Table of Contents

Go Back

Using System Dynamics in Modelling Health and Social Care Commissioning in the UK

Professor Eric Wolstenholme, Douglas McKelvie, Gill Smith and David Monk

OLM Consulting, Cairns House, 10 Station Road, Teddington, Middlesex, TW11 9AA

All of the authours can be contacted on 020 8973 1179

eric.wolstenholme@olmgroup.com

douglas.mckelvie@olmgroup.com

gill.smith@olmgroup.com

david.monk@olmgroup.com

Abstract

Over the past two years OLM Consulting, initially in partnership with Cognitus, have used System Dynamics (SD) modelling in a wide range of health and social care settings to shed light on a number of difficult and complex issues and to influence and interpret health and social care policy in the UK. This work has been instrumental in causing health legislation to be modified in the Upper House of Parliament as well as helping local health communities implement sustainable performance improvement.

This paper describes the work done in 2003 with two local health economies. It shows the commissioning models that resulted from applying a nationally-developed template in a local context, as well as some of the findings obtained from running those models. The emphasis has been on demonstrating strategies that achieve efficiency improvements for all agencies across whole patient pathways. By modelling whole pathways from primary care through acute care to post acute care, and focusing on admission prevention and delayed discharges, it has been possible to show that significant resources can be saved within agencies along the pathways, without influencing performance.

A glossary of terms is provided at the end of the paper.

Context of the Work

The socio-political dimensions of this type of work are complex (Wolstenholme et al, 2004). Joint planning in health and social care is everywhere; government correctly exhorts those responsible for working together locally to adopt a “whole systems” approach to that task. There is a need to develop and adopt suitable tools to aid this process. OLM Consulting and Cognitus initially built a number of templates and models to simulate patient pathways using information provided by national agencies. From these, a profile of a notional local health economy, based around a fictional Primary Care Trust (PCT), district general hospital and Social Services Department (SSD), was distilled. The initial focus was on demonstrating that “delayed discharge” (a key issue in the National Health Service (NHS) currently) is just one aspect of the behaviour of a whole commissioning system. The next step was to populate the template with data from an actual health economy, and in the past nine months, a number of agencies have worked with OLM to adapt the model template to suit local circumstances.

Probably few service planners in social services use computer simulation. Those that do not, therefore, probably do not realise that a model can be a powerful tool for gaining understanding of the kind of complex systems that health and social care planners need to comprehend. Without experience of models, they are unlikely to commission much model-building, or to recognise when they are dealing with problems that could be better understood through the medium of a model. Models are more prevalent in health care, but perhaps mainly applied to discrete parts rather than whole processes.

At the same time, most managers have technology that can run whole systems models sitting on their own desktops.

Perhaps ironically, managers who are also qualified social workers will almost certainly be familiar with versions of systems theory, and will have studied organisation theory. Social workers are trained to take an interest in process, to look at the structure of communications rather than just the content. So an approach to strategic planning that moves away from a simple concern with objectives and targets, and focuses on understanding and improving operational processes, should be attractive to them.

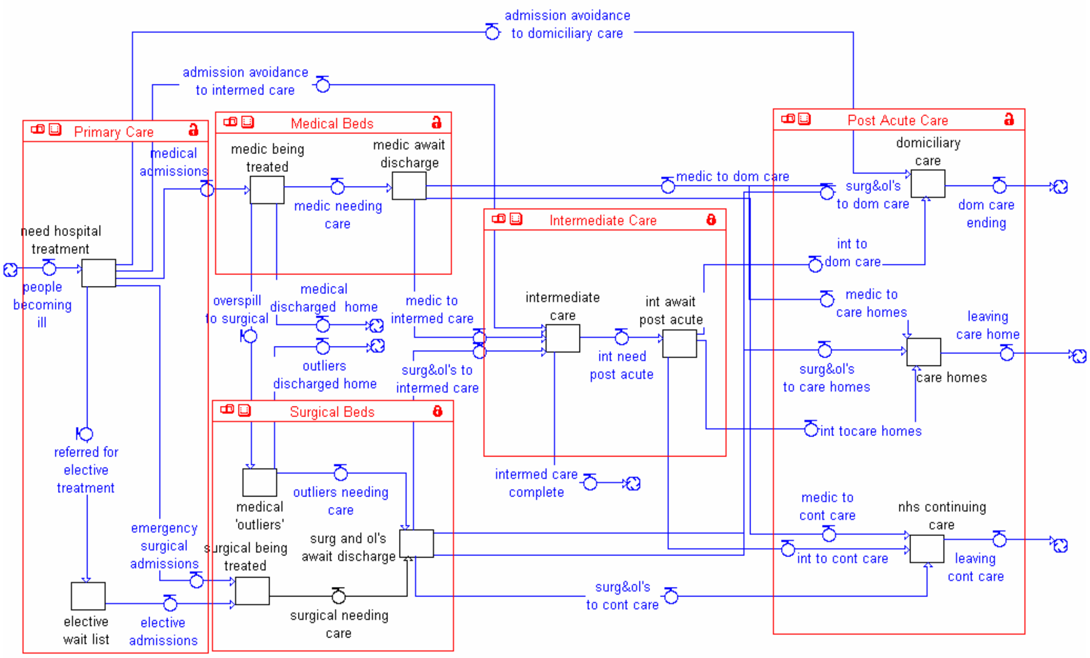
Moreover, the use of models is commonplace in other settings where experimentation on the real-world is impractical or dangerous. Model-building is a key stage in the design of new products and simulation is a normal part of learning new skills (including social work). A new shape of aeroplane would not be trusted without testing models of it in a wind-tunnel. It would also be surprising if airline pilots did not spend substantial amounts of time in flight-simulators before taking to the real sky. But somehow, when health and social service managers are trying out a new plan, or considering making major changes in a complex network of care services, they rarely think about building a model to check assumptions about how things might operate.

The modelling described here uses the ***ithink***© system dynamics software and is mainly concerned with finding and testing policies relating to long patient pathways across multiple agency boundaries, seeking to improve the performance of all agencies. This approach is truly "whole systems thinking" and contrasts strongly with the more usual unilateral actions of the more powerful agencies. The models are used to demonstrate the effects of policies on all performance measures over time and the models are equipped with easy-to-use interfaces containing a range of input devices (slider bars and buttons) and output devices (graphs and tables) to facilitate this end.

The Commissioning Model Described

OLM Consulting started with the model of a fictional health economy referred to above and, over the past year, has worked with real health and social care partnerships, to adapt the model to suit local circumstances and run real data. Each application has resulted in a new template which has been used as a starting point for further applications. Figure 1 shows a highly simplified overview of the latest template.

Figure 1: Simplified Version of Model



Description of the Model

Complexity arises by the interaction of the proportion of patients flowing, their lengths of stay in, and the capacities of, each pathway. This is exacerbated by the effects of actions taken by the agencies owning each piece of the pathway, often influenced by the way in which patient movements impact on their own organisation's resources and performance targets. The resultant behaviour of the whole system over time can be very counter-intuitive, which is a major justification for the use of computer analysis to clarify the structure and improve understanding and communication.

Primary Care

This sector represents only patients who are becoming ill and will be referred for hospital treatment. There are three main types of admission: medical, emergency surgical and elective surgical. It is also possible for some admissions to be avoided or diverted. If space allows, a proportion of those needing admission can instead receive certain forms of intermediate care, or a domiciliary care package. This is normally at the expense of the resource being allocated to a patient awaiting discharge.

Medical Beds

The model uses the shorthand term "beds" to mean essentially the total number of hospital places of a particular kind. In reality, hospital capacity is constrained by a number of factors, including availability of staff and operating theatre time. Similarly, the concept of hospital "capacity" is simply a number, expressed in terms of so many beds. This is set at the acute hospital's intended level of occupancy (somewhere between 85% and 95% of the number of beds available).

In the model (but not shown in this diagram), there are various pathways through medical beds representing patients with differing characteristics (those with relatively straightforward conditions and few onward care needs, those with more serious conditions and who are more likely to have onward care needs, and older people with mental health problems who are a significant sub-group of patients within medical beds with very specific onward care needs). These pathways cluster the medical admissions into groups, each having different characteristics with respect to length of stay, treatment time and assessment time.

There is a fixed capacity of medical beds. The model will allow patients to be admitted only if space is available. If medical beds are under pressure, two things happen. First, some patients are discharged home earlier than they ideally should be. In the short term, this creates some space for new admissions. However, since a higher proportion of those discharged early will need to be readmitted, there is a limit to how effective this strategy will be. Second, if there is spare surgical capacity, a number of medical admissions will be admitted to and treated in surgical beds, as "medical outliers". Again, this has the effect of dealing with the immediate problem of patients needing to be admitted, but at a cost, because operations may have to be cancelled and fewer elective surgical patients are admitted.

Once patients have been treated, if they have no onward care needs, they go home. Those with onward care needs, and who therefore cannot be discharged unless a further service is available, go into the "await discharge" stock. They will remain in that stock, occupying a hospital bed, until the correct (intermediate or post-acute) service becomes available for them.

Surgical Beds

There are three kinds of surgical admission.

Emergency surgical patients are always admitted, regardless of how much spare capacity exists. This means that at times the hospital will exceed its target occupancy rate, which resembles reality for the current users of the model.

Those needing **elective surgery** go on the waiting list. They are admitted as surgical beds become available. One of the buffers in the whole system, therefore, is the elective waiting list, which increases when the hospital is full, and reduces when the hospital has spare capacity.

Medical outliers are medical patients occupying surgical beds. These admissions take place only when there are no more medical beds available but there are surgical beds.

Patients are admitted to surgical beds according to this order of priority: emergency patients (always admitted), medical outliers (if medical beds are full and there are spare surgical beds), and then elective patients (if there are spare surgical beds). As with medical beds, once patients have completed their treatment, they either go home, or, if in need of onward care, wait in an acute bed until a suitable resource is available.

Intermediate Care

In this simplified version, there is one entity called "intermediate care". The actual model represents a number of services, some of which are intended as alternatives to admission, and others that assist timely discharge. Some intermediate care services do both. Intermediate care is time-limited. On completing a period of intermediate care, patients/service users either go home with no further service, or are discharged with a post-acute package. The latter might involve waiting in intermediate care until the post-acute care is available.

Post-Acute Care

There are three main types of post-acute care, each with a separate capacity: domiciliary care, care homes (whether residential or nursing), and NHS continuing care. Patients are discharged to these services either directly from acute hospital or from intermediate care. Some are referred directly to the domiciliary care service as an alternative to acute admission.

Each post-acute service has a fixed capacity. Vacancies become available according to an average length of stay (care home or continuing care) or duration of package (domiciliary care). This figure includes those whose care package ends only on their death.

The Model and Data

There are two main challenges when implementing the model locally. Firstly, there is a need to secure agreement about the structure of the model, which means mapping out in some detail (normally directly on **ithink©)** the main care pathways and especially how they connect across service types. This may well turn out to be revelatory for some managers, who may not fully understand how some connected services actually operate.

Secondly, there is a need to attempt to populate the model with real data. This turns out to be particularly challenging, because the (extensive) categories of data that are currently collected, normally as required by government, do not neatly match the data categories required by the model. Again, the modelling exercise can be a revelation to managers about the data they **really** need to manage a process. Interestingly, the modelling task is not driven by data. The starting point is to focus on operational processes, then use this understanding to specify what data is required. In fact, most of the model can be built without real data, concentrating on mapping processes. Mapping using stocks and flows is a more rigorous method than other types of process mapping. Stocks and flows force specific decisions about operational realities. Once participants are satisfied with the structure of the model, the focus shifts to finding the right data.

The model requires in-depth understanding of data. There are some data items that are simple “inputs” to the whole system, and which **can** be entered into the model as such. These would include:

- The capacity of each service
- Average treatment or service lengths
- Demand data at the entry point to the whole system (i.e. numbers of people requiring acute care)
- The percentage of people using a particular service who will go on to have a need for a further type of service

However, most data items are not in themselves “inputs” to the whole system. They are indicators of how the whole system is operating but they do not drive it. A useful learning point is that, even where agencies have accurate data for a particular part of the process, it is not always possible simply to put these numbers into the model. For example, the daily admission rate to intermediate care may be known, but it is not an input. It should be clear from the diagram that once the model is running, the rate of admissions to intermediate care is a product of the rate at which people are being referred to intermediate care (mainly) from the acute hospital, and the rate at which people are leaving intermediate care. And that rate is itself dependent on the availability of some post-acute services. The admission rate is therefore used, not as an input, but to check that the model is correctly deriving this figure.

As the model is running, variables are checked to ensure that their behaviour within the model corresponds to how they “really” behave.

In general, agencies hold more data about stocks (such as how many people there are in most categories at various points in time), when it would be more useful to know about flows (that is, the rates at which people move between different stages in a process).

The model reveals an absence of data that is absolutely critical for joint planning to improve the hospital discharge process. For example, there is very detailed information about “lengths of stay” in hospital for all categories of patient but the model shows that it is equally important to know:

- Length of stay up to the point where a patient is deemed ready for discharge (which would be a model input called “treatment time”), and
- The length of time spent awaiting discharge (which would be a model output that would vary according to whether there is space available in a given post-acute service).

Other useful data (currently unavailable) are the proportions of patients being discharged to the main post-acute options of domiciliary care, care homes and NHS continuing care. These categories do not neatly map against the standard discharge data available.

The Modelling Process and Joint Planning

The model provides a focus for discussion between managers and other stakeholders with operational responsibilities for the various parts of the whole system. The modelling process itself contributes to this dialogue. A typical modelling project will consist of a number of stages:

- Formation of a strategic inter-agency group to oversee the model-building process
- Initial workshops with a broader, operational group to test out assumptions about the main model elements, concentrating on care pathways
- Specifying/agreeing the model structure (or adaptations that might be needed to a pre-existing model template)
- Detailed work with a range of operational staff and those responsible for data / information management to confirm the detailed structure of the model, and to

determine the best data sources, or proxy-measures in the event of certain data being unavailable

- Building the revised model and entering data
- Testing the revised model with a range of operational and information management staff
- Using the model as the basis for strategic scenario-testing exercises

In the case of the commissioning model, an established model template has been introduced initially and then customised to suit local circumstances. Some of the main changes that have been made to the original template have included: the addition of resources and patient pathways for older people with mental health issues, more detailed representation of intermediate care, and more detailed representation of hospital admission processes.

The commissioning model simulates the daily operation of the whole system outlined above over a period (normally three years is seen as a useful planning horizon) and incorporates three periods of winter pressure. These cause the system to be under capacity in the summer, but over capacity in the winter. Although the predictive element of modelling can be compelling, the modelling exercise is as much about constructing a rich learning environment for informing stake-holders about how a complex system operates, and how the various parts of that system fit together. Participants probably learn as much from the development stage of the model as they do from testing scenarios on the final product.

Some Results Using the Commissioning Model

The model shows that the whole system of acute, intermediate and post-acute care in a given locality can be characterised as a series of services each having its own capacity, pattern of usage and average duration of care/treatment. Whenever there is a mismatch between the rate at which people finish one service-stage, and the rate at which places become available at the next one, there will be problems such as delayed discharges. Throughout the system, there are a number of buffers, bottlenecks, and ways of dealing with pressure. Even without running any data, the model structure itself makes explicit the complexities of planning health and care services, and the challenges facing planners and managers.

The model was initially set up with a number of fixed runs, to introduce people to the range of experiments that yielded useful insights into the behaviour of the whole system. From there, they were encouraged to devise their own runs and develop their own theories of useful interventions and commissioning strategies. The fixed runs are described on the following pages.

Base Run

In this run, the economy is "stretched" by seasonal demand for admissions. It is barely coping, since elective wait times are significantly above target (Figure 2). Delayed discharges are also a cause for concern, and would result in the local authority having to "reimburse" the health sector more than £3 million over the three-year period.

Fig 2: Base Run – elective wait time

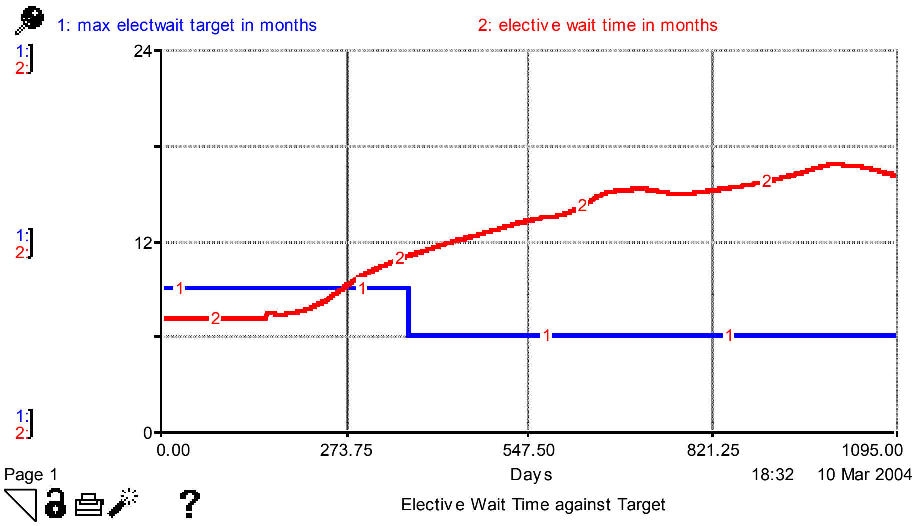
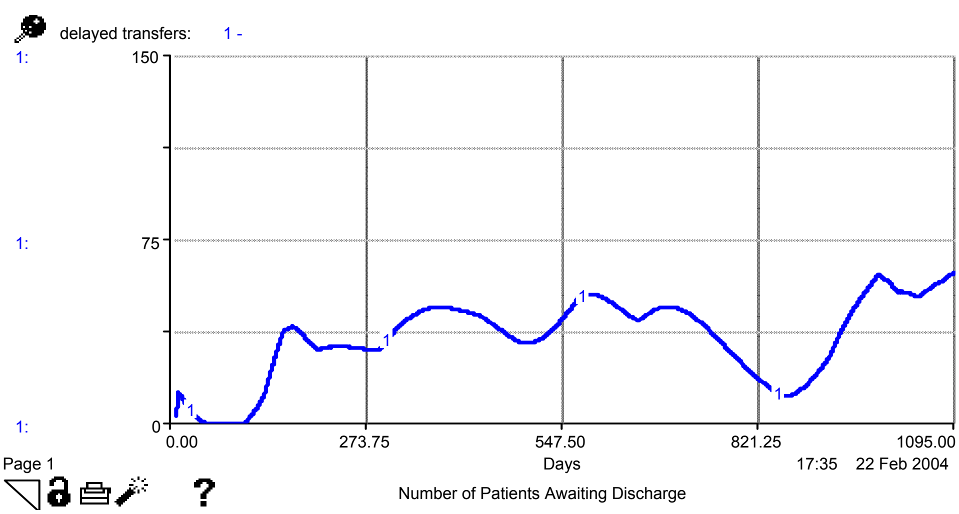


Fig 3: Base Run – Delayed Discharges



Adding 10% Acute Capacity

The acute trust's likely response to the base run would be to try and find additional in-patient capacity, to bring elective wait times down. In this run, they add 10% across the board (medical and surgical beds) from day 250. This has some beneficial effect on the elective wait times (Figure 4), but dramatically increases delayed discharges (Figure 5).

Fig 4: Acute Run – elective wait time

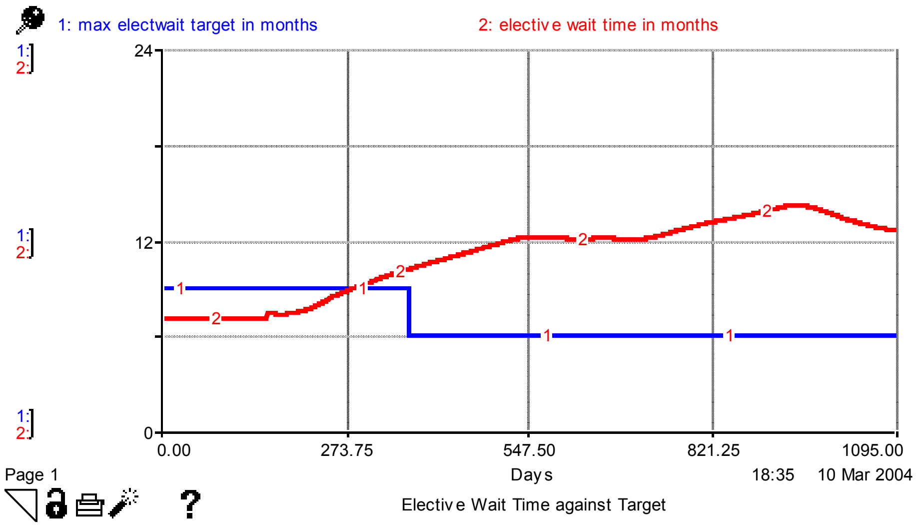
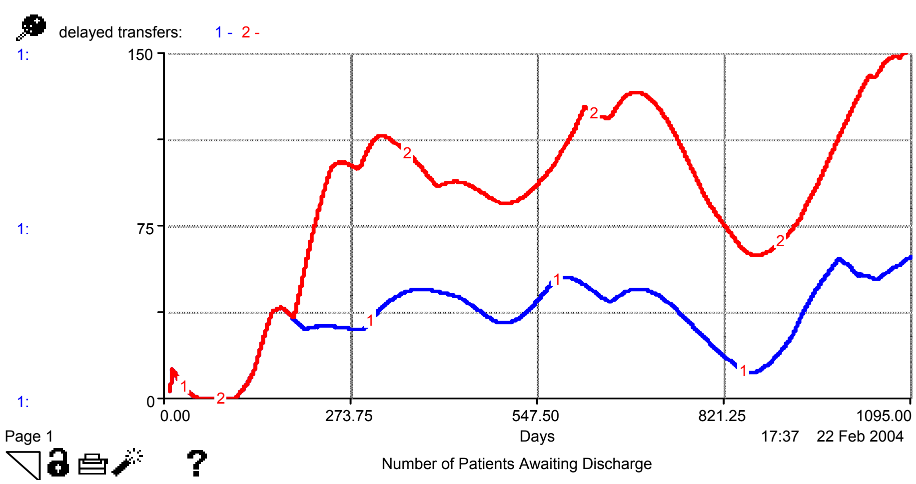


Fig 5: Acute Run – delayed discharge



Line 2 is delayed discharge resulting from 10% additional acute beds

Adding 10% Post-acute Capacity

A more considered response to the initial problem might have been to improve throughput (by adding post-acute capacity to alleviate delayed discharge). This has very good results for both elective and delayed discharge (Figures 6 and 7).

Fig 6: Post-acute Run – elective wait time

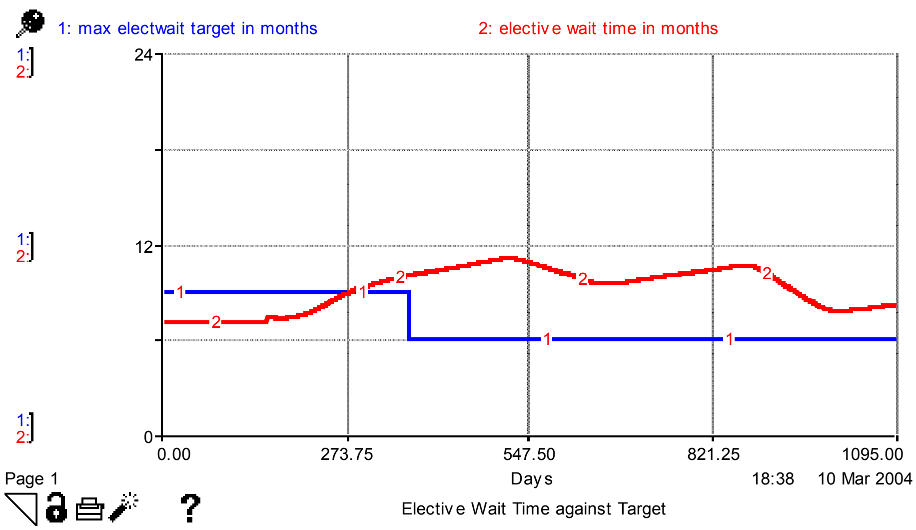
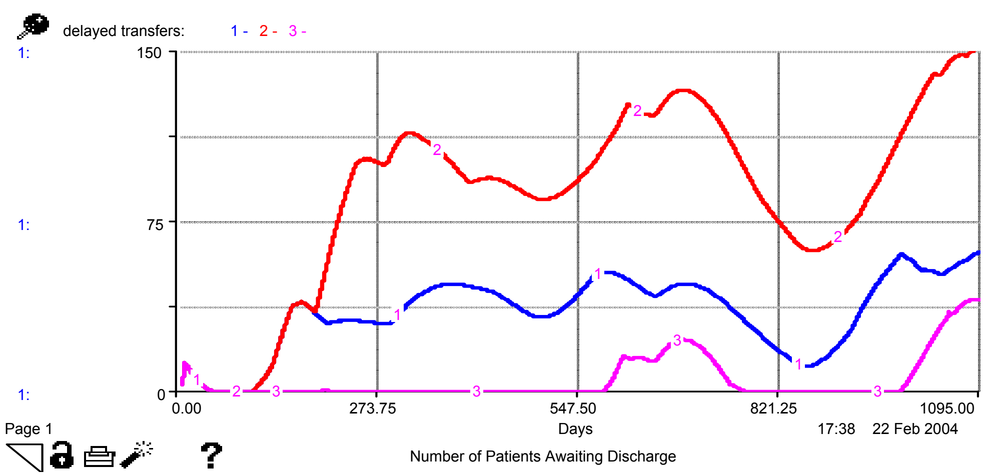


Fig 7: Post-acute Run – delayed discharge



Line 3 is delayed discharge after 10% post-acute capacity added.

Adding 10% Capacity for both Acute and Post-acute

In order to see whether the beneficial effects of adding acute capacity and post-acute capacity are summative, both capacity increases are tried simultaneously (a very expensive investment). Elective wait times are improved further (Figure 8). However, delayed discharge results are disappointing: adding all this extra capacity has actually worsened the situation (compare line 4 to line 1 in Figure 9).

Fig 8: Acute and Post-acute Run – elective wait time

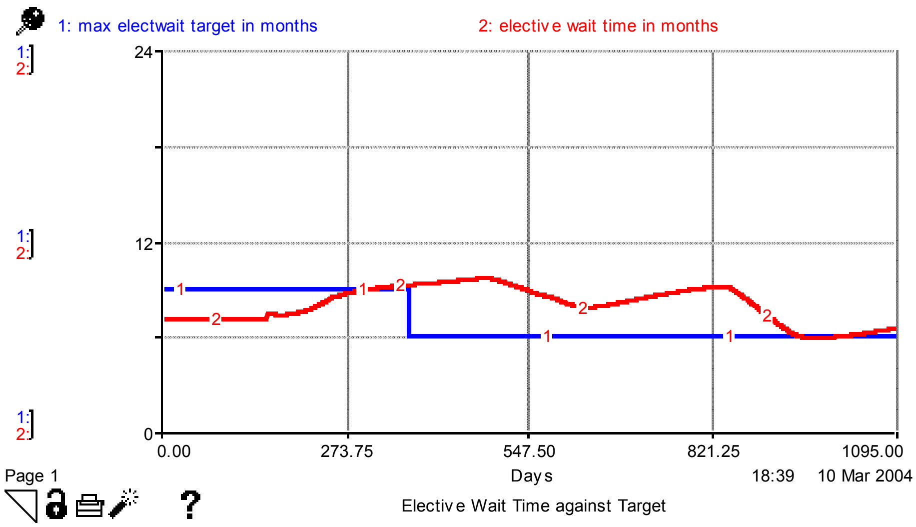
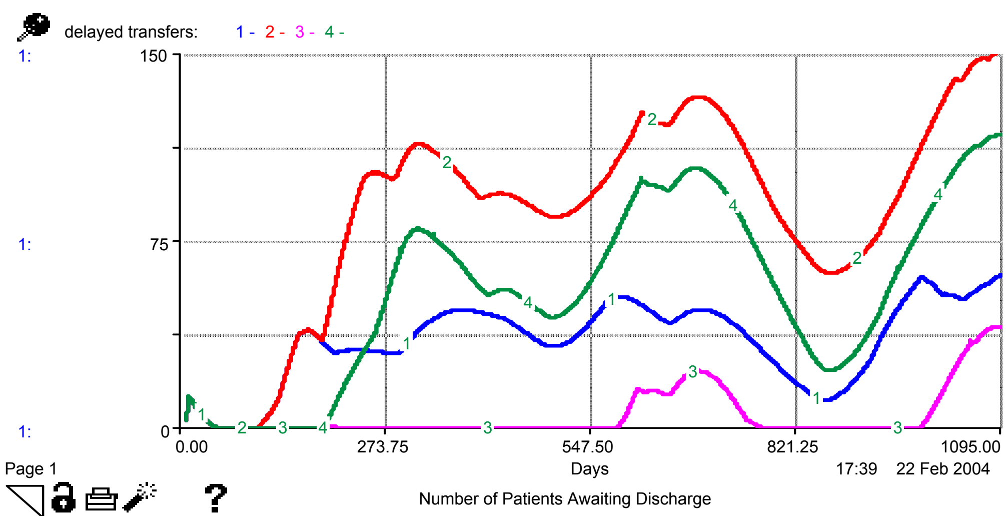


Fig 9: Acute and Post-acute Run – delayed discharge



Line 4 is delayed discharge after 10% acute and post-acute capacity added.

Diverting 5% Admissions within Primary Care

Instead of adding capacity, this run tests the principle that it is better to prevent people getting into hospital than to make capacity available to accommodate them. The percentage diversion is modest (5%), but results are good for elective wait times (Figure 10) and similar to base run for delayed discharge (line 5, Figure 11).

Fig 10: Diversion Run – elective wait time

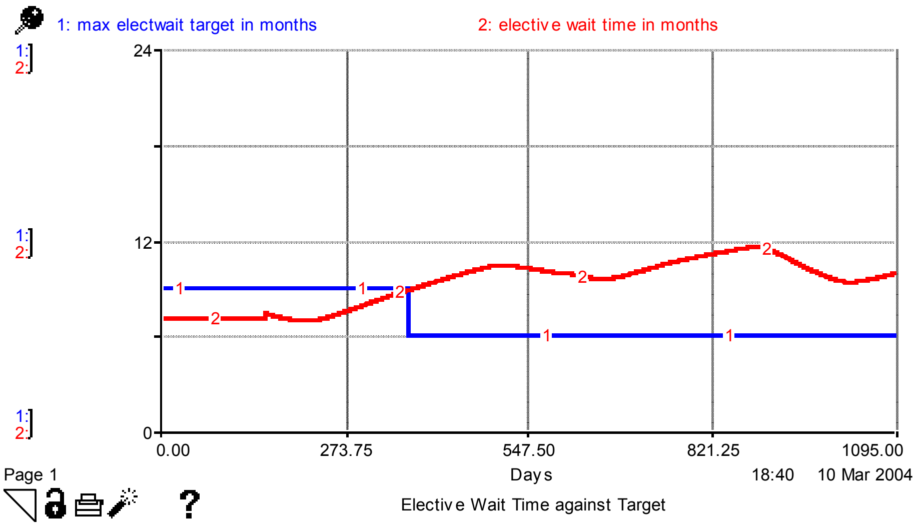
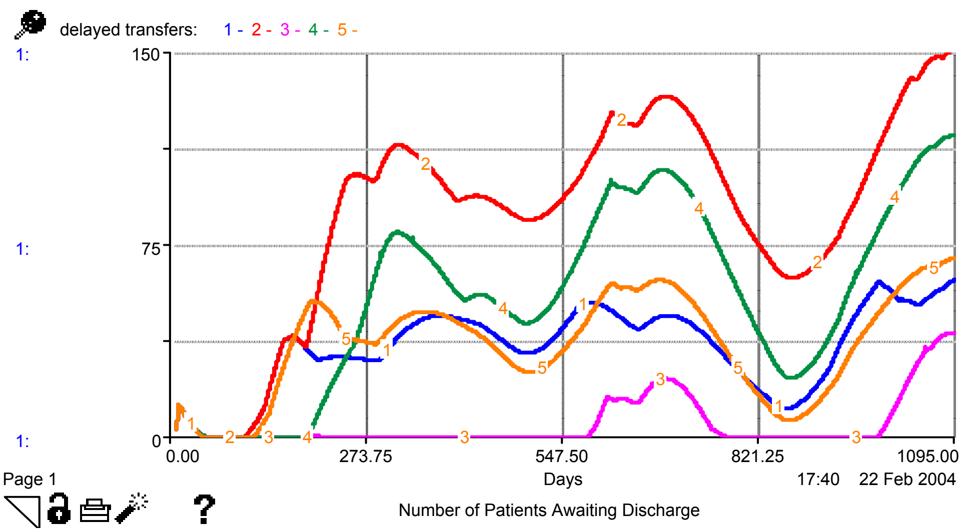


Fig 11: Diversion Run – delayed discharge



Line 5 is delayed discharge with 5% diversion

Diverting 5% Admissions and Adding 10% Additional Domiciliary Care

Since the base run indicated that the post-acute sector was under pressure, this run explores the benefit of relieving that pressure with 10% additional post-acute resource of one type – domiciliary care (at the same time as diverting 5% patients from admission). Domiciliary care is home-based and one of the most flexible forms of post-acute care, since it can be provided to prevent admission or to facilitate discharge. Results are encouraging (Figures 12 and 13).

Fig 12: Diversion and Dom Care Run – elective wait time

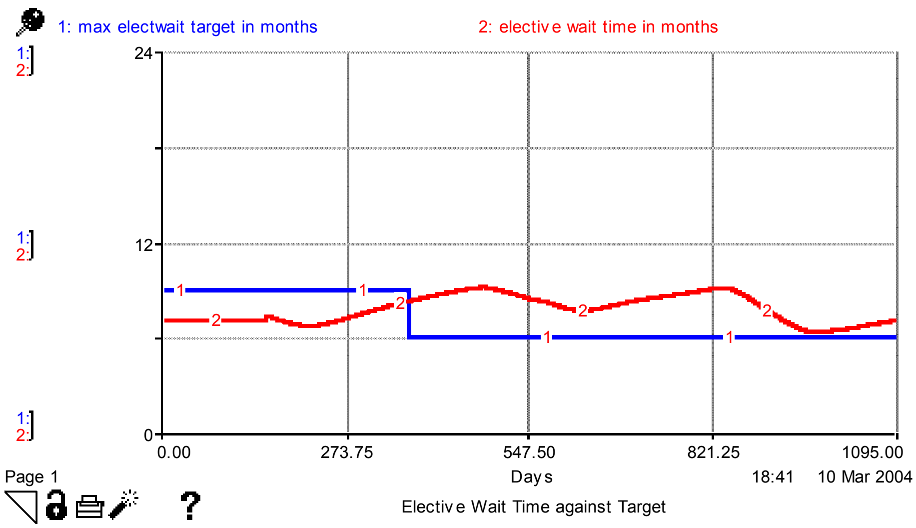
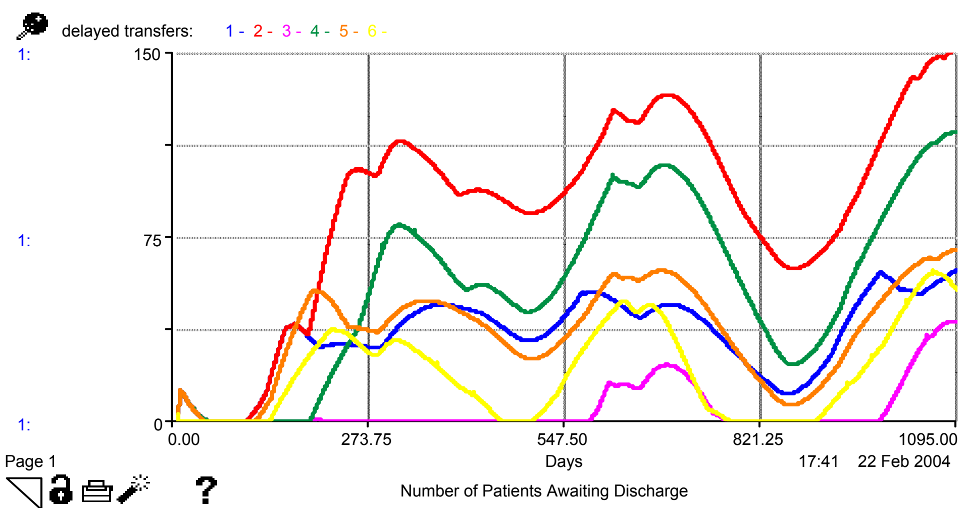


Fig 13: Diversion and Dom Care Run – delayed discharge

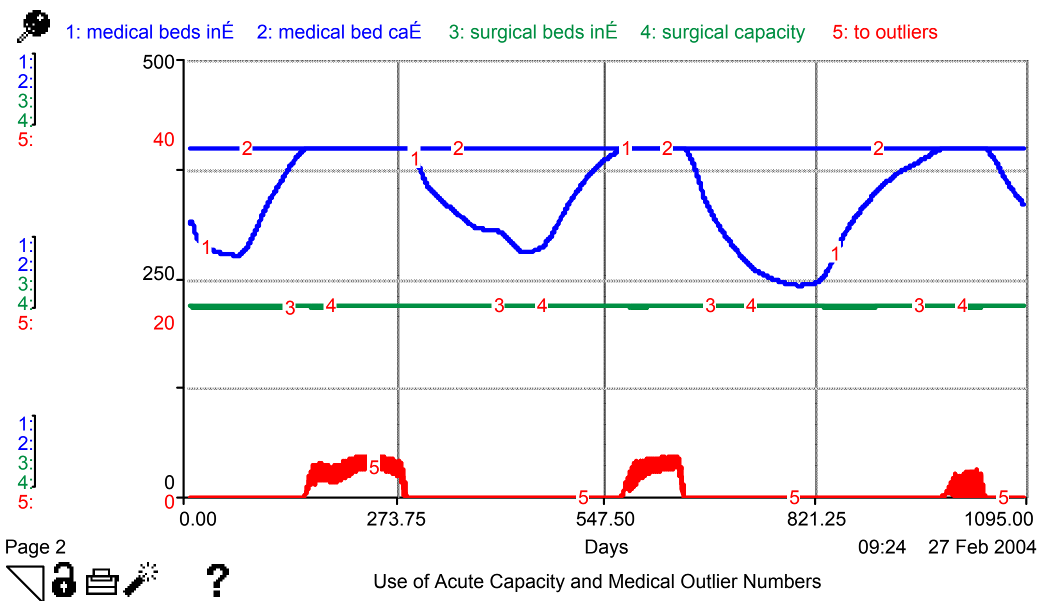


Line 6 is delayed discharge with 5% diversion and 10% additional domiciliary care

Interpreting the Results: Some Learning Points

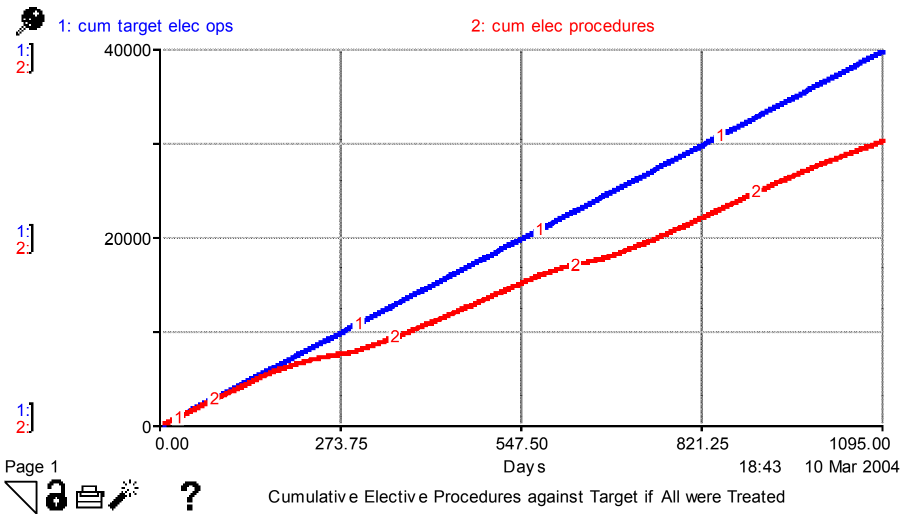
The model has a large number of sliders for varying inputs and a large number of graphs for assisting in interpreting results. The section above shows the graphs that are most useful for a quick diagnosis of useful interventions. However, another useful graph is the one which shows the loading on medical beds, and illustrates how surges in medical demand result in overspill into surgical beds (medical outliers). In the base run, demand for medical admission regularly outstrips spare beds (lines 1 and 2 in Figure 14).

Fig 14: Effect of seasonal demand on medical capacity



As the demand "bumps up against" capacity (line 2), then medical outliers (line 5) are created. Note that surgical capacity (line 3) is fully committed (line 4) at all times, so any additional demand (from medical outliers) will mean that operations have to be cancelled. This is what is causing the problem with the elective wait time. Further evidence is provided by a graph of elective activity (Figure 15).

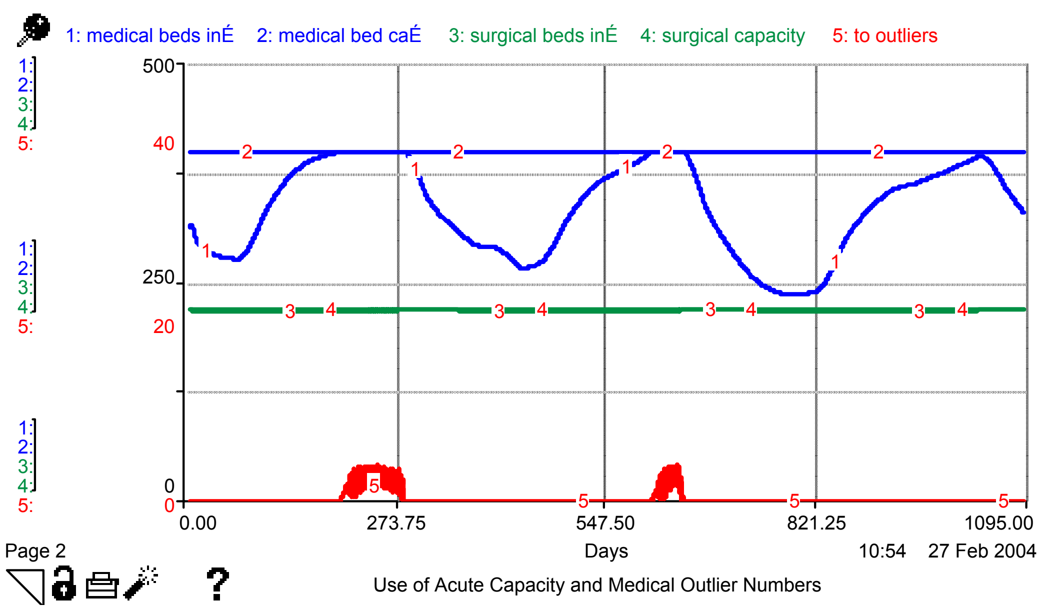
Fig 15: Cumulative surgical procedures performed



This graph shows the cumulative number of elective procedures compared with the "target" number, which is the total that would result if everyone who needed treatment was treated. This is essentially another way of illustrating how the actual total of surgical procedures is affected by medical outliers being admitted to surgical beds; the line falls away from the target at these times.

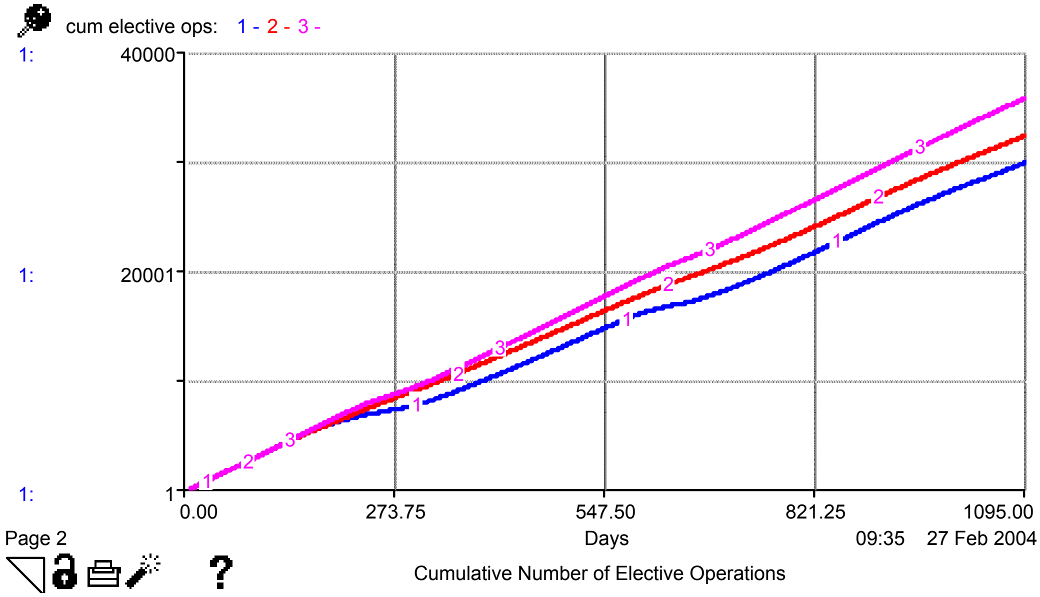
These two additional graphs below illustrate the beneficial effect of adding post-acute capacity. Medical outliers are significantly improved and surgical operations increase (Figures 16 and 17).

Fig 16: Improved medical bed usage following additional post-acute capacity



Reduction in medical outliers (line 5) when post-acute capacity is added

Fig 17: improved elective activity after adding post-acute capacity



Increase in surgical procedures when post-acute capacity is added (line 1 is base run; line 2 is with additional acute capacity; line 3 is with additional post-acute capacity)

As a learning environment, the model shows how to balance capacity across the whole system, and demonstrates that the real cause of a problem may be significantly removed from where it surfaces. So, for example, if elective waiting lists are high as a result of a large number of delayed discharges, different strategies can be compared to resolve this. It transpires that increasing hospital capacity will tend to produce a short-term gain as more patients are treated, followed very quickly by development of a bottleneck in the discharge process which blocks further admissions. By contrast, the addition of post-acute care, particularly domiciliary care, results in not only fewer delays, but also perhaps a greater reduction in waiting lists.

Without a model, it is very difficult to describe or envisage how changes in one part of the system will affect the whole. Indeed, it is hard to imagine how such a conversation could take place between managers responsible for different types of service, without some kind of process model to facilitate the discussion. Where the model has been presented back to a group of strategic managers drawn from across a whole health economy, it has provided a qualitatively different type of discussion. Participants become much more sympathetic towards understanding a systemic problem from the point of view of colleagues from other agencies and settings. They also accept much more readily the model's projections of the unintended consequences that might arise from individual, single-agency initiatives.

Conclusions

Given that whole systems are complicated, and behave in unpredictable ways, managers need new tools to equip them to undertake joint planning. No single method can do everything, but dynamic modelling provides a new way of understanding how whole systems operate. It provides a safe environment within which representatives from across agencies and functions can make explicit their own assumptions, understand the impact that their initiatives might have on other parts of the system, and develop ways of collaborating to achieve maximum benefit for service users.

Some of the benefits are shown in Figure 18 and 19.

Fig 18: General Findings from System Dynamics

Diagnosis	• Problem is often at some distance from where symptoms appear • Best action may be in organisation A, but results benefit organisation B • Aim for maximum BALANCED flow (don't just optimise one bit)
Complexity	• This is caused by interaction of flows, capacity and time delays • Rules set up to deal with expected situation may EXACERBATE problem if people have no freedom to take appropriate action
Capacity	• Capacity is rarely the (only) answer to bottlenecks: try changing the process (eg split the flow, change the timings of the flow) • Ill-judged use of additional capacity can make matters WORSE • Plan required capacity on forecast demand (not reaction to problem)
Variation	• This causes bottlenecks or under-usage of resource (often alternately) • One way to deal with peaks in demand is to have a way to "buffer" • Another solution is to divert away from the main flow
Performance	• Measure performance of Whole System – beware perverse incentives • Important to measure what is happening at the LEVERAGE points (the most sensitive parts of the process, where change has most effect)

QUANTIFY likely results and beware UNINTENDED CONSEQUENCES

Fig 19: What the Template Demonstrates

Diagnosis	• Best solutions for delayed discharges are ***preventive services and post acute care*** (ie the front and back of the process) • This also improves elective throughput and wait times • Adding surgical capacity may actually ***decrease*** elective activity
Complexity	• The whole system is very sensitive when A&E is over-stretched: o Medical outliers significantly reduce elective activity o Early discharges may increase readmissions/reduce throughput
Capacity	• Capacity rapidly fills up and often only 'buys time' • Adding 80 acute beds and 140 post acute places resulted in little improvement as the gearing was insufficient
Variation	• Good results are obtained by diverting a small percent of 'slow route' • Need the ability to buffer demand at A&E (eg outliers, early discharge) • Intermediate care can buffer for post acute if throughput is maintained
Performance	**Measure** • Early discharges, outliers/day • Readmissions/week • Average LOS for each flow • Utilisation • Elective wait list/times • Number of operations • Delayed transfers • Reimbursement total

The results with delayed discharge have been sufficiently compelling to attract at least one significant adopter – a Strategic Health Authority (SHA). There are 28 SHAs in England and they play a key role in managing performance of the primary care and acute sectors within their region. They can also play a part in developing centres of excellence in various strategic, management or operational skills. If other SHAs were to follow suit, the dissemination of SD within the health service would be greatly accelerated.

References

Ann van Ackere, Towards a macro model of Health Service Waiting Lists (System Dynamics Review, 1999, Vol. 15, No 3).

Dangerfield, B. and Roberts, C. 1999, Optimisation as a statistical estimation tool: an example in estimating the AIDS treatment free incubation period distribution (System Dynamics Review, Vol. 15, No. 3).

Dangerfield, B., Fang, Y and Roberts, C. (2001), Model based scenarios for the epidemiology of HIV/AIDS: the consequences of highly active antiretroviral therapy. (System Dynamics Review, Vol. 17, No. 2).

Lane, D. C., Monefeldt, C. and Rosenhead J. V. 2000. Looking in the Wrong Place for Healthcare Improvements: A system dynamics study of an accident and emergency department (Journal of the Operational Research Society 51(5): 518-531).

Roysten G., Dost A., Townsend J. and Turner H. Using System Dynamics to help develop and implement policies and programmes in Health Care in England (System Dynamics Review, 1999, Vol. 15, No 3).

Vennix, J. (1996) Group Model Building: Facilitating Team Learning Using System Dynamics, Chichester, England: John Wiley and Sons.

Wolstenholme E. F., Monk, D., Smith G. and McKelvie D. Using System Dynamics to Influence and Interpret Health and Social Care Policy in the UK. Paper to be presented to the 2004 International System Dynamics Conference

Wolstenholme, E.F. A Case Study in Community Care using Systems Thinking (Journal of the Operational Research Society, Vol. 44, No. 9, September 1993, pp 925-934).

Wolstenholme, E.F. A Management Flight Simulator for Community Care (Enhancing Decision Making in the NHS, Ed. S. Cropper, 1996, Open University Press, Milton Keynes).

Wolstenholme, E.F. A Patient Flow Perspective of UK Health Services (System Dynamics Review, 1999. Vol. 15, no. 3, 253-273).

Wolstenholme, E.F. Patient Flow, Waiting and Managerial Learning - a systems thinking mapping approach (Working paper, Cognitus 2002).

Wolstenholme, E. (1990) System Enquiry-A System Dynamics Approach. Chichester, England: John Wiley and Sons.

Glossary of Terms

Accident and Emergency – A&E

General Practitioner – GP

Health Authority – HA

Information Technology – IT

Local Government Association – LGA

Mental Health - MH

National Health Service of the UK – NHS

National Institute for Mental Health in England - NIMHE

Primary Care Trust – PCT

Social Services Department – SSD

Strategic Health Authority – SHA

System Dynamics – SD

For the benefit of non-UK readers, unless otherwise indicated the word "national" means "England-wide". The word "government" refers to the UK-government, which controls health policy for England only. A variety of devolved arrangements applies in Northern Ireland, Scotland and Wales.