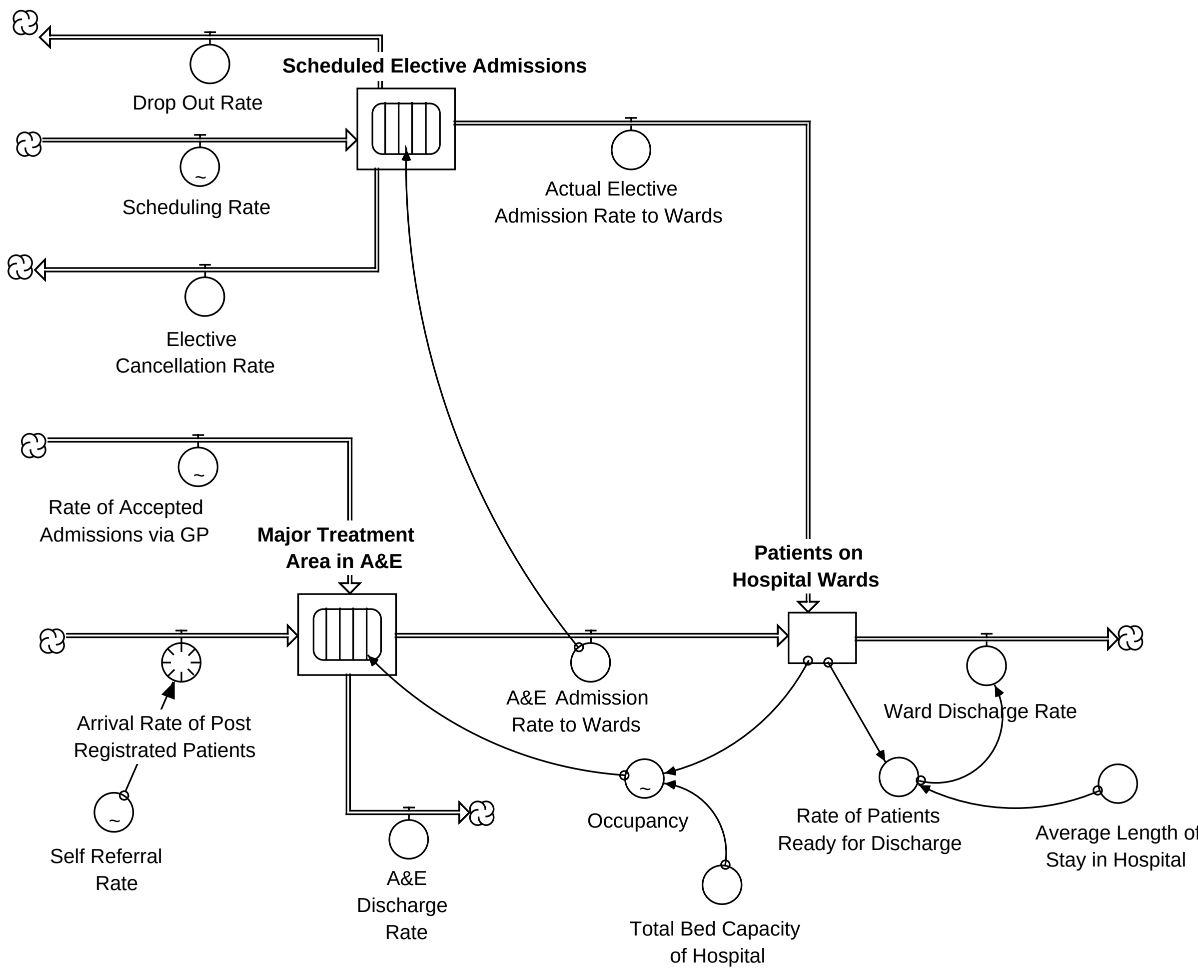

Figure 3 Stock/flow diagram of the A&E system dynamics model. Only the high level map is shown: the two boxes with vertical grills open to reveal the full, 194 equation model.

The Base Case Simulation: calibration, analysis and model validation

In this section the focus moves to simulation output. First the calibration of the two patient arrival rates is described. We then report on how the completed simulation model was used to generate a Base Case, what summary statistics were recorded and of the broad-brush analysis of the system that resulted. In the following two sub-sections we elaborate on some points of formulation necessary to understand adequately the functioning of the model. The section closes with comments on the validity of the model.

Arrival rate calibration

The model treats the arrival of emergency and elective patients into the A&E department as exogenous. The calibration of these two rates is therefore crucial to the model's realism. Below we describe the sources of data used to estimate them.

Emergency admissions. In the most recent year nearly 8,000 patients arrived without pre-arrangement at the St. Dane's A&E department. This total was known to be unevenly distributed, depending on the time of day and day of week. However, at the time of this study the hospital computer system had not been used to record hourly data on these emergency patients, only monthly totals being available. Fortunately, the recently arrived Registrar had started to compile more detailed information (in order to examine the appropriateness of the A&E doctor rosters then in place). We were therefore able to obtain written archival data which recorded A&E arrivals at hourly intervals for the period September to November 1995 (Figure 4). This data was incorporated in the model after averaging over days of the week (since day to day variation is dominated by within-day variation) and adjusting upwards by 7% to reflect annual increase in usage. The result was a cycle which produced a total of 220 emergency patients per day, of whom about 40 would require hospital admission and therefore a bed.

Scheduling of Elective Admissions. Data for the average rate of scheduling elective patients for admission was obtained from the hospital Bed Manager. He estimated that each day 110 waiting list patients were scheduled for admission during the normal hours of 9:00 to 17:00, for treatment the following day. In the absence of more detailed information this total was assumed to be distributed evenly over that period.

Both emergency and elective admissions make use of the finite bed capacity, taken to be 800 beds in the Base Case simulation. The relative priorities afforded to A&E patients awaiting admission and patients scheduled for elective admission are affected by the extent of the backlog of each category of patients. The flow of admissions onto wards accounts for the relative priorities so that, in practice, all other things being equal, as the backlog of A&E patients rises, the free beds are increasingly allocated to A&E patients and more scheduled elective admissions are cancelled. This policy was confirmed by all those we spoke with and was therefore represented in the model.

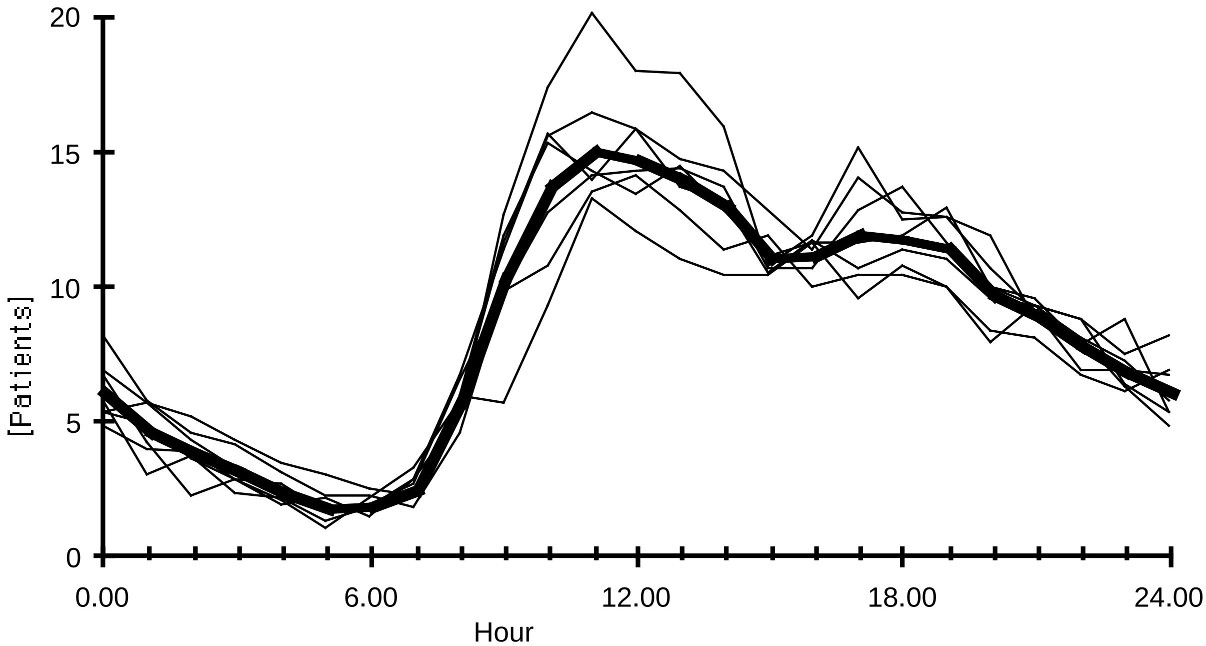

Figure 4 Average number of emergency patients arriving in A&E department in hourly intervals. Thin lines: arrival numbers for each of the days of the week. Bold line: hourly data averaged across days of the week.

Base Case output generation, performance measures and preliminary analysis

This sub-section describes how the simulation model was used to produce results and what summary statistics were computed. Some preliminary analysis of the system is also described.

The model was run for six simulated 24 hour cycles to obtain steady state data. Summary statistics from these runs are shown in Tables 1 and 2. Unless otherwise stated, these performance measures have been calculated from the steady state region of the run, all initialisation transients being excluded, and are averages across the daily cycle.

Since the starting point of the study was delays experienced in A&E, the measures calculated included daily averages across all patient types for the delays from registration to: consultation with an A&E doctor; decision to admit; and admission to wards. The daily minimum and maximum of the last of these - the total waiting time - was also calculated. Other measures track daily average figures for percentage of elective cancellations, proportion of A&E doctors' time spent with patients and hospital bed occupancy.

The output of the 'Base Case' simulation was also used to analyse the functioning of the system in finer detail. Emergency patients go through many procedures in A&E (see Figure 1) before it is possible to decide whether or not hospital admission is required, and delays in these activities contribute to the total time spent in A&E. Figure 5 displays these delays experienced in reaching these different stages. Such data may be presented in a variety of ways; we have chosen to show the average time taken to reach the completion of each stage, plotted against the hour of the day when that stage is completed. For example, at 6.00 the patients have waited about half an hour from registration to first consultation, from which we may deduce that they completed registration around 5:30. Similarly, at 00:00 patients about whom a decision to admit has been made have spent nearly five hours from registration to reach this point, from which we may deduce that they completed registration around 19:30 the previous day.

The waiting times averaged across all those emergency patients subsequently admitted to a ward is therefore given by the top line in Figure 5. At 24:00 patients admitted to a ward have spent more than six hours from registration to reach this point. From this we may deduce that they completed registration before 18:00. We can see that under normal conditions patients spend at least four and as much as eight and a half hours in the A&E department, depending on their time of arrival, before being given a bed. Note that patients who leave the department for hospital admission early in the evening have therefore experienced the longest waits, while patients who are transferred to beds on wards in early to mid-morning have spent relatively shorter times in A&E.

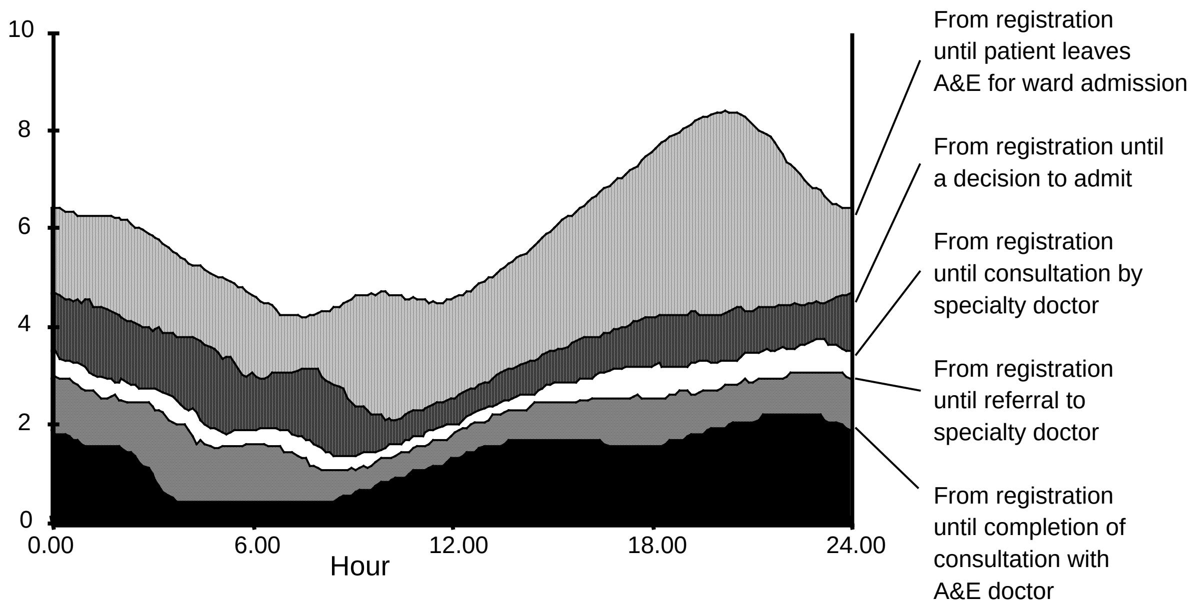

Figure 5 Emergency patient waiting time to reach different stages in A&E (Base Case simulation output).

According to the model results, the current hospital resources produce an average daily occupancy level of 95% (Table 1). Nevertheless, on average 16% of the scheduled elective admissions are cancelled to give room for emergency patients. Considerable variation may be seen in two components of total waiting time for emergency patients: the time from registration until being seen by an A&E doctor and the time waiting in A&E from the decision to admit until leaving A&E for hospital admission (Figure 5). The underlying reasons for these two delays are analysed in more detail below.

	Average time to A&E Dr. Consult. [Hours]	Average time to DTA [Hours]	Total Waiting Time (min, avg, max) [Hours]	Average % Elective Cancell-ations [%]	Average daily hospital Occupancy [%]	Average daily A&E Dr. Util. [%]
Base Case 800 beds and normal demand	1.3	3.6	4.2, 5.9, 8.4	16.2	94.6	92.1

Table 1 Performance measures for the Base Case model run.

Waiting time from registration until completion of A&E doctor consultation

The graph for time waiting to be seen, and then being seen, by an A&E doctor (Figure 5) shows a single deep trough in the morning and gentle afternoon and evening peaks. This is only partly explained by the emergency patient arrival pattern. Another factor is the variation in the roster arrangements for doctors in A&E. At the time of this study, the number of A&E doctors on duty at St. Dane's ranged from a peak of seven in the afternoon to a trough of two in the early hours of the morning. (The endogenous control of this capacity is the subject of an additional study[49]).

The trough in waiting time can now be analysed in relation to the provision of A&E doctors. At midnight a backlog of post-registration patients has built up and the four A&E doctors on duty are fully utilised (Figure 6, graph 1). Delays of nearly two hours result. After midnight the number of doctors falls to three but further A&E arrivals also tails off (Figure 4). Consequently, the backlog falls and the delay reduces. This continues until 3:00, when patients are being seen with a delay of only 25 minutes (the assumed minimum time spent having a consultation with a patient) and doctor utilisation begins to fall away. Although the number of doctors reduces to two at 4:00, utilisation rises to compensate and the delay remains minimal. By 6:00 doctor utilisation is at its lowest. Utilisation and delay rise after this but patients who arrive in casualty as late as 8:00 have waited on average only 25 minutes for a consultation after registration.

Although the number of doctors steps up steadily from 7:00 until early-afternoon, there is a considerable rise in emergency patients to contend with. Doctor utilisation, patient backlog and delay all therefore rise from 9:00 onward. Even with the maximum compliment of doctors in the early afternoon, there is insufficient capacity to keep up with the workload. The situation is exacerbated by a crowding effect as the cubicles in A&E become full. It then becomes more difficult to administer the patients who are waiting, the time expended by a doctor finding and giving an initial consultation to a patient rises and so the backlog is drained less quickly. A&E doctors therefore work flat out for the rest of the day. Only in the late afternoon do they begin to catch up with the backlog of patients, and the resulting slight dip in delay around 18:00 causes the afternoon peak. This persists only briefly as a second, evening, rush of arrivals produces a new backlog of patients waiting to be seen. During the evening, therefore, delay in consultations by doctors again rises, producing the second peak.

The waiting time from registration to A&E doctor consultation and the associated utilisation of A&E doctors is therefore seen to result from the combination of the pattern of arrival of emergency patients and the specific A&E doctor roster.

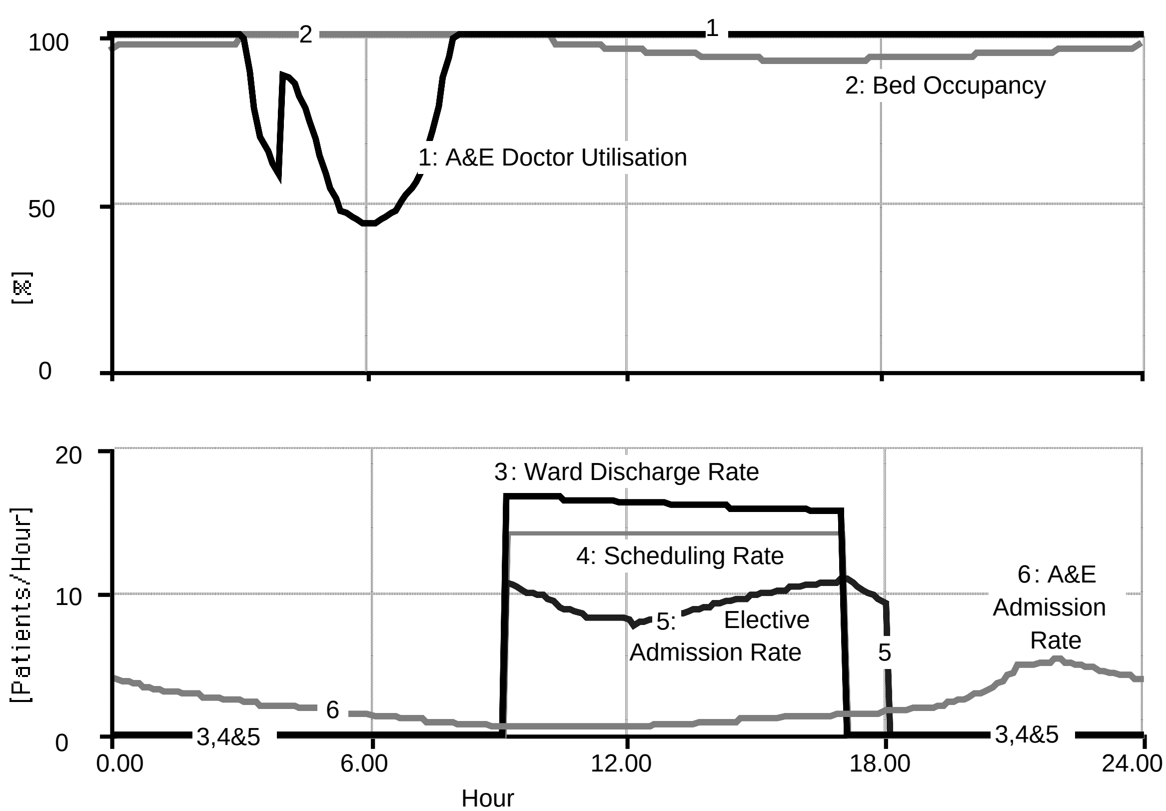

Figure 6 Outputs from the Base Case run of the model.

Waiting time from decision to admit until admission to ward

The longest waits experienced by patients are primarily attributable to delays in the ward admission process, this delay displaying a steep peak and trough pattern. This variability is not solely due to the backlog produced by the arrival pattern: the turnover rate of cleared beds and the management of elective patients combine to co-produce this behaviour. We now consider the causes of this delay in more detail (numbered graphs refer to Figure 6).

Patients are discharged from the hospital from 9:00 until 17:00. In the model these discharges are taken to occur after an average stay of six days, a figure which necessarily aggregates a wide variety of surgical and medical cases. As stated previously, there is a delay, the turnover interval, between patient discharge and the time when a bed is ready for a new patient. Hospital staff estimated that in the late evening and early morning the turnover interval is relatively low (average one hour), while from midday until late in the afternoon this interval is generally high (up to five hours). This difference is clearly generated by endogenous effects; procedures involving the interaction of administrative, cleaning and nursing duties. However, for simplicity this effect was represented in the model as an exogenous time series.

During the daytime, this component of waiting time is at its highest around 20:00 and relates to patients who registered in A&E at 12:00. In fact, this component rises from 8:00 onwards (Figure 5). During the day there is interaction with the elective patients that have been scheduled for admission, whilst increased bed turnover times and a second A&E crowding effect also contribute to the rise in delay. These three effects are discussed further below.

At 9:00 patients begin to be discharged from wards (Graph 3). However, the turnover time on free beds starts to increase as the activities of the day keep ward staff busy. Therefore, just as more patients start to arrive in A&E (Figure 4) and those requiring admission begin to accumulate, the time to re-cycle a free bed starts to increase. Bed occupancy falls (Graph 2) and delays to A&E patients mount. We now consider another reason for the rise in delays - crowding. An A&E patient must be accompanied to a ward by a porter and a nurse from A&E. As A&E becomes more pressured during the late afternoon and early evening and more potentially urgent cases must be attended to, it becomes increasingly difficult to spare nurses to do this. Consequently, although A&E patients have priority over elective patients for admission, this crowding effect limits the former's ability to take up the opportunity of an available bed. These beds are snapped up by elective patients, as long as they are seeking admission.

Elective patients intended for treatment the following day are scheduled to arrive throughout normal daytime hours (Graph 4) and those that can be admitted are a component of the elective admission rate (Graph 5). However, from 9:00 to 12:00 priority is afforded to those elective patients held overnight whose treatments are scheduled on this day. Such patients make up the remainder of the elective admission rate. After 12:00 such patients cannot be prepped in time and their treatment is cancelled. Consequently, all of the elective admission flow in the afternoon consists of patients intended for treatment the following day. This continues until 18:00, when some will be held over in the hotel unit for admission on the day of their treatment and the remainder are cancelled.

During the evening and the early hours of the morning these dynamics have a knock-on effect on the time spent by A&E patients awaiting an available bed. Although no patients are discharged from wards during this time, the delay resulting from the lengthened turnover interval means that beds released by earlier discharges are continuing to become available. In the absence at these hours of elective patients seeking admission, these are filled by waiting A&E patients. There is a surge in such admissions around 22:00 (Graph 6). This is associated with the late evening reduction in turnover time and also with the accumulated wave of A&E arrivals, phase-shifted beyond the period of elective patient admissions. In consequence, through midnight and into the early hours we see occupancy steadily rising and the lowest values of delay from decision to admit and admission (Graph 2 and Figure 5).

Model Validation

Previous sub-sections have described key aspects of the Base Case simulation output. Here we outline the processes that were undertaken to validate the model. Two introductory remarks are necessary. Firstly, model validity cannot be established by a single test, or at a single moment in a modelling study. Rather, system dynamicists - along with other simulators[50] - accept that validity, "accumulates gradually as the model passes more tests" (reference 51, p. 209). Therefore the treatment of the validity issue would ideally run in parallel with the description of the model and its runs. We have avoided such a clumsy structure and have instead chosen to consolidate the discussions of validation at this point. Secondly, because this paper focuses on the technical details of the model and because validation and the modelling process are the subjects of separate, more detailed, publications,[52,53] the treatment here is brief.

The validation tests specific to a system dynamics model are well established[39,51] and may be divided into those focusing on model structure and those relating primarily to simulated behaviour. We first outline the application of these two types of tests to the A&E model and then describe the team involved in performing them.

Structure-based validation tests. These validation tests are concerned with formulation and ensure, firstly that the model is suitable for its purpose and, secondly that it is consistent with the real system. To test suitability for purpose we focus inwards on the model. These tests ensure, firstly that the model is suitable for its purpose and, secondly that it is consistent with the real system. To test suitability for purpose we focus inwards on the model, testing that all variables and outputs of equations have sensible dimensions and that equations hold true for extreme input values. In this way we confirmed that the model was well posed.

Testing that the model structure is consistent with the real system involves judging the model's representativeness.[50] The close links with St. Dane's allowed us to discuss the structure with people familiar with the A&E system. Considerable time was spent on choosing variable names and on documenting the assumptions in algebraic formulations. Validity was further enhanced by confirming the correspondence of model parameters to information available.

Behaviour-based validation tests. These tests use model simulations to probe further the validity of its construction. The paucity of recorded data made it impossible to perform a full behaviour reproduction test[54,55] and so a process of triangulation involving subjective, qualitative time series analysis and objective, quantitative summary data was used. Graphs of all of the model variables were presented to our collaborators (Figures 5 & 6 show examples). This output, along with the performance indicators, was judged to be realistic and convincing by those with day-to-day experience of the real system.

A brief comment should be made on one of the quantitative validation points, 'average percentage elective cancellations', which is the ratio of those patients cancelled to those scheduled. This performance measure corresponds to the only available data, which was obtained from the relevant Health Authority, converted into the appropriate form and then cross-checked with estimates from the St. Dane's Bed Manager. Some adjustment is necessary because, in situations of great demand pressure, adroit bed management allows some beds temporarily vacated by patients undergoing surgical treatment to be used by new patients spending only a brief time in the hospital for some clinical treatment. This 'bed doubling' effectively boosts the actual occupancy of beds, though to a strictly limited extent. However, the Registrar and the Bed Manager felt that a satisfactory accounting of model behaviour had been achieved without an explicit modelling of bed doubling. One might also observe that the insights generated by the model result, essentially, from sensitivity analysis and that the qualitative insights that flow from this would not be altered by using adjusted data.

The above judgements were made from the Base Case. However, model validity was further enhanced by applying tests to the policy analysis runs described below. These scenarios can be seen as extreme condition tests,[51] the diagnosis of surprise behaviour[56] and the generation of insights.[39,51] The group process used to create the model meant that these simulations could be studied and judged reasonable by our collaborators.

The modelling team. For a model to be valid, one must be confident that it is, "suitable for its purposes and the problem it addresses ... [and] ... consistent with the slice of reality it tries to capture"(reference 39, p. 312). To make such judgements both experienced modellers and individuals with knowledge of the actual system are needed. The core team for this study consisted of the authors (acting as modellers and facilitators) and staff from St. Dane's. Our collaborators from St. Dane's included: the Registrar in the A&E department, other physicians from A&E, the Bed Manager, the Site Nurse Practitioner, other nurses and staff from other specialisms and from the test laboratories. In addition, aspects of the model formulation were checked with staff from the Southwark Community Health Council and we also benefited from the experience of staff in Casualty Watch, the London Ambulance Service and the Emergency Bed Service.

The details of how this team worked to create the model are recorded elsewhere.[53] Here we merely comment that this spread of individuals gave us access to both hospital databases and a range of judgmental estimates. However, it was the LSE team and the Registrar of St. Dane's who were responsible for checking each model assumption, element of formulation and parameter value used. By conducting the tests described above we aimed to ensure that the model accurately incorporated the variables, parameters and feedback effects necessary to analyse the underlying reasons for long waits in A&E. The model presented here was the result. We might also comment that it was a model which our collaborators at St. Dane's used with confidence (and enjoyment) to study the real system.

Policy Analysis: Exploring scenarios using model simulations

The initial motivation for the research described in this paper was to test the hypothesis that restrictions of bed capacity would not lead to increased waiting times in A&E. However, in the course of developing the model it became evident that waiting time is only one of a number of measures of system performance. Similarly, discussion with our collaborators indicated that other simulation experiments might be of interest. The developed model was therefore used to conduct a range of simulation experiments for comparison with the Base Case. The values of the key performance measures for these scenarios are shown in Table 2. The scenarios are discussed in turn below.

	Average time to A&E Dr. Consult. [Hours]	Average time to DTA [Hours]	Total Waiting Time (min, avg, max) [Hours]	Average % Elective Cancell-ations [%]	Average daily hospital Occupancy [%]	Average daily A&E Dr. Util. [%]
Bed Capacity Scenarios						
700 beds	1.3	3.6	4.2, 5.9, 8.4	30.4	95.4	92.1
800 beds (Base)	1.3	3.6	4.2, 5.9, 8.4	16.2	94.6	92.1
900 beds	1.3	3.6	4.2, 5.9, 8.4	7.8	90.0	92.1
Demand Pattern Scenarios						
Demand Increase						
0%	1.3	3.6	4.2, 5.9, 8.4	16.2	94.6	92.1
1%	1.4	3.7	4.3, 6.0, 8.4	16.5	94.6	93.6
2%	1.6	3.8	4.7, 6.2, 8.5	16.8	94.6	95.5
3%	1.7	4.0	4.9, 6.3, 8.5	17.1	94.7	97.5
4%	1.9	4.2	5.3, 6.6, 8.6	17.4	94.7	99.4
≥ 5%	→ ∞	→ ∞	→ ∞	n/a	n/a	n/a
Combined Scenario						
700 beds and 7% increase in demand (+ A&E Dr response)	2.3	4.6	6.0, 7.1, 8.8	32.5	95.5	99.5

Table 2 Consolidated performance measures for the various policy analysis runs of the model.

Bed Capacity Scenarios

System behaviour was simulated for levels of bed capacity between 700 and 900, the range chosen by our collaborators to be of interest. These scenarios were aimed at investigating whether the government was right in its assertion that closure of hospital beds would not affect the standard of service provided to the community. The effect that changes in hospital bed numbers have on A&E performance is shown in Table 2.

The performance measures, and analysis of the various outputs of the runs, reveal the extreme similarity of the three simulations; there are only trivial differences in output graphs. This surprising, counter-intuitive result[57,58] might appear to support the hypothesis that the model was built to test, namely that reductions in bed capacity can be compensated by increases in occupancy so that A&E waiting times do not increase. However, the performance measures reveal that average daily occupancy is increased only slightly as bed numbers are reduced (Table 2), and such changes are quite insufficient to substitute for the lost beds. The explanation for this apparent paradox lies in the figures for elective cancellations. This measure is far more sensitive to bed capacity. These results indicate that as hospital beds are removed more elective patients have their treatments cancelled. In order to deliver approximately the same waiting time to emergency patients with 100 beds less, cancelled non-emergency treatments almost double in absolute terms.

This response can be explained using the causal loop diagram of Figure 2. Balancing Loops B1, B2 and B3 together control the flow of patients onto wards. Specifically, B3 controls the balance between B4a and B4b. By shifting greater weight to B4b and cancelling elective admissions, the hospital manages - even with reduced bed numbers - to keep ward occupancy at a level which leaves virtually unaffected the waiting time for emergency patients. The elective cancellation ’safety valve’ removes the expected distortive effects on B2, and hence B1. Patients will not experience additional delays in the various A&E activities as a result of reduced bed numbers in hospital wards. The operation of loop B1 is unaffected by the changes in bed capacity and so no more patients will occupy trolleys in casualty than previously.

The response of the hospital occupancy performance measure has a less complex explanation, the slow decline indicating that the various processing capacities involved in the different admission activities result in a limited ability to take advantage of further beds.

Demand Pattern Scenarios

The response of the model to changes in the number of emergency patients presenting was also examined in two ways. By applying permanent changes we considered possible future demand environments for the A&E department; and by studying the model's response to a sharp transient increase we both tested its ability to reproduce past behaviour in a plausible way and explained the system's response to a crisis event.

Permanent changes in demand. The aim of these scenarios is to explore how the system - with today's staff numbers and bed resources - would behave in situations of increased load. These might be caused either by the permanent closure of a casualty unit in a nearby hospital, or simply by a generalised increase in demand. (Further increases in emergency arrivals and admissions would be in line both with short term experience - a 7% rise in the previous year - and long term trends[59]). Reductions in demand were deemed unrealistic and therefore irrelevant by our collaborators. For reasons explained below, the model was simulated for permanent changes in emergency demand of up to 5%. The results are shown in Table 2 and Figure 7, with data for demand reductions plotted only for comparison purposes.

Elective cancellations once again increase but to a smaller extent than in the previous set of scenarios. This is because only a small fraction of A&E patients actually require admission and so increasing this fraction adds only slightly to the total number of all patients seeking admission. However, small changes in demand do have appreciable effects on patients' total waiting time, primarily because of the increase in delay before consultation with an A&E doctor. With a 4% increase, patients spend an average of three quarters of an hour more in A&E. Higher levels of arrivals also produce increases in the daily averaged utilisation of A&E doctors. However, with increases up to 4%, there is still free doctor capacity at certain hours of the day, though this decreases steadily. Although patients experience delays for consultations there is just sufficient staff to cope with the workload over a 24 hour cycle. Note, however, that with a 4% increase in demand doctor utilisation virtually reaches 100%.

Beyond 4%, this gradually deteriorating balance collapses. There is no slack A&E doctor capacity to cope with the extra arrivals, a backlog of patients awaiting initial consultation accumulates without bound and, even with staff working at maximum capacity, the waiting time escalates from hour to hour. The system therefore fails to reach a steady state and the performance measures can no longer be calculated. Interestingly, the cause of this collapse is not bed capacity, but rather insufficient provision of A&E doctors.

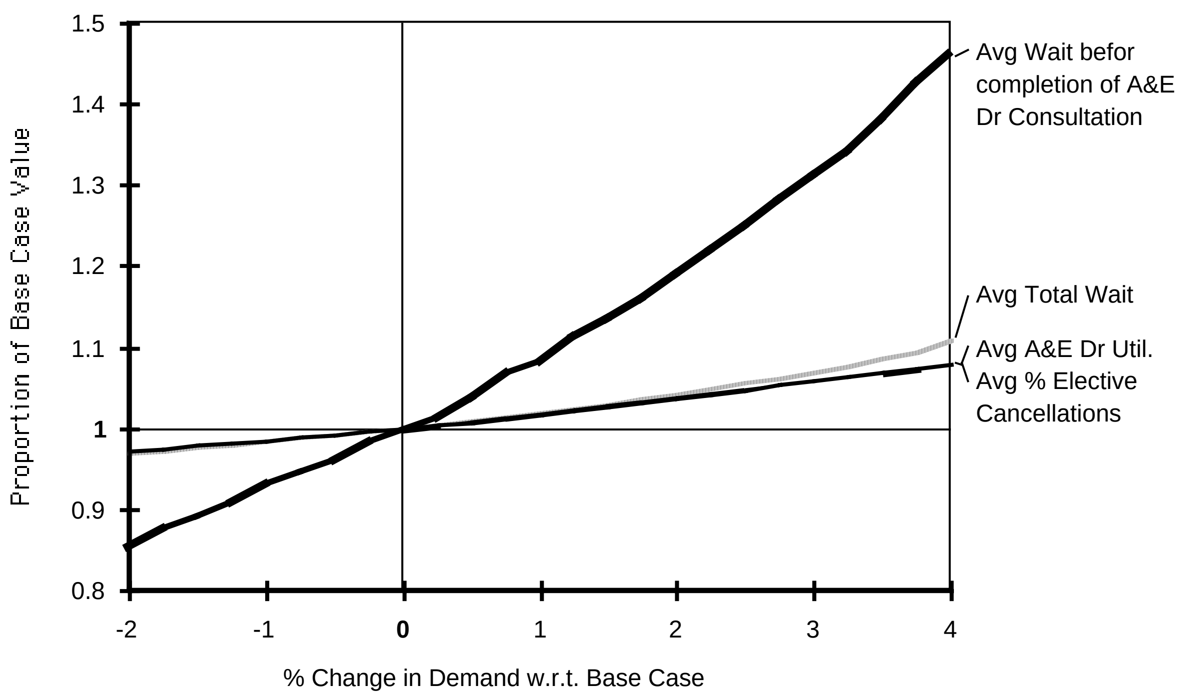

Figure 7 Spiderplot showing the proportional effect of permanent demand changes on a selection of performance measures.

A crisis event.

N.B. Removed in this shortened paper.

Combined Scenario

In the final scenario a situation was simulated with a combination of changes from the Base case, namely 700 hospital beds and a 7% increase in demand for A&E services. However, the simulation described above has shown that A&E doctors would be overwhelmed by demand increases above 4%. To bridge this gap we applied an increase of 3% to the consultation rate. This crudely simulates A&E doctors working faster in response to the increased demand (possible deleterious effects on the quality of their subsequent diagnoses are not considered). De-bottlenecking this component of the model was necessary if the behaviour of the other activities was to be examined. This scenario therefore explores system behaviour under reasonable predictions of annually increased demand if, at the same time, a policy of further reduction in bed numbers to achieve NHS economies was to be implemented (Figure 8 and Table 2).

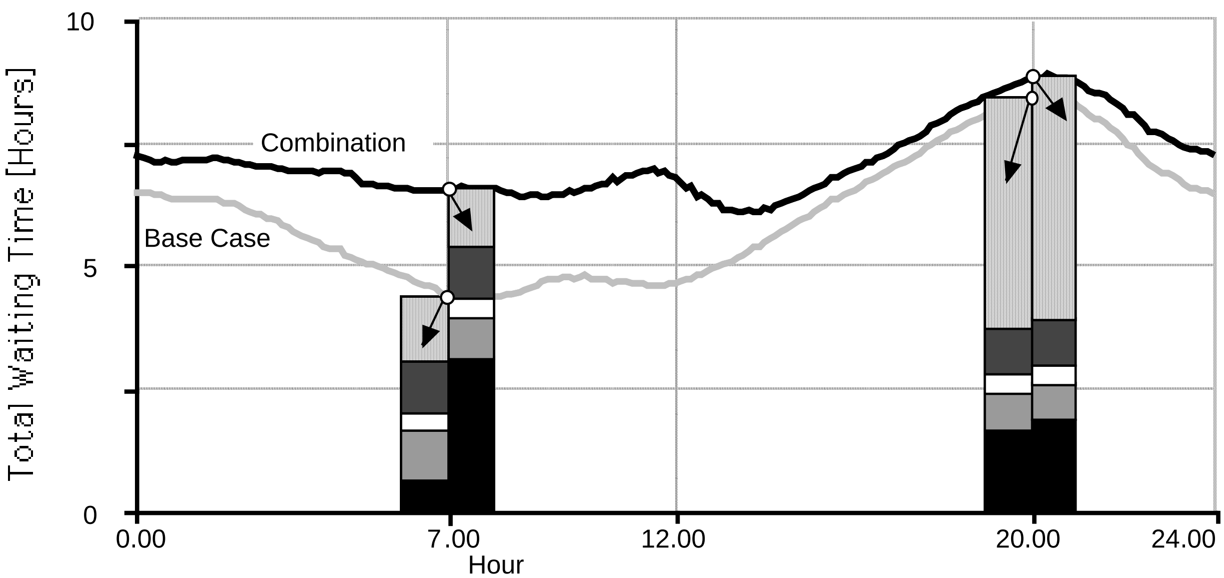

Figure 8 Comparison of total waiting time under normal conditions (Base Case) and in the scenario of combined demand increase and bed reduction, with illustrative decomposition into stages of waiting.

As might have been expected, total waiting time for patients increases. The greatest contribution to increased waiting time occurs in the morning, the experience of patients arriving on wards around 7:00 exemplifying the effect (Figure 8). In this scenario, these emergency patients entered A&E around 00:30 and have therefore waited six and a half hours before admission. This compares with the Base Case in which patients only waited four and a half hours, having entered A&E just before 3:00 the same day. This increase has a single main component, the delay before obtaining a first consultation, which has increased fivefold, reaching three hours. The analysis of the Base Case and of the scenarios with permanent changes in demand allow us to see that this results from the loading on A&E doctor capacity caused by the demand increase.

The increase in total waiting time in the afternoon and evening is smaller. Patients admitted to wards at 20:00 have waited less than half an hour extra in total. This increase is largely in the delay prior to first consultation, with a small contribution from the delay between decision to admit and actual admission (Figure 8). From the previous bed capacity scenarios and analysis of this run, we can determine that this latter change is associated with the small increase in patients needing beds, rather than the reduction in bed capacity. Patients arriving on wards at 20:00 have waited almost nine hours since arriving in A&E but during this period (from 11:00) the 'safety valve' of elective patient cancellations is used to cope with the reduced bed capacity. This is evidenced in the sharp increase in elective cancellations shown in Table 2, up to 32.5% across a 24 hour cycle. At times when loop B4b is available (Figure 2) it is used to control waiting times and the penalty of a mismatch between demand and resources is substantially transferred to a different performance measure.

Comments and conclusions

The development of the model of an A&E department reported in this paper, and the simulation runs which have been carried out, offer lessons at a number of different levels. Firstly, there are direct, practical implications from the Base Case and the policy analysis runs; secondly, there are more general lessons as to the connectedness of the system of which A&E is a part; and finally there are proposals for the elaboration and further use of the model.

Lessons from the Simulations

Certain conclusions can be drawn from the Base Case and policy analysis runs described in the previous sections; these are detailed below, against the scenarios which give rise to them. Although they apply formally only to situations comparable to those studied at St. Dane's, their wider applicability is quite evident.

The Base Case. The analysis indicates that much of the waiting time experienced before admission is inevitable: the constituent processes (Figure 1) are many and simply take time. However, at St. Dane's the restricted A&E doctor capacity in the morning and the limited availability of beds in the afternoon produce avoidable additional delay. The average daily occupancy level of 92% - above the 80-85% recommended by the British Association of Accident and Emergency Medicine[60] - and the high A&E doctor utilisation indicate an intensive use of resources. Nevertheless, capacity is insufficient to deal with demand: waiting times are high and elective patients are regularly cancelled. This is a system with little room for manoeuvre and with few 'efficiencies' waiting to be squeezed out.

Bed capacity scenarios. The bed occupancy measure was shown to be relatively insensitive to total bed capacity. These scenarios clearly indicate that without changes in both A&E and ward staffing, there are limits to how fast new patients and used beds can be made ready. These simulations demonstrate that enhanced bed occupancy levels cannot compensate for further bed reductions. Instead, the burden is borne elsewhere: it is the 'safety valve' of elective cancellations that is used to a greater extent to compensate for any bed loss. The rapid increase in these cancellations as bed capacity falls is what prevents a further rise in daytime A&E delays. It follows that the appropriate choice of indicators is crucial in judging the performance of the system. Taking an holistic view of the healthcare system within which A&E is embedded, it becomes evident that delays to patients in both casualty and elective cancellations must be assessed - they are linked and compensating measures.

Demand pattern scenarios. Scenarios with permanent demand changes show that despite increased elective cancellations, bed occupancy and utilisation of A&E doctors approach 100%. Eventually the system collapses, though it is the earliest process - consultation with an A&E doctor - which is overwhelmed first.

Combined scenario. This scenario serves to emphasis the error of the dynamic hypothesis - that bed occupancy levels can absorb a relative increase of demand pressure on resources. There is even less slack in this system, and it can reasonably be inferred that it is highly exposed to a crisis event or to a reduction in staff capacity due to illness.

General policy conclusions

The principal message of this study is that, while A&E waiting times may in practice be excellent measures of the effectiveness of acute hospital, using them alone to judge the effect of bed reductions is systemically naive. Recent developments in multiple performance indicators[61] recapitulate an old idea: complex systems must be monitored using a corresponding variety of signals.[62] In the context of A&E the observation has been made before: "studies should be based on appropriate outcome measures" (reference 11, p. 41). By providing a rigorous description of the complexity of an A&E department, our model acts as a formal platform from which conclusions may be drawn which address the complexity of the system's signals. By concentrating on A&E delays - important as these are - the original dynamic hypothesis provides misleading, indeed dysfunctional guidance to policy. The hypothesis encourages policymakers to look in the wrong place for healthcare improvements because it implicitly discounts the effects of cancelling elective patients. However, this 'safety valve' is not without its costs. Even with current bed numbers, headlines such as "Girl's heart surgery cancelled five times" (*Guardian*, 12.09.96) are all too numerous. Patients waiting on lists for elective admission are the victims in the battle for beds on wards, because of the priority necessarily allotted to emergency patients. In the light of this, and of the low amount of slack at St. Dane's and elsewhere, the consequence of a further reduction in beds must be to increase cancellations and push hospitals closer to being mere 'emergency wards'.

This study reveals the interconnectedness of A&E service levels, bed provision and the experience of patients on waiting lists. Bringing that interconnectedness to life via simulation can, in principle, also provoke a discussion on appropriate measures of performance. A&E departments do not exist in isolation. Policy must be based on an understanding of how they relate to pre-hospital circumstances, to the rest of the hospital and to care in the surrounding community,[1] and a set of performance measures reflecting this broad view should therefore inform policy making.

Possible elaborations of the model

Any model has limitations, and the model reported in this paper is no exception: it concentrates on short timescale effects, aggregates patient attributes and has a simplified representation of a hospital. These limitations appeared, both to the study team and its collaborators, to be appropriate to the questions which were being addressed. Nevertheless, there are a number of model elaborations that could throw light on a range of related issues.

i) De-bottlenecking. The simulated system collapse provoked by inadequate provision of A&E doctors masks other effects. With few changes, the model can be run with this capacity scaled up in order to reveal the effects of increased demand on other system elements. These runs could be used to inform judgement on priorities for the provision of additional resources.

ii) Longer timescale effects. Models operating with a longer timescale might have the potential to explore a range of effects. Scenarios in which staff operate at permanently high utilisations could incorporate the effect of 'worker burnout', as downward spiralling productivity serves to exacerbate individual workload and so further degrade productivity.[63] It might also be possible to investigate the longer term effects of cancelling elective admissions, since those so dislodged cycle back into an ageing chain, emerging with a higher priority, either because of the duration of their wait or because they become emergency cases. High occupancy levels and/or high demand levels can lead to great pressure to free up beds, leading to inappropriately early discharge. Modelling this effect would allow the consideration of the extra demand created by the subsequent relapse of some of these patients.[64] Lastly, there are interesting systemic interactions involving the phenomenon of so-called 'bed blocking' caused by the low provision of community care services for elderly people.[8] The addition of new care capacity would merit study since it might have the effect of increasing effective bed capacity in hospitals.

iii) Hospital clusters. To test the ability of A&E services to respond to demand surges, it would be desirable to consider the resilience of a network of hospitals across which a demand surge would be shared. (Possible sources of such a surge include a major accident, or the closure of A&E to admissions at one or more hospitals within the network). Such a 'Cluster Model' is currently under development.

iv) Feedback control of A&E doctor roster. In its current form the model stimulates a complex system with inputs in order to understand the nature of the response mechanism. The system dynamics approach prompts us to see how endogenous feedback effects might be improved; to this end we have been developing improved policies for controlling the rostering of A&E doctors.[49]

In conclusion, we would say that the system dynamic model reported in this paper helped our collaborators to consider the interconnectedness of factors affecting the performance of A&E at an appropriate level of detail and to explore the effect of changes in parameter values. We hope that the model attracts wider attention. We believe that it offers those involved with the planning and provision of health services an opportunity to base their decisions on a systemic analysis and so improve the quality of the healthcare that is delivered.

Acknowledgements - We would like to offer our thanks to staff at Casualty Watch, London Ambulance Service, the Emergency Bed Service, Southwark Community Health Council, the Department of Health and particularly all of the staff at the collaborative hospital for the time that they contributed to this study. Without the access that they afforded us, and their personal involvement, this work would not have been possible.

References

N.B. References 25,39,42,43 and 58 have been republished by Productivity Press, Portland OR; See References 65, 66 and 67.

1 Audit Commission (1996). *By Accident or Design: Improving A&E Services in England and Wales*. HMSO: London.
2 Department of Health (1996). *The Patient's Charter*. HMSO: London.
3 NHS Confederation/Royal College of Physicians (1997). *Tackling NHS Emergency Admissions: Policy into practice*. NHS Confederation: Birmingham.

4 Department of Health (1998). *Outpatients and Ward Attenders, England: Financial year 1996-1997*. HMSO: London.

5 Hadfield, J, Yates, D and Berry, A (1994). The Emergency Department and the Community: A model for improved cooperation. *Journal of the Royal Society of Medicine* **87**: 663-665.

6 Wolstenholme, EF (1993). A case study in community care using systems thinking. *Journal of the Operational Research Society* **44**: 925-934.

7 Lane, DC (1994). System dynamics practice: a comment on 'a case study in community care using systems thinking'. *Journal of the Operational Research Society* **45**: 361-363.

8 Audit Commission (1997). *The Coming of Age: Improving care services for older people*. Audit Commission Publications: Abingdon.

9 Derlet, RW, Nishio, DA, Cole, LM and Silva, J (1992). Triage of patients out of an emergency department: three year experience. *American Journal of Emergency Medicine* **10**: 195-199.

10 Dale, J, Green, J, Reed, F and Glucksman, E (1995). Primary care in the accident and emergency department: I. Prospective identification of patients. *British Medical Journal* **311**: 423-426.

11 Leydon, GM, Lawrenson, R, Meakin, R and Roberts, JA (1996). *The cost of alternative models of accident and emergency care: a systematic review*. Report to North Thames Regional Health Authority:

12 Meislin, HW, Coates, SA, Cyr, J and Valenzuela, T (1988). Fast track: Urgent care within a teaching hospital emergency department: can it work? *Annals of Emergency Medicine* **17**: 453-456.

13 Middleton, EL and Whitney, FW (1993). Primary care in the emergency room: a collaborative model. *Nursing Connections* **6**: 29-40.

14 Davies, R (1985). An assessment of models of a health system. *Journal of the Operational Research Society* **36**: 679-696.

15 Davies, R and Davies, H (1994). Modelling patient flows and resource provision in health systems. *Omega* **22**: 123-131.

16 Paul, R (1995). Outpatient Clinics. *OR Insight* **8**: 24-27.

17 O'Kane, PC (1981). Simulation model of a diagnostic radiology department. *EJOR* **6**: 38-45.

18 Riley, J (1995) Visual Interactive Simulation of Accident and Emergency Departments. In: Kastelein, A, Vissers, J, van Merode, GG and Delesie, L (eds). *Proceedings of ORAHS 21: Managing health care under resource constraints*. Eindhoven University Press: Maastricht, pp 135-141.

19 Huang, X-M, Taylor, C and Stewart, I (1995). Computer simulation of an A&E department. *OR Insight* **8**: 18-23.

20 Altinel, IK and Ulas, E (1996). Simulation Modelling for Emergency Bed Requirement Planning. *Annals of Operational Research* **67**: 183-210.

21 Romanin-Jacur, G and Facchin, P (1987). Optimal planning of a pediatric semi-intensive care unit via simulation. *EJOR* **29**: 192.

22 Worthington, D (1991). Hospital waiting list management models. *Journal of the Operational Research Society* **42**: 833-843.

23 Davies, R (1995). The growing need for renal services. *OR Insight* **8**: 6-11.

24 Mejia, A, Shirazi, R, Beech, R and Balmer, D (1998). Planning midwifery services to deliver continuity of care. *Journal of the Operational Research Society* **49**: 33-41.

25 Forrester, JW (1961). *Industrial Dynamics*. MIT Press: Cambridge, MA.

26 Taylor, KS and Lane, DC (1998). Simulation Applied to Health Services: Some experiences so far and opportunities for applying the system dynamics approach. *Journal of Health Services Research and Policy*, to appear.

27 Richardson, GP (1991). *Feedback thought in social science and systems theory*. Univ. Pennsylvania: Philadelphia.

28 Richmond, B (1985) STELLA: Software for bringing system dynamics to the other 98%. In: Andersen, DF, Forrester, NB and Warkentin, ME (eds). *1985 International Conference of the System Dynamics Society: Volume II*. System Dynamics Society: Boston, MA, pp 706-718.

29 Richmond, BM, Vescuso, P and Peterson, S (1990). *iThink™ Software Manuals*. High Performance Systems, 145 Lyme Road, Hanover, NH 03755, USA.: Hanover, NH.

30 Morecroft, JDW (1988). System dynamics and microworlds for policymakers. *EJOR* **35**: 301-320.

31 Lane, DC (1992). Modelling as learning: a consultancy methodology for enhancing learning in management teams. *European Journal of Operational Research* **59**: 64-84.

32 Vennix, JAM (1996). *Group Model-building: Facilitating team learning using system dynamics*. Wiley: Chichester.

33 Forrester, JW (1992). Policies, decisions and information sources for modelling. *EJOR* **59**: 42-63.

34 Senge, P (1990). *The Fifth Discipline: The art and practice of the learning organization*. Doubleday/Currency: New York.

35 Senge, P and Sterman, JD (1992). Systems thinking and organizational learning: acting locally and thinking globally in the organization of the future. *EJOR* **59**: 137-150.

36 Lane, DC (1993). The road not taken: observing a process of issue selection and model conceptualization. *SDR* **9**: 239-264.
37 Lane, DC (1995). On a resurgence of management simulations and games. *Journal of the Operational Research Society* **46**: 604-625.
38 Lane, DC (1997) Diary of an Oil Market Model: How a system dynamics modelling process was used with managers to resolve conflict and to generate insight. In: Bunn, DW and Larsen, ER (eds). *Systems Modelling for Energy Policy*. Wiley: Chichester, pp 205-240.
39 Richardson, GP and Pugh, AL ({1981} 19--). *Introduction to System Dynamics Modelling with DYNAMO.* Productivity: Cambridge, MA.
40 Sterman, JD (1988) A skeptic's guide to computer models. In: Grant, L (ed) *Foresight and National Decisions*. University Press of America: Lanham, MD, pp 133-169.
41 Lane, DC and Smart, C (1996). Reinterpreting 'generic structure': evolution, application and limitations of a concept. *SDR* **12**: 87-120.
42 Randers, J (1980) Guidelines for model conceptualisation. In: Randers, J (ed) *Elements of the System Dynamics Method*. MIT Press: Cambridge, MA, pp 117-139.
43 Goodman, MR (1974). *Study Notes in System Dynamics*. MIT Press: Cambridge, MA.
44 Tomlinson, B (1992). *Report of the inquiry into London's health service, medical education and research*. HMSO: London.
45 King's Fund Commission (1992). *London Health Care 2010: Changing the future of services in the capital*. King's Fund: London.
46 Department of Health (1993). *Making London Better*. HMSO: London.
47 Homer, JB (1996). Why we iterate: scientific modelling in theory and practice. *SDR* **12**: 1-19.
48 Peterson, S (1994) Software for model building and simulation: an illustration of design philosophy. In: Morecroft, JDW and Sterman, JD (eds). *Modeling for learning organizations*. Productivity Press: Portland, OR, pp 291-300.
49 Monefeldt, C and Lane, DC (1997). A&E Model: Initial investigation of improved rostering policies. *LSE Research Notes*
50 Lane, DC (1995) The Folding Star: a comparative reframing and extension of validity concepts in system dynamics. In: Shimada, T and Saeed, K (eds). *Proceedings of the 1995 International System Dynamics Conference: Volume I - Plenary Program*. Gakushuin University and the International System Dynamics Society: Tokyo, pp 111-130.
51 Forrester, JW and Senge, PM (1980) Tests for building confidence in system dynamics models. In: Lagasto, AA, Forrester, JW and Lyneis, JM (eds). *System Dynamics: TIMS Studies in the Management Sciences*. North-Holland: Oxford, pp 209-228.
52 Monefeldt, C and Lane, DC (1997). A&E Model: Validation notes. *LSE Research Notes*
53 Monefeldt, C and Lane, DC (1997). A&E Model: Collaboration meeting notes. *LSE Research Notes*
54 Lane, DC (-). Can We Have Confidence In Generic Structures? *Journal of the Operational Research Society*, to appear.
55 Sterman, JD (1984). Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* **10**: 51-66.
56 Sterman, JD, Repenning, N and Kofman, F (1997). Unanticipated side effects of successful quality improvement programs: exploring a paradox of organizational improvement. *MS* **43**: 503-521.
57 Mass, NJ (1991). Diagnosing surprise model behavior: a tool for evolving behavioral and policy insights (1981). *System Dynamics Review* **7**: 68-86.
58 Forrester, JW (1970) Counterintuitive behaviour of Social Systems. In: *Collected Papers of Jay W. Forrester*. Wright-Allen Press: Cambridge, MA, pp 211-244.
59 Fry, L (1960). Casualties and casuals. *Lancet* **1**: 163-166.
60 Royal College of Surgeons of England (1993). *British Association for Accident and Emergency Medicine Directory 1993*. Royal College of Surgeons of England: London.
61 Kaplan, RS and Norton, DP (1992). The balanced scorecard - measures that drive performance. *HBR* **Jan./Feb.**: 71-79.
62 Ashby, WR (1956). *An Introduction to Cybernetics*. Methuen: London.
63 Homer, JB (1985). Worker Burnout: A Dynamic Model with Implications for Prevention and Control. *System Dynamics Rev.* **1**: 42-62.
64 Audit Commission (1992). *Lying in Wait: The Use of Medical Beds in Acute Hospitals.* HMSO: London.
65 Lane, DC (1997). Invited Review and Reappraisal: 'Industrial Dynamics' by Jay W. Forrester. *Journal of the Operational Research Society* **48**: 1037-1042.
66 Lane, DC (1997). Four Invited Book Reviews. *Journal of the Operational Research Society* **48**: 1144-1148.
67 Lane, DC (1997). Invited Reviews on System Dynamics. *Journal of the Operational Research Society* **48**: 1254-1257.