

Emergency - but no Accident

- a system dynamics study of an accident and emergency department

David Lane, Camilla Monefeldt and Jonathan Rosenhead

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 analysis indicates that while reductions in bed numbers increase waiting times for emergency admissions to a small extent, their principal effect is 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.

-oo0oo-

Whenever high winter approaches, Britain expects a crisis in the National Health Service. So it was last year, and the year before that... But not in the beginning. For there is nothing natural about this cyclical phenomenon. It is not like leaves falling off trees, or even the midwinter blues engendered by daylight deprivation in northern climes. The spectacle of sick people whose conditions merit admission to hospital waiting overnight or longer on trolleys in Accident and Emergency, or even being treated in ambulances parked outside, is unnatural. Or, to put it another way, it is the unplanned and unintended consequence of a set of consciously taken decisions. These decisions - government decisions about priorities and budgets, management decisions about allocation of resources - are designed to achieve other outcomes. But possibly this is small consolation to the patients on the trolleys.

Right now A&E is the health service problem in the political and media spotlight. Indeed, A&E waiting times are a particularly sensitive indicator of stress in the health service, demonstrating in the most graphic way that the system, taken as a whole, isn't working as it should. But it is only one problem among many. Waiting lists, cost containment, cancellations of routine surgery, bed closures, bed

blocking, community care ... they queue behind each other, and as action is taken to tackle the one at the front, the next one in line reveals itself. The thrust of this article is that this merrygoround of failed palliatives will continue until the interconnections between these issues are understood systemically.

The scale of the problem

Accident and Emergency (also referred to as Casualty) provides access to hospital for urgent cases which need to bypass the normal process of referral, outpatient appointment, and pre-scheduled admission to a hospital bed. Individuals may be brought to A&E by ambulance, or walk in and present themselves for treatment. The latter, in particular, include people whose medical conditions vary widely in severity. A&E therefore has to perform a sorting function, dealing itself with some patients, and assessing more serious cases for possible admission as hospital inpatients. Those eventually admitted to hospital wards make up only 15 to 20% of those arriving at A&E (Audit Commission, 1996). It is long waits for admission for these patients which has been the focus of public concern, and which is the topic on which this article centres.

According to the Patient's Charter, patients admitted to hospital through an A&E department are supposed to be given a bed within two hours (Department of Health, 1996). In practice this standard is routinely broken in many hospitals. Waits of eight hours or longer are by no means exceptional - a distressing experience for severely ill patients, and still more so for their relatives when, as can happen, death arrives before a hospital bed. At times A&E units have become so congested that they have been 'closed to blue lights' - even emergency ambulances are barred and diverted to other hospitals.

Quite a range of explanations have been offered for the crisis in A&E, though little is known about the scale of the various effects. The closure of hospital beds is one obvious target. The British Medical Association estimated that 9000 acute beds were closed in England and Wales during 1991 to 1995. Evidently shortage of beds can have both direct and indirect effects on A&E. Directly, a patient in A&E may find that there is just no vacant bed available. Indirectly, bed shortages cause cancellations of scheduled non-emergency admissions. Quite possibly some of these may, as a result of the delay, become emergency cases themselves. Again, the pressure to free up beds may lead to cases of inappropriately early discharge of current inpatients - who as a result may need unscheduled re-admission. It has also been suggested that some part of the increased attendances at A&E constitute a behavioural response by patients and their GPs to the difficulty of securing admission to hospital through the normal referral process. Certainly attendances at A&E are up sharply - from 3,260,000 in financial year 1992-93 to over 3,825,000 in 1996-97 in London (Department of Health, 1998)

Bed shortage is not the only possible culprit in the A&E crisis. Too many of the beds in acute hospitals are 'blocked' by patients fit enough for discharge but unable to leave because of the lack of resources to care for them in the community. Paradoxically, another possible contributory factor is the politically driven campaign over recent years to bring the maximum waiting time for non-emergency admission to hospital down to 18 months. The effect of initiatives like this, targeted at those who have been waiting longest, is to give precedence to less serious cases, dislodging more urgent ones which may then require emergency admission. Yet another level of argument blames competition on price between hospitals, promoted since 1990 by the internal pseudo-market in health services; or the accounting balance between fixed and variable costs. Each has been seen as driving bed occupancy upwards to a level where there is simply insufficient slack to cope with demand variations.

Handling complexity with system dynamics

This thumb-nail sketch of the tangle of issues which may be implicated in the problems afflicting A&E should be sufficient to indicate the perils of taking remedial action based on just one or a small subset

of the relevant factors. The difficulty has been to develop a systematic understanding of the ways in which these, and other, factors interact to produce the observed, and unwelcome, behaviour. This difficulty is the focus of a continuing research project based at the Department of Operational Research at LSE. The approach which we have been using is called System Dynamics, a computer-based method which might have been tailor-made for the intricacies of the A&E problem - except that it is in fact the end result of 40 years of sustained development.

The creation of Jay Forrester of MIT in the late 1950's, system dynamics is a modelling approach which is used both to understand how social systems - in which human actors take decisions which affect events - change over time, and how these changes might best be influenced. System dynamics models represent the flow of information and materials which connect the state of the system to the decision points, and also the ramifying consequences of the decisions made. These closed loops of causal links frequently involve delays and non-linear relationships. Computer simulation of systems of this kind is crucial since humans can conceptualise such complex models but do not have the cognitive capacity to predict the dynamic consequences unassisted. Simulation enables us to calculate the consequences over time of complex causal structures, so that counter-intuitive behaviour may be explored in a rigorous way. Such models can then be used to craft new guiding policies for the system and to advance the intuition of those who will implement them (Forrester, 1961 and 1968, Lane, 1997a).

Assisted by modern, icon-based software, system dynamics has provided causal theories for a wide variety of phenomena. Forrester's pioneering work ranged from the management of growth in a start-up company to the interaction of population and resources on a global scale (Lane, 1997b). Contemporary applications have included, for example, the diffusion of new medical technologies, the dynamics of software development, the long-term management of life insurance companies and the collapse of the Mayan empire.

Operational researchers have normally used the discrete event simulation approach to healthcare modelling (Davies, 1985, Huang, et al, 1995). System dynamics modelling offers many other benefits (Lane, 1992) playing a different role in

probing the operation of healthcare systems. A general discussion of this role is available elsewhere (Taylor and Lane, 1998).

System dynamics, therefore, seemed tailor-made as an approach to use when staff at LSE began their involvement with the problems confronting A&E. LSE's involvement started through contacts with Casualty Watch, a project based in the Southwark Community Health Council but now with a remit which is London-wide, and beyond. Casualty Watch monitors the performance of hospital A&E departments on a monthly basis; it operates by sending volunteers to make simultaneous co-ordinated visits to casualty units, where they record the length of time that the patients currently in A&E have been waiting.

The Casualty Watch surveys can be used to track trends in waiting times, and publicise them if appropriate. What they cannot do is pin the responsibility for increases on any particular factor or combination of factors. The work carried out at LSE has been designed to fill this gap. We have built a system dynamics model whose purpose is to explore how factors internal and external to A&E contribute to delays at casualty.

To make such a model credible and authoritative it is necessary to root it in observed real world behaviour. This we have been able to do through intensive collaboration with the A&E department of a major inner-London teaching hospital, which for confidentiality reasons we will refer to as 'St Dane's'. Aspects of the model formulation were also checked with Community Watch, the London Ambulance Service and the Emergency Bed Service.

Modelling A&E

How can we understand the way in which the factors influencing casualty unit performance could combine to cause havoc? One way is to focus in on the detailed procedures which need to be carried out within the A&E department before a patient can be admitted and sent up to an available bed. (The emphasis in this research is on these, rather than less seriously ill attenders at A&E.) A representation of how this works is given in Figure 1.

Another way, which provides greater insight, is to look at the feedback loops which control the flow of

patients both into and out of A&E. A patient can only leave A&E for the wards when there is an available bed; the availability of beds depends on the discharge of existing patients; and an available bed may be filled either by an admission from casualty or by a scheduled ('elective') admission. (Particular beds are not dedicated to emergency patients.) The logic of these relationships can be displayed in a 'causal loop diagram', such as that shown in Figure 2.

In brief, loop B1 acts to drain emergency patients from A&E by admitting them to hospital wards, once necessary activities within the casualty unit have been completed. Loop B2 ensures that emergency patients are admitted to a bed on a ward when, and only when, a bed is available. Loop B3 performs a similar function for elective admissions. Loops B4 and B5 together control the backlog of patients scheduled for elective admission. Loop B4 acts to admit such patients into a ward, when there are free beds. If hospital occupancy is too high for all scheduled elective patients to be admitted, then loop B5 operates and cancels the surplus.

A stock and flow diagram was derived from this outline representation, expressing the interconnections in more detail. From this a computer model could be developed which was used to explore how the different feed-back loops combine to produce the system's performance. Models like this can help to provide conceptual clarity about the processes which generate the waiting times which patients experience in A&E. They become much more useful, however, if the various rates of flow and the feedback effects between them are based on reliable evidence. In this study we received access not only to hospital databases at St Dane's, but also to the considered judgements of a wide range of staff both within the A&E department and elsewhere in the hospital, as well as from relevant outside organisations. The end result was a model which was validated in terms both of its inputs and of its internal relationships, and whose outputs were accepted by our experienced collaborators as realistic.

What causes long waiting times?

A model has the advantage over the real world that one can play with its parameters fairly freely to explore what would happen if the system were to be quite drastically changed, without having to worry about the chaos that might ensue. The methodology

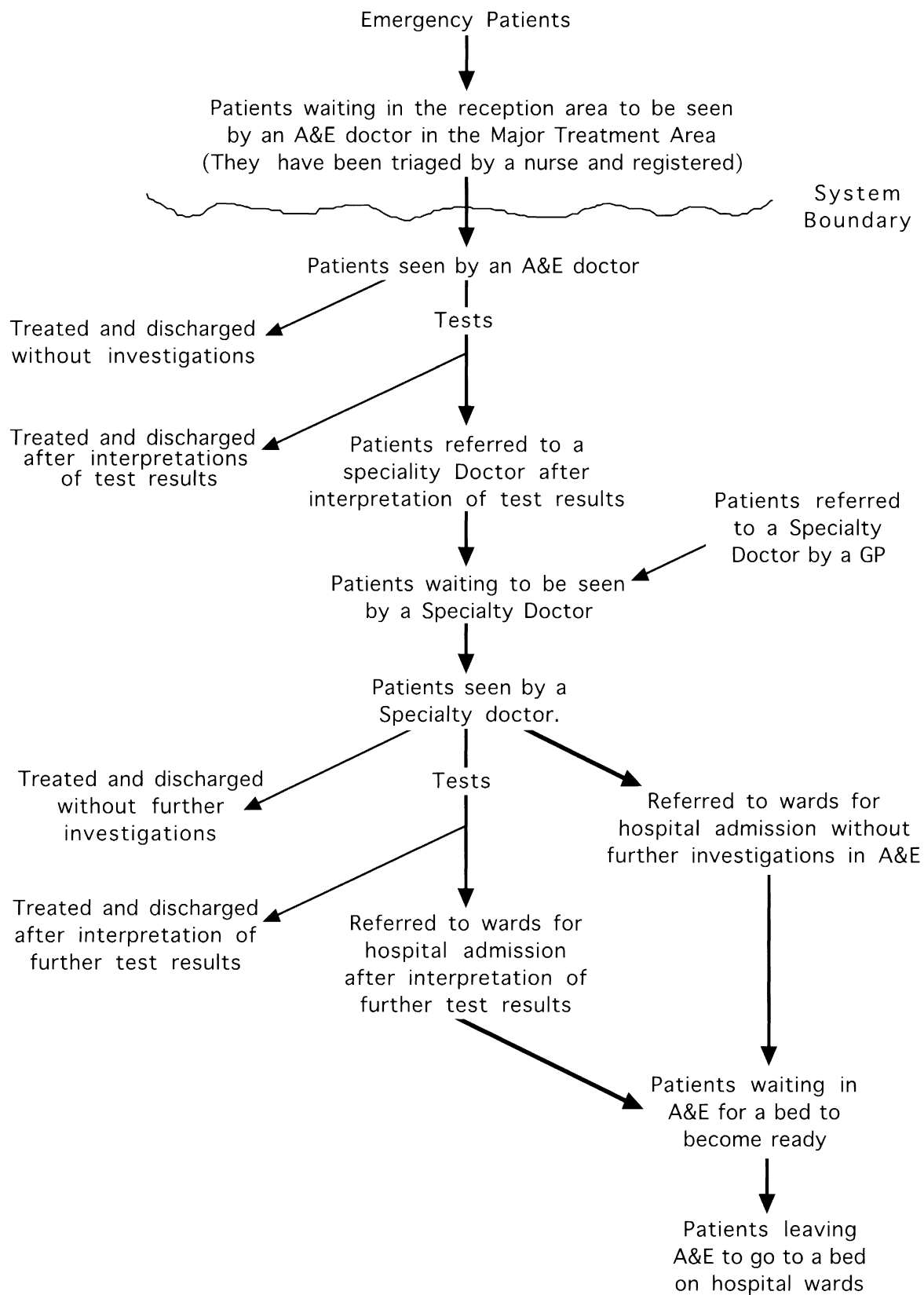


Figure 1: Schematic representation of A&E elements, processes and pathways included in the system dynamics model.

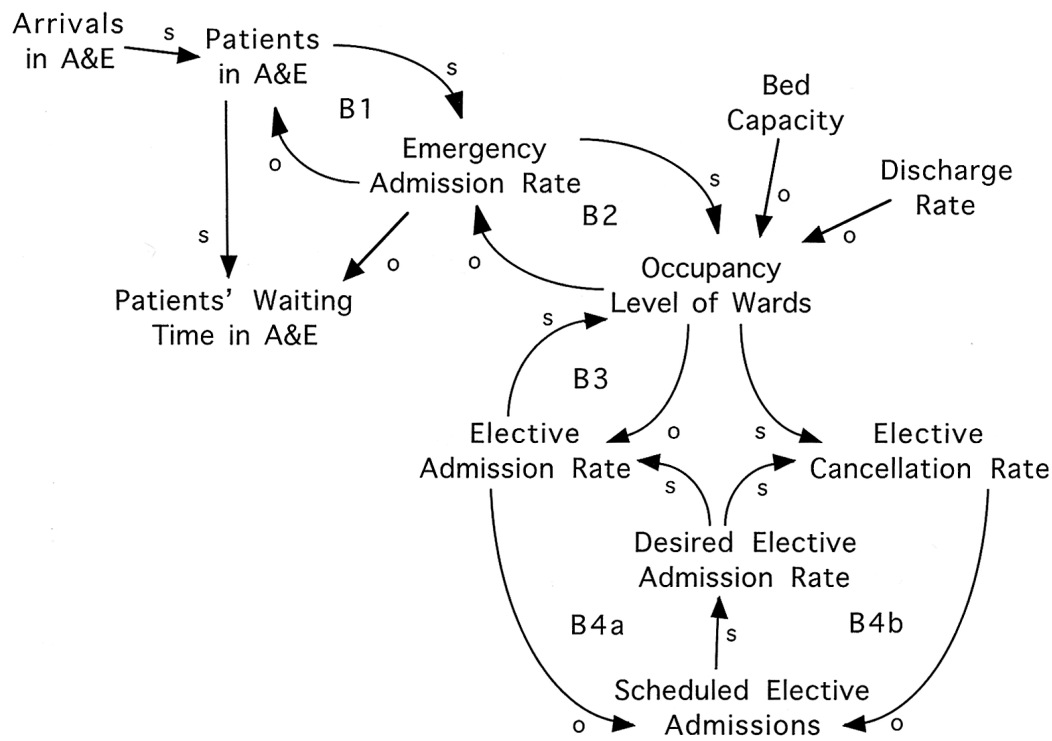


Figure 2: Causal loop diagram of the main effects determining waiting times in an A&E department. The polarities of the causal links read as:

s = variables move in the same direction, ceteris paribus o = variables move in opposite directions, ceteris paribus

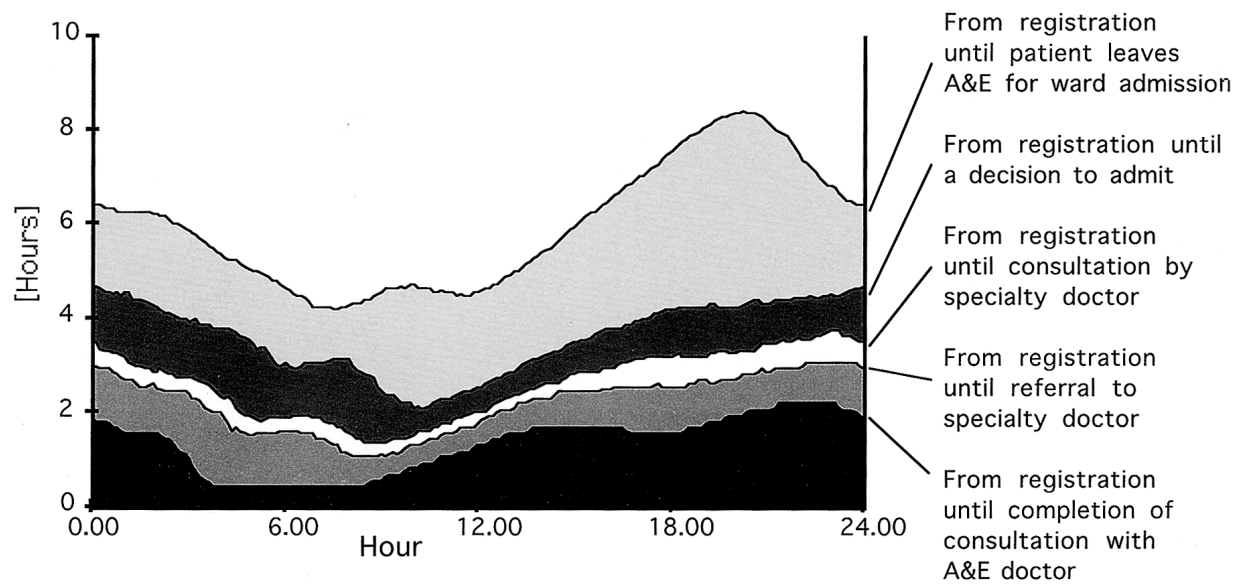


Figure 3: Emergency patient waiting time to reach different stages in A&E.

Beds	Demand	Avg. Total Waiting Time [hours]*	Elective Cancellations [%]	Bed Occupancy [%]	A&E Dr. Utilisation [%]
700	Standard	5.9	30.4	95	92
800 (base)	Standard	5.9	16.2	94	92
900	Standard	5.9	7.8	90	92

800	Standard +2%	6.2	16.8	95	96
800	Standard +4%	6.6	17.4	97	99
800	Standard +5%+	∞	n/a	n/a	n/a

∞ (i.e. tends to infinity)

Table 1. Performance measures for model policy analysis runs

*Total waiting time measures elapsed time from entry to A&E until admission to a bed as an inpatient

consists of, first, carrying out a base case simulation - running the model with data and parameters representing the most recent experience of the real world system; and second, conducting a range of policy analysis runs - systematically changing parameters or relationships of interest to explore their effects on model outputs. This process and its results can only be summarised here; a more extensive paper is available describing the analysis in more detail (Lane, et al, submitted).

The base case simulation demonstrated the strong effect of time of day on total waiting time. This was the result partly of the variation in arrival rate, but also of the varying number of A&E specialists on the roster. Figure 3 shows the daily pattern of the several components of total waiting time. It is noticeable, for example, that (under the conditions of this run) the initial period of waiting to be seen by an A&E doctor is about one quarter to one third of the total. However the continual reduction in the numbers of A&E doctors on duty in the evening and in the early morning produces delays of two hours for most of the day. It is not until the early hours of the morning that this time drops appreciably, and this lasts but a few hours before the new day's onrush of arrivals pushes the delay back up. Thus the management of the hospital's staff resources can have a significant impact on the effectiveness of the service delivered to patients.

More striking are the lessons of the policy analytic runs with the model. In these explorations, the model was run with different bed capacities, and with different demand levels. For the former, we looked at the effect of operating a 700 bed or 900 bed hospital, rather than the 800 beds of the base case. The demand scenarios considered a range of possible percentage increases on the base case demand level. (Reductions in demand were considered an unrealistic prospect.) The results of these simulations are shown in Table 1. (All the results are average performance measures.)

The first three runs show that reducing the number of beds does not influence the average total waiting time, at least at the levels of patient pressure on resources at which St Dane's is operating. Strikingly, the main strain is taken by cancellations of elective admissions; with 100 beds fewer, the percentage of elective cancellations almost doubles.

This outcome, surprising at first glance, is explicable in terms of the hospital management's operating policies as captured by the model. In terms of the causal loop diagram of Figure 2, loop 5 serves as a safety valve, preserving the relative priority of emergency patients and preventing the final stage in their delay process (waiting for a vacant bed) from escalating. When there are bed shortages, scheduled admissions suffer disproportionately.

A very different picture emerges when increases in the level of demand from emergency patients are considered. Without any bed closures, an increase of 2% in emergency admissions pushes up the average waiting time in casualty only a little; but a further 2% generates in total an extra three quarters of an hour before admission to a hospital bed. (It is worth noting that the most recent year-on-year increase at St Dane's was 7%.) Most ominously, the doctors in A&E are now working almost flat out. There is no slack capacity left to handle any further increased flow. The simulation run with a 5% increase in emergency admissions demonstrated this graphically - the backlog of patients awaiting initial consultation with an A&E doctor accumulates without bound, and the waiting time escalates from hour to hour. The system cannot achieve a steady state, and the performance indicators therefore become incalculable.

These results make the overall message more complex. With the system, however precariously, in balance, bed shortages affect elective cancellations rather than waiting times for emergency patients. Emergency patients are given priority. However, further increases in demand on A&E (and the demand is growing inexorably year by year) threaten a more catastrophic collapse.

A range of other simulations of the system centred on A&E have been run. For example, we have tried out a scenario combining bed closures with increased demand - which not surprisingly produces worse performance than either of the factors separately. Many more could be carried out. Strictly speaking, the results from such policy explorations apply only to St Dane's, whose data has been used to calibrate the model. However there are some clear general implications. There are conclusions from the model which support and are supported by evidence from the acute sector of the health service as a whole. And there are general lessons about the way in which policy may, or may not, be safely driven by performance indicators.

Policy lessons

We have recently lived through a general election at which the victorious party promised to reduce waiting lists but a few months on found that the opposite was happening. More recently still the new government, unshakeably committed to the low spending commitments of its predecessor, has had to

round up an extra £300 million to avoid an impending disaster in A&E departments this winter. Evidently in the short run there is no alternative. Yet waiting lists, waiting times in A&E, and cancelled operations do not exist in parallel universes. They are all aspects of health service performance, and their linkage needs to inform any longer-term strategy.

The study we have carried out is still at an early stage. However we believe that there are already some clear lessons for policy formation:

- 1) The staff and bed stock of our acute hospitals are already being utilised at an intensity which can be argued to be counterproductive. There is virtually no slack left at least in this sector of the NHS out of which further 'efficiency savings' can be squeezed. In some hospitals skilled bed management may still have some contribution to make. However there is already so little slack in the system that its ability to cope with normal day to day fluctuations, let alone major incidents, without disproportionate disruption is severely compromised. There is no reasonable prospect of reaching the Patient's Charter standards for waiting time. Further bed closures would cause further deterioration.
- 2) There is some scope for achieving improvements in the fit of the deployment of A&E resources, particularly specialist A&E doctors, to the daily profile of demand for emergency services. Doctor rosters can both cause increases in the average delay before emergency patients' first consultation, and under conditions of bed scarcity generate knock-on effects later in the admission process. The potential for improvement will be extremely limited, however, without additional staffing.
- 3) In a complex system under stress, hitching policy to a single headline performance indicator succeeds only in transferring the problem. Give priority to long stayers on waiting lists, and more serious cases are delayed. Give priority to emergency patients, and the strain is taken by cancellations of elective admissions. Adopting a single indicator can give the illusion of progress, as in the case of waiting list reductions. However some of those improvements will be illusory, given the managerial temptation to respond to perverse incentives (playing the numbers game);

while real improvements will result in a game of musical chairs. The next time the music stops, some other aspect of health service performance will be found wanting. Should reductions in A&E waiting times be bought at the expense of headlines such as “Girl’s Heart Surgery Cancelled Five Times”? There needs to be a coherent use of multiple indicators and a more sophisticated awareness, in practice, of their interconnections.

- 4) We have lived through 18 years in which planning and analysis have been downgraded in favour of the virtues of competitive entrepreneurship linked through a market mechanism. In the public sector where real markets are hard to engineer, the government installed control processes in which performance indicators stood in for the non-existent price signals. This response to the overselling of planning in the postwar period was a clear over-reaction. But in any case the market mechanism has at most a limited role to play in managing complex systemic interactions of the type we have been considering - it is far too blunt an instrument. A government without ideological blinkers needs to re-instate analysis in an appropriate but close relationship with the policy-making process.

Forrester J W (1968) *Principles of Systems*, MIT Press, Cambridge, MA.

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

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

Lane D C (1997a) ‘Invited Review and Reappraisal: ‘Industrial Dynamics’ by Jay W Forrester’, *Journal of the Operational Research Society* 48, 1037-1042.

Lane, D C (1997b) ‘Invited Reviews on System Dynamics’, *Journal of the Operational Research Society* 48, 1254-1257.

Lane D C, Monfeldt C and Rosenhead J V (Submitted) ‘Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department’, *Journal of the Operational Research Society*

Taylor, K S and Lane D C (1998) ‘Simulation Applied to Health Services: opportunities for applying the system dynamics approach’. *Journal of Health Services Research and Policy* 3, 1-7.

Acknowledgements

We are grateful for the assistance both of Malcolm Alexander of Casualty Watch and the Southwark Community Health Council, and of many members of staff of ‘St Dane’s’.

For the interested reader

Audit Commission (1996) *By Accident or Design: Improving A&E Services in England and Wales*, HMSO, London.

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

Department of Health (1996) *The Patient’s Charter*, HMSO, London.

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

Forrester J W (1961) *Industrial Dynamics*, MIT Press, Cambridge, MA.