Simulation analysis of the consequences of shifting the balance of health care: a system dynamics approach

Kathryn Taylor, Brian Dangerfield¹, Julian Le Grand

Social Policy Department, London School of Economics and Political Science, UK

¹Centre for OR & Applied Statistics, University of Salford, UK

Kathryn Taylor PhD, Doctoral Student, Department of Social Policy, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK, Julian Le Grand PhD, Richard Titmuss Professor of Social Policy, Brian Dangerfield PhD, Professor of Systems Modelling, Centre for OR & Applied Statistics, University of Salford, Salford, M5 4WT, UK

Correspondence to: KT via email (k.taylor-alumni@lse.ac.uk)
Abstract

Objectives: The shift in the balance of health care, bringing services ‘closer to home’, is a well established trend. This study sought to provide insight into the consequences of this trend, in particular the stimulation of demand, by exploring the underlying feedback structure.

Methods: We constructed a simulation model using the system dynamics method, which is specifically designed for the analysis of feedback structure. The model was calibrated to two cases of the shift in cardiac catheterisation services in the UK. Data sources included archival data, observations and interviews with senior health care professionals. Key model outputs were the basic trends displayed by waiting lists and average waiting times, cumulative patient referrals, cumulative patient activity, and cumulative overall costs.

Results: Demand was stimulated in both cases via several different mechanisms. We revealed the roles for clinical guidelines and capacity changes, the typical responses to imbalances between supply and demand. Our analysis also demonstrated the potential benefits of changing the goals that drive activity by seeking a waiting list goal rather than a waiting time goal.

Conclusions: Appreciating the wider consequences of shifting the balance of care is essential if services are to be improved overall. The underlying feedback mechanisms of both intended and unintended effects need to be understood. Using a systematic approach is essential if the editor specifically wants to avoid the reference to joined-up thinking then we feel that...
“systemic” is the next best term to use] approach, more effective policies may be designed through coordinated programs rather than isolated initiatives, which may only have a limited impact.
Introduction

The balance in health care in many systems is shifting towards the primary sector and services delivered ‘closer to home’. These changes include the development of outreach clinics, near patient testing, GPs undertaking minor surgery, telemedicine and day surgery. This trend has been motivated by the broad desire to improve the provision of services by alleviating pressure on hospital resources, by expanding capacity and thus reducing waiting times, and by providing more accessible care. However, the efforts to improve services overall may be undermined by the unintended consequences of service shifts, particularly if improvements in access stimulate demand. This general phenomenon has been frequently reported, but only anecdotal evidence is available with respect to service shifts. Demand increases are particularly problematic in the UK health care system, the National Health Service (NHS), where there are persistent concerns about excessive waiting lists and waiting times, and rising costs. Additional complexity arises as views about the value of stimulated demand vary.

Underlying the stimulation of demand are feedback structures, or closed causal relationships, since access influences demand and vice versa. The consequences of feedback structures are not always obvious. They may be delayed, counterintuitive, and cross organisational boundaries. Therefore, ‘joined-up’ or ‘whole systems’ thinking is required to understand them. However, existing analyses of the consequences of service shifts and the stimulation of demand in general have been limited to isolated parts of the system, based upon economic appraisals and surveys. Furthermore, these analyses have led to calls to control demand with stricter clinical guidelines or to meet demand with capacity increases. However, these interventions may have only a limited impact due to the complexity of the underlying feedback mechanisms.
The system dynamics simulation approach\(^1\) is specifically designed for the analysis of feedback structures. System dynamics contrasts with more familiar simulation modelling (e.g.\(^2\)) in several ways. In particular, the aims of system dynamics are more qualitative; that is to understand and improve patterns of behaviour (the basic trends), rather than produce point predictions. In this paper, we describe a system dynamics modelling study\(^3\) of a tertiary to secondary shift in the NHS: the shift in cardiac catheterisation (CC) services for low risk investigations.\(^4\) The majority of elective cases are defined as low risk. We focus on the development of CC services at two English district general hospitals. For reasons of confidentiality, they are referred to as ‘Ribsley Hospital’ and ‘Veinbridge Hospital’.

Prior to the introduction of district services, the elective demand for Veinbridge’s patients was met. However, there was a persistent problem of under-capacity for Ribsley’s patients. Consequently, temporary capacity increases had been provided at the tertiary level.

At Veinbridge Hospital, the district CC service formed part of a long-term strategy to expand capacity and to develop cardiac services. Consequently, the district CC service was introduced and developed into a permanent service. Meanwhile, at Ribsley Hospital, due to resistance by purchasers, only temporary district CC services were provided.

**Methods**

Data collection included accessing hospital records, observational work, informal discussions and formal interviews with consultant cardiologists and hospital managers at the district hospitals and their main referral centre, with commissioners of health services and with managers of the company that provided a mobile catheterisation laboratory to the district hospitals. Information was elicited regarding the processes of care and the various factors that influenced decision making, which together formed the mechanisms underlying the
consequences of the shift in CC services. In considering ‘access’, we addressed both factors of ‘having access’ and ‘gaining access’.

System dynamics simulation models, which have a mathematical underpinning based upon differential equations, are most conveniently constructed using a purpose-built software package. Our simulation model was constructed using STELLA/ithink software. The model generates the supply and demand for the outpatient clinic (the main source of referral for an elective CC investigation) and for an elective CC investigation (Figure 1).

In calibrating the model to our two case studies we produced two sets of parameters to reflect their different circumstances. Model calibration procedures involved the use of hospital data, expert estimates, simple calculations and preliminary simulations. Several parameters were calculated from the actual activity rates and assuming equilibrium with the desired waiting time goals maintained. [Note to authors: please rephrase for greater clarity. We propose the changes shown]

For the Veinbridge case, we used the information that waiting times for CC were under control to calculate several parameters for which data was not available. Preliminary simulations derived several starting values and were used as a vehicle for standard techniques including sensitivity analysis. Confidence was gained in the model via established methods. One key test confirmed the model’s ability to replicate the problematic behaviour. This involved superimposing and comparing plots of simulated and actual behaviour.

We provided retrospective analyses by generating insight into how the shifts in services helped and hindered the provision of NHS cardiac services over time, and into how NHS providers and health service commissioners could have more effectively intervened to improve the provision of services. The experiments were conducted by making parameter and structural changes to the model. The interventions investigated included the use of stricter
clinical guidelines, various approaches to capacity changes, and driving activity by seeking waiting list goals rather than waiting time goals.

We evaluated and interpreted the experiments broadly in order to illustrate the trade-offs involved, and encapsulate the concerns of both NHS providers and health service commissioners. We considered several different aims. First, the need to improve health; i.e. identify patients in need of treatment (some patients with severe heart disease only display minor symptoms and these patients may thus fail to be identified by other diagnostic methods), provide health care promptly and appropriately, direct resources towards the most urgent cases, and increase activity and thus meet higher activity targets. Second, the need to control the overall costs incurred; i.e. ensuring that services are used appropriately. Third, the need to improve efficiency; i.e. deliver care at the lowest cost per case. Both supply and demand variables were considered across the cardiac referral chain. The key model outputs were the basic trends displayed by the waiting lists and average waiting times, cumulative patient referrals, cumulative patient activity, and cumulative overall costs.

Whilst maintaining waiting time goals is a key policy aim, it is also important to control the waiting list length. Meeting the waiting time goal does not necessarily mean that the waiting list goal is met. The waiting list could be rising and this would imply that the waiting time goal is only being maintained by increases in patient activity. Therefore, controlling the waiting list is of importance to those involved in funding services and in dealing with the challenges of distributing limited resources.

Results

Changes in access

At Ribsley, the district service prolonged the provision of capacity increases and thus produced dramatic improvements in access to elective services (Figure 2a). However, this
trend was reversed when the district service was withdrawn. A further temporary service was offered at a later date. This achieved further temporary improvements in access.

Demand was stimulated in both cases and for both outpatient and CC services. The stimulation of demand was described during the interviews and informal discussions, and it was also reflected in the hospital data. There was sufficient supply to meet the CC investigation waiting time goals whilst the district CC services were in place. However, in the Veinbridge case, the increase in demand for an outpatient appointment (Figure 2b) produced severe detrimental effects on access to this service. This loss in access had not been anticipated.

Basic feedback mechanisms

Feedback mechanisms may be balancing, representing efforts to control and seek particular goals, or reinforcing where self-sustaining growth or decay occurs. The collaborators provided rich descriptions of several feedback mechanisms by which changes in access influenced demand. These mechanisms included those involving the average waiting time, the knowledge of patients and GPs of CC and the new CC service, and the skills and confidence of the doctors who referred patients for CC (Table). No other changes occurred during the 40-month observation period that could have accounted for the changes in demand.

The basic feedback mechanisms may be summarised (further details are provided elsewhere\textsuperscript{19}) by a ‘fix that fails’\textsuperscript{24} (Figure 3) which is a simple structure that explains how problems increase over time. The increases in patient activity, facilitated by the use of district services, may be considered as efforts to ‘fix’ the ‘problem’ of poor or unsatisfactory access to cardiac services (a balancing process) with the stimulation of demand (a reinforcing
process) representing the side effects that limit the effectiveness of this fix. The model-based policy analyses thus focused on strengthening the fix and/or controlling the side effects.

**Clarifying the roles for capacity increases and stricter clinical guidelines**

The Veinbridge case illustrated that capacity increases are not necessarily the most effective way of improving access. For even if it had been possible to increase outpatient capacity, this would only have eliminated excessive waiting time, not the excessive waiting list. This is because efforts would have been made to maintain the desired waiting time but without controlling the waiting list length. In general, seeking a desired waiting time will only control the waiting time if there are sufficient resources available to respond to increases in the waiting list by removing patients commensurately. **[Question to authors: is this OK? No. The text, following the editorial changes, refers to controlling the waiting list length. We specifically wish to refer to efforts to control the waiting time. We then refer to the desire to control the waiting list length in the next sentence. We suggest the change shown above]** The desired waiting list length will not necessarily be met at the same time as the desired waiting time. With sufficient slack in the system (by weakening the reinforcing loop in Figure 3), both goals could be met simultaneously by changing the forces that drive activity rates; i.e. seeking a desired waiting list length rather than a desired waiting time (strengthening the balancing loop).

There was spare capacity for elective CC services and introducing controls on demand would have produced some slack for outpatient services (Figure 4). In fact, the combined policy would have overcompensated by producing an average waiting time that was lower than required. Of course, slack for outpatient services could also have arisen from removing the capacity constraint. However, this would not have been feasible as the purchasers were unwilling to further fund significant increases in outpatient activity.
The combined policy would have led to improvements in access and reductions in costs. There would have been trade-offs as the reductions in referrals and activity would have been in conflict with the desire to meet higher activity targets and identify more high risk patients. The mobile-based district CC service was as efficient as the tertiary-based CC service. When an integrated laboratory opened at Veinbridge, replacing the mobile facility, it provided the opportunity to improve the efficiency of the district service by increasing volume and thus reducing the cost per case.

The extent of the inherent imbalance between supply and demand at Ribsley was such that frequent capacity increases would have been necessary to provide permanent improvements in access (strengthening the balancing loop in Figure 3). Coordinating this policy with efforts to manage demand (weakening the reinforcing loop) would have ensured that the benefits of increasing supply were not cancelled out by stimulated demand. Whilst this would be obvious to NHS managers and clinicians, the usefulness of the analysis lay in the ability to explore different approaches to capacity increases and gain insight into the trade-offs involved between short-term and long-term effects, and between localised and broader consequences (Figure 5). On balance, it seemed that further temporary district services would have been the most practical approach to capacity increases, compared to a permanent district service and an expanded tertiary service.

**Discussion**

The feedback mechanisms, associated with service shifts have been neglected in spite of the increasing emphasis in health care on the need for ‘joined-up’ or ‘whole systems’ thinking. This paper presents empirical evidence of shifts in services stimulating demand derived from participants’ reports and graphical evidence derived from hospital activity data. The participants’ reports were tested and supported by our model. This paper also illustrates the
insights that system dynamics can offer and thus supports previous claims made regarding its usefulness in health care\textsuperscript{18} and of operational research methods in general\textsuperscript{25-27}. As with all models, this model is based upon certain simplifying assumptions. For example, in reality, there would be additional feedback mechanisms. Nevertheless, the model still successfully replicated the problematic behaviour for both cases. Furthermore, as the interactions between supply and demand are widely applicable, we believe that our findings may be generalised to other locations, times and service shifts, especially to shifts across the secondary/primary interface, which is the main current focus of this trend. Therefore, a number of broad policy lessons can be derived.

The shift in the balance of care is a continuing trend, which is currently being promoted by government policies in the UK such as the new GP contract\textsuperscript{28} [Note to authors: reference needed – The reference has been inserted] where GPs are being encouraged to provide services that are traditionally carried out in hospitals. Our study provides insight into both the desirable and less desirable consequences of such service shifts. Whilst analysts have some understanding of the consequences of service shifts, they are still not universally appreciated by NHS managers and clinicians.

Our study also demonstrates the limitations of the use of stricter clinical guidelines and capacity increases in improving access. This echoes and develops previous research\textsuperscript{29-28}. The Veinbridge case showed that individually these interventions would have provided similar leverage in improving behaviour (in terms of reducing the waiting list and average waiting time), but significantly better leverage could have been obtained by combining the use of stricter guidelines with changes to the goals that drive activity. This result arose from the existence of spare capacity and was not specific to the case of the shift in CC services. It was a general consequence of the interplay between supply and demand that determines the waiting list length and average waiting time. Although discussions of spare capacity in the
UK may seem incongruous with customary reports of long NHS waiting lists and waiting times, [Note to authors: However, waiting times are falling rapidly at the moment, so our view of the NHS should change too! – Please see changes] some spare capacity is often released by service shifts, through their ability to provide additional capacity and/or by prioritising patients. By contrast, the Ribsley case illustrated the superior leverage of capacity increases in cases of extreme imbalance between supply and demand.

We challenge the persistent tendency in health care towards a narrow focus on isolated events, short-term results and single performance measures, as well as the current emphasis in the NHS on waiting time rather than the length of the list. For example, the Veinbridge analysis showed how maintaining a waiting time goal did not necessarily mean that the system was not under pressure; the rise in the waiting list suggested that it was and the waiting time goal was only maintained because more money was poured in to raise activity levels. Another example arose from the Ribsley analysis. We showed that a permanent district service at Ribsley would have only produced similar short-term qualitative improvements in access to CC services, compared to an expanded tertiary-based service. In addition, the former would have produced later problems of access to outpatient services.

In the Veinbridge case, we illustrated how access problems can arise from the inability to cope with patient pressures. This related to the effect of changes in knowledge about CC on demand for services (see Table). It has been argued that clinical decisions should be driven by the preferences of patients. Questions thus arise about how the shift in the balance of care can continue whilst providing high quality care to patients whose expectations are traditionally high. [Note to author: I have removed this paragraph because it did not seem to be integrated into the rest of the discussion. OK? – No. We are referring to the effect of changes in knowledge about CC on demand for services (please see key factor]
“Knowledge” in Table) which was a key factor underlying the outpatient access problems that arose in the Veinbridge case. We propose reinstating the text with some changes as shown above. Please note that (a) this involves removing one reference to allow for the new GP contract reference and (b) by reinstating the text we still remain within the word limit.

Whilst the NHS (a publicly funded, cash-limited system) formed the context to our case studies, our research findings may apply to shifts in health services in other countries in spite of the differences between their health systems. The stimulation of demand in response to improved access is a common response in the NHS since services are free at the point of delivery. However, services in most other health systems, whether private or insurance-based, are largely free at the point of use. The individual feedback mechanisms that we have discussed can be generalised to other health systems. Even the impact of waiting times on referrals can be generalised in spite of the fact that waiting lists are a typical characteristic of the NHS; in other systems, the waiting time may be translated into the price customers pay for services.

In seeking improvements to the provision of services, we did not focus on changes in patient outcomes. The procedural complication rates did not change, so safety was not compromised in shifting services away from the tertiary centre, which offered surgical backup. In the Ribsley case, the waiting time for catheterisation dropped and this contributed to shorter delays for those requiring invasive treatment (coronary angioplasty or coronary bypass surgery). This could offer the potential for health benefits as treatment delays can lead to some patients deteriorating. The fraction referred on for a CC investigation increased in both cases. As the case mix of the patient populations did not change, this meant that the threshold for a CC investigation had changed to include less severe cases (defined without the benefit of a CC investigation). Catheterising more of these patients could lead to
improvements in health if it identified further patients in need of invasive treatment. The fraction of patients referred on for invasive treatment remained constant so associated with the increased catheterisation rate was an increased invasive treatment referral rate. This would suggest that, from the less severe cases, further patients in need of such treatment had, indeed, been identified. However, for these cases, it might have been better to delay bypass surgery until the disease was more advanced, since repeat bypass surgery, which can occur given the progressive nature of heart disease, carries higher risks. Therefore, delaying catheterisation (and therefore delaying bypass surgery) could, paradoxically, be beneficial in the long-term management of the disease.

Applying our methods to explore shifts in other services would serve to test the generalisability of our findings and provide further insight into this health care trend. Further work could also consider the possible shifts in demand between different hospitals, and associated loss in income, as a result of service shifts.

Acknowledgements

The collaborators’ invaluable input is gratefully acknowledged. This work formed part of a doctoral research project that was sponsored by The Wellcome Trust (Reference Number 041243). Our thanks are also due to the two anonymous referees for their helpful comments.
References


20. A working group of the British Cardiac Society with the Royal College of Physicians of London and Edinburgh and Royal College of Physicians and Surgeons of Glasgow. The changing interface between district hospital cardiology and the major cardiac centres. Heart 1997; 78: 519-523


25. Buhaug H. Long waiting lists in hospitals: operational research needs to be used more often and may provide answers. BMJ 2002; 324: 252-253

26. Smith R. What doctors and managers can learn from each other: a lot. BMJ 2003; 326: 610-611


<table>
<thead>
<tr>
<th>Key factor</th>
<th>Feedback mechanism</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity drivers</td>
<td>Adjustments in activity to meet desired waiting time. <em>Balancing feedback loops.</em></td>
<td>Policy analysis involved changing goal to desired waiting list length.</td>
</tr>
<tr>
<td>CC waiting times</td>
<td>Demand suppressed by high waiting times and stimulated by low waiting times. <em>Balancing feedback loop.</em></td>
<td>Loop did not apply in the Veinbridge case because waiting times were not considered a factor in making decisions about referrals. In the Ribsley case, the suppression of demand was insufficient to control access.</td>
</tr>
<tr>
<td>Knowledge</td>
<td>By developing a local CC service, GPs and patients become more knowledgeable about the benefits of CC and overcome anxiety about the risks, and thus more demanding for this service. The extent of this effect on demand increases as the district service grows generating more publicity and, through ‘word of mouth’, more reports of patients who have benefited. <em>Reinforcing feedback loops.</em></td>
<td>Affected CC and outpatient services. The increase in pressure produced outpatient capacity shortages at Veinbridge.</td>
</tr>
<tr>
<td>Skills</td>
<td>As junior staff gain skills and confidence as CC operators, they refer more. Accelerated learning by capacity increases and periodic effects occur due to the existence of training schemes and the staff rotation between different hospitals (e.g. periodic drops in referrals with the replacement of experienced juniors with novices). <em>Reinforcing feedback loops.</em></td>
<td>Expands upon a previous description by considering effect of confidence on referrals and variation of effects. Only refers to junior CC operators. Loop did not apply in the Veinbridge case because the consultant cardiologist (an expert CC operator) made all final decisions about referrals.</td>
</tr>
<tr>
<td>Other waiting list removals</td>
<td>Removals following review of the waiting list e.g. deaths, patient moved to another area, patient’s condition improved. Modelled as a constant fractional rate of the waiting list length as a simplifying assumption. <em>Balancing feedback loops.</em></td>
<td>Only significant in the Veinbridge case for outpatient services.</td>
</tr>
</tbody>
</table>

Several other feedback mechanisms were modelled exogenously as simplifying assumptions.
Figure 1 Model overview
Figure 2 Consequences of district services: Improving access and stimulating demand. Black blocks indicate the use of district services. In (b), note that the stimulation of demand lags behind the introduction of district services due to the perception delays of GPs and patients.
A fix, effective in the short-term, has side effects which may require even more use of the same fix.

Figure 3 Fixes that fail
Figure 4 Meeting the outpatient waiting list and waiting time targets for the Veinbridge case. District services were available from month 13 onwards. The outpatient waiting list goal was 338 patients and the waiting time goal was 3.5 months.
Figure 5 Increasing elective CC capacity for the Ribsley case. The outpatient waiting list graphs for runs 0 and 3 are the same. District services were available from months 14 to 23 and months 34 to 38. The outpatient waiting list goal was 424 patients. The CC investigation waiting list goal, 60 patients, was halved during the periods when the district service was present, reflecting how expectations rose and efforts were made to ‘squeeze’ more out of the system.