A System Dynamics Approach to Analysing NHS Waiting Times

Ann van Ackere* and Peter Smith**
London Business School, ** University of York

The United Kingdom National Health Service (NHS) delivers 95% of the country's health care. It is a public sector organization which delivers health care free of charge at the point of access. With the exception of emergency services, patients cannot refer themselves directly to NHS hospitals. Instead they must be referred by a family practitioner, who acts as the "gatekeeper" to hospital health care. If referrals to hospitals give rise to excess demand, rationing takes place in the form of waiting lists for non-emergency procedures. There is a small private sector which is used by some patients who have the means to bypass the NHS queues.

The issue of NHS waiting lists and the associated problem of waiting times for elective surgery has been a source of acute public and political concern since the inception of the NHS in 1948. Although most patients were treated reasonably quickly (about 2/3 within 3 months) there were always a small minority who had to wait very long times for surgery. In response to this, the UK Government set up in 1991 a "Patient's Charter" which, amongst other things, placed a duty on local hospitals to ensure that all patients received treatment within two years of being put on a waiting list. This has had a dramatic effect in eliminating the very long waits experienced by some patients, and the Patient's Charter has now been amended to reduce waiting times for certain conditions still further.

Thus waiting times play a key role in the functioning of the NHS. On the demand side, a long expected waiting time may persuade a patient to forego treatment or to seek private health care. On the supply side, long waiting times reflect poorly on local management. In response to long waiting times, management might therefore either devote more resources to inpatient surgery, or seek to use those resources more efficiently.

Previous analyses of this problem mostly consider a static framework. Martin and Smith [1] model waiting times for elective surgery, using comprehensive data for 1991-92. They assume an equilibrium situation, and estimate a demand and a supply equation. The resulting coefficients are estimates of the elasticity of the dependent variables (demand and waiting time) with respect to the explanatory variables. This paper builds on this work by adding the dynamic dimension, using a system dynamics approach. The estimates provided by Martin and Smith enable meaningful parameterization of the model.

Figure 1 shows a simple causal loop diagram illustrating the key feedback structure, which consists of two feedback loops: longer waiting times lead to pressure for more resources (and more effective use of available resources), and lower demand, both of which yield shorter waiting times. Figure 2 shows the stock and flow diagram, as well as the equations. The waiting list increases by referrals, a reflection of demand. The list is depleted as patients are treated. The treatment rate depends on the number of beds (a measure of capacity) and efficiency (inefficiency is measured as the average length of stay). Changes in demand, beds and inefficiency result from changes in the perceived waiting time, and their elasticity with respect to waiting time. The elasticities of beds and inefficiency are assumed constant, while the elasticity of demand depends on perceived waiting time in a highly non-linear way (see figure 3). The "external change in beds" is used to subject the model to a step change in resources (beds).

Figure 4 shows the results of 4 simulation runs. Runs 1 and 2 consider a 10% increase in resources, staring from an average waiting time of 3 and 4.5 months respectively. Runs 3 and 4 consider a 10% decrease in resources for the same cases. Note that combining cases 2 and 3 can be interpreted as transferring resources from a 'short wait' region to a 'long wait' region.

Figure 4A shows that it is comparatively easy to achieve significant reductions in waiting time when the initial waiting list is shorter: a reduction of 0.8 months (just over 25%) in run 1, compared to a reduction of 0.5 months (barely 12%) in run 2. The deterioration resulting from a reduction in resources is of the same order of magnitude for both cases: respectively 28% and 25% in runs 3 and 4.

Figure 4B shows the impact on demand: changes in resources are partially off-set by changes in demand when the average waiting time lies in the "sensitive" region (with our parameterization, between 3 and 5.5 months).

Figure 4C shows the internal pressures on the reallocation of resources when average waiting times change due to a sudden change in resources. Only in run 4 (long initial waiting time and resource increase) does the externally induced change resist fairly well to internal pressures.

The present model needs further refining. For instance, the time required for changes in waiting time to affect demand, number of beds and efficiency is assumed to be the same, un unrealistic assumption. Further work will consider a more detailed model of the demand side, including referral patterns, and a more realistic representation of changes in resources.

Reference

[1] Martin and Smith, Modelling Waiting Times for Elective Surgery, 1995

Acknowledgment

We are grateful to Stephen Martin for providing us with further estimates of the various elasticities, not included in [1]

Figure 1

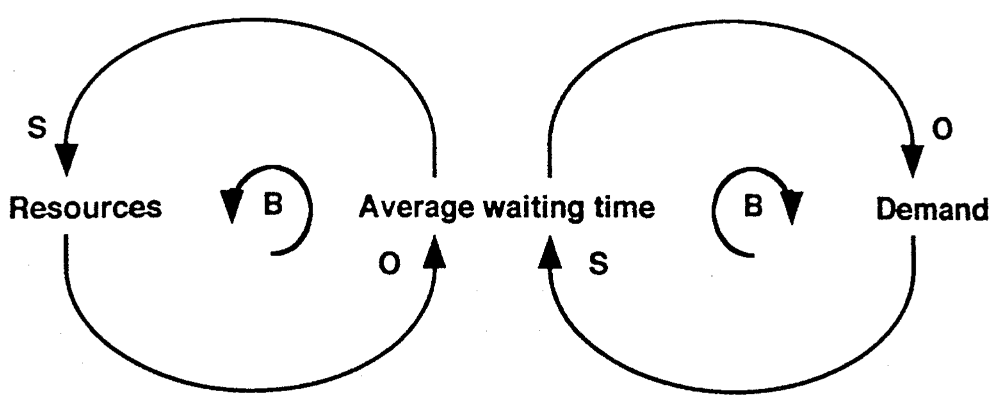
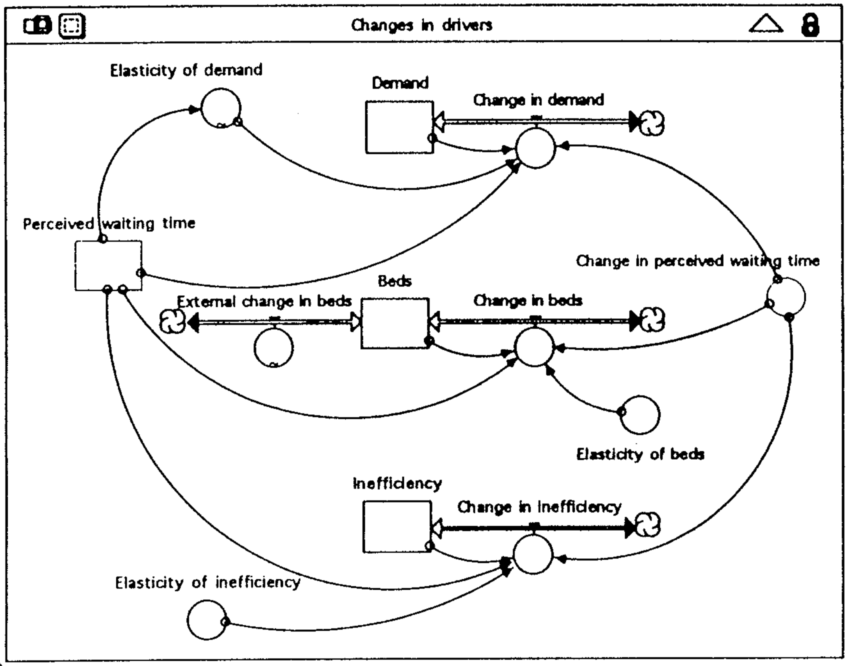


Figure 2



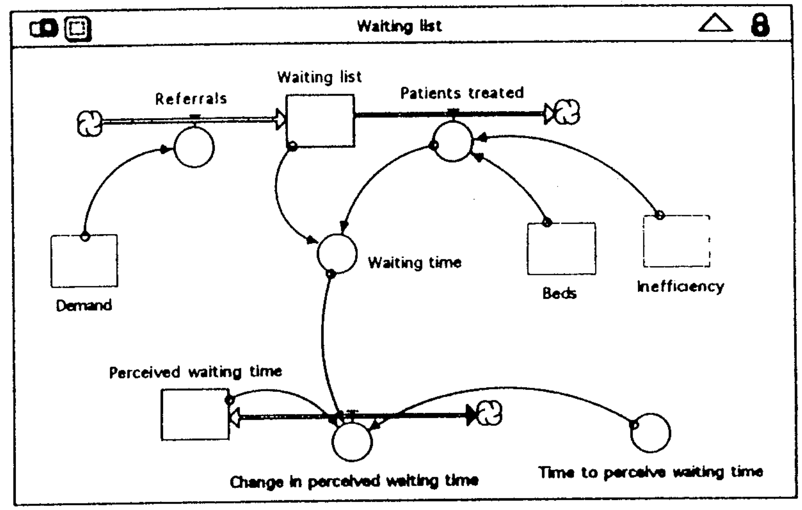
Changes in drivers

☐ Beds(t) = Beds(t - dt) + (Change_in_beds + External_change_in_beds) * dt
INIT Beds = 10 {beds}
INFLOWS
Change_in_beds = Elasticity_of_beds*Beds*Change_in_perceived_waiting_time/Perceived_waiting_time {beds per month}
External_change_in_beds = GRAPH(Time)
(9.00, 0.00), (10.0, 1.00), (11.0, 0.00), (12.0, 0.00)

☐ Demand(t) = Demand(t - dt) + (Change_in_demand) * dt
INIT Demand = 100 {patients per month}
INFLOWS:
Change_in_demand = Elasticity_of_demand*Demand*Change_in_perceived_waiting_time/Perceived_waiting_time {people per month}

☐ Inefficiency(t) = Inefficiency(t - dt) + (Change_in_inefficieny) * dt
INIT Inefficiency = 0.1 {months, inefficiency is approximated by the average length of stay}
INFLOWS:
Change_in_inefficieny = Elasticity_of_inefficiency*Inefficiency*Change_in_perceived_waiting_time/Perceived_waiting_time {months per month}

○ Elasticity_of_beds = .29 {constant}
○ Elasticity_of_inefficiency = .03{constant}
⊘ Elasticity_of_demand = GRAPH(Perceived_waiting_time)
(0.00, 0.00), (0.5, 0.00), (1.00, 0.00), (1.50, 0.00), (2.00, 0.00), (2.50, 0.00), (3.00, 0.00), (3.50, -0.09), (4.00, -0.4), (4.50, -0.4), (5.00, -0.113), (5.50, 0.00), (6.00, 0.00)



Waiting list

☐ Perceived_waiting_time(t) = Perceived_waiting_time(t - dt) + (Change_in_perceived_waiting_time) dt
INIT Perceived_waiting_time = Waiting_list/Patients_treated {months}
INFLOWS:
Change_in_perceived_waiting_time = (Waiting_time-Perceived_waiting_time)/Time_to_perceive_waiting_time {months per month}

☐ Waiting_list(t) = Waiting_list(t - dt) + (Referrals - Patients_treated) * dt
INIT Waiting_list = 300 {patients}
INFLOWS:
Referrals = Demand {patients per month}
OUTFLOWS:
Patients_treated = Beds/Inefficiency {patients per month; see document for comment on units}

○ Time_to_perceive_waiting_time = 3 {months}
○ Waiting_time = Waiting_list/Patients_treated {months}

Figure 3

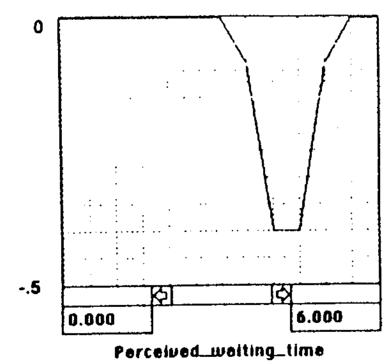


Figure 4

A. Waiting time

B. Demand

C. Beds