Agent based modelling and policy design: the case of elective surgery in the UK National Health System

Donald Franklin, Senior Economic Advisor, Department of Health
Martin Chalkley, Department of Health and Department of Economics, University of Dundee
James Raftery, Director, Health Economics Unit, University of Southampton
Ellie Cooper, Volterra Consulting
Paul Ormerod, Volterra Consulting
Matt Salisbury, Volterra Consulting

September 2005

corresponding author: Paul Ormerod e-mail: pormerod@volterra.co.uk

Summary

Elective surgery refers to procedures such as hip replacements which are not urgent. There are private sector providers of health care in the UK, but most is funded and provided by the public sector National Health Service (NHS).

At present, individuals using the NHS for elective surgery are essentially obliged to use their local hospital. It is proposed to introduce instead a policy regime under which individuals can choose their hospital, and in which the NHS will fund treatment at any hospital, private or public, offering care at a fixed tariff which meets minimum quality standards.

This agent-based model was developed for the economists in the Department of Health to inform on policy advice as to how this market is likely to work. The model contains decision rules for both consumers and providers. These agents are heterogeneous, operate with imperfect information and do not necessarily maximise.

The model has been calibrated to actual data on 13 hospitals currently providing primary hip replacement surgery in the Birmingham and Black Country Strategic Health Authority. The results have been presented to professionals in the NHS both inside and outside this area.

In this market, the price available to producers is fixed. Consumers have imperfect information and do not necessarily take their optimal decision even with respect to the limited degree of information which they have. Producers adopt relatively myopic decisions rules. Nevertheless, in general the introduction of consumer choice results in an increase in the quality of the product, a fall in waiting times and a rise in consumer welfare.

Key words: agent-based; heterogeneous agents; health; imperfect information

1. Introduction

The UK National Health Service (NHS) was founded in 1948, and its organisational principles have in general remained closer to those of a centrally planned system than a market-oriented one. However, an important change is proposed to take effect in the near future. Elective surgery refers to a class of operations which are not urgent and which can be scheduled flexibly. Well known examples include hip replacements and cataract operations. To date, consumers within the NHS have essentially been assigned to the waiting list of their local hospital.

Under the proposed new system, consumers will in principle be able to choose the hospital where their operation will be carried out. A tariff (price) for each particular type of operation will be fixed centrally, and the hospital which actually performs the operation will receive this tariff.

The proposal has attracted widespread criticism. Examples include Appleby et.al. (2003), who argue that increased patient choice will have negative effects, mainly to do with equity of use, and Webb (2005). However, these lack any formal model of how choice might operate.

The present model was designed to assist the economists in the Department of Health formulate their policy advice on how the new system might operate. The model is a general one, which can be calibrated to a specific example, both in terms of the initial conditions and the behavioural rules themselves. Section 2 describes the operation of the model and the behavioural rules. Section 3 discusses the properties of the model, and in section 4 we provide examples of the sensitivity of the properties to different assumptions. In section 5, we give an example of a calibration of the model to actual data, using information on hip operations in hospitals in the Birmingham and Black Country district in the English Midlands.

2. The operation of the model and its behavioural rules

There are M suppliers initially. These are placed on a circle, and the distance between each one is d. The maximum distance between any pair is scaled to be equal to 1, so d $\in$ [0,1]. There are N consumers, where N >> M. These are geographically based, and are initially obliged to use the nearest supplier. (For simplicity, we assume that the distance between each consumer and his/her nearest supplier is 0).

We allow the model to move forward in time on a period by period (week by week) basis. Consumers are allowed to choose the hospital where their operation will be carried out. Once the choice is made, no further switching is allowed.

The number of consumers coming forward each period to register for the operation is fixed for each area at the outset. We can easily incorporate growth in the overall size of the market over time into the model, but this is not done in the results presented here. They could therefore be interpreted as showing the market share of each hospital in a growing market. The 'period' in the model is deemed to be a week, though this could be varied if required.

In this version of the model, the number allocated to each area is drawn at random from a uniform distribution on [x, y]. These can be thought of either as individual agents or as representative agents of a larger number who have identical preferences.

We specify the initial capacity of each hospital, which is set to be equal to the number of consumers coming forward each period in the relevant area. So initially, every hospital is operating at full capacity and waiting lists are stable.

Quality and distance are both measured over the interval [0,1], and waiting times are scaled into this interval for the purposes of consumer choice. A wait of W_{max} weeks is deemed to be equal to 1, and waits below this level are scaled by (wait time in weeks/ W_{max}). A hospital with a wait time in excess of W_{max} is regarded as unacceptable by all consumers, regardless of its quality. Similarly, distance and time are scaled such that hospitals with q = 0 or d = 1 are regarded as unacceptable.

Both distance and waiting times are perceived perfectly for all hospitals by each consumer (who is given a specific admission date for a specified hospital at the point of choice). The quality of the local hospital is also perceived perfectly. However, the quality of all other hospitals is calculated by the j th consumer as $q_{kj} = q^*_k - \varepsilon_j$, where q^*_k is the true quality, and ε_j is drawn from a random uniform distribution on $[0, \mu]$. In other words, consumers are aware that they have imperfect information. They therefore make an allowance for this in their calculation of quality.

The utility obtained by each consumer at each hospital is calculated. If this is maximised by the local hospital, the consumer goes there. If not, a consumer from a given area chooses the best hospital with probability σ_k , where this parameter is drawn from a uniform distribution on [0,1] at the start of each model solution, and is allocated to all consumers in a given area throughout the course of the solution. The value of σ varies across localities. The parameter σ is a simple way of introducing into the model a complex range of factors relating to the ability and propensity of different social groups both to gather information and to exercise choice based on that information, as well as the loyalty of patients are their advisers to historic referred providers.

There are 4 types of consumer, and their preferences are calibrated on the basis of stated preference research carried out by the Department of Health (2004). They vary in the weights assigned to quality, waiting lists and distances. When each consumer enters the model, he or she is allocated to Type A with probability p_A, to Type B with probability p_B, and to Type C with probability p_C, and to Type D with probability 1 - $p_A + p_B + p_C$. The probabilities follow from Department of Health research findings on the stated preferences of consumers.

Each week, each hospital treats a number of patients which is equal to its capacity, if the waiting list is larger than capacity. If the waiting list is less, it simply treats this number.

Each hospital receives an identical and fixed amount of money for each consumer treated. For simplicity, this is normalised at 1, so that revenue per week is simply the number of consumers treated.

Individual hospitals differ in their level of quality, their effort, their efficiency, the number of consumers they treat, and their capacity. So each hospital has a special case of a more general cost function. It is important at this point to distinguish between effort and efficiency. Effort here is designed to indicate how hard hospitals aim to keep their costs down. Efficiency is intended to capture the structural efficiency of a hospital –determining the level of fixed costs. Both concepts indicate measures of how well a hospital turns cost into output, but the efficiency parameter is a fixed random draw, whereas effort is a variable that the hospital has the ability to change.

The cost function is based upon expert knowledge within the Department of Health and is given by:

$$(1 + q)^2*(1/(1 + e))*N_{op}*w_1 + w_2*N_{capac}$$

where q is the quality of the hospital and e is its effort, both of which are scaled to be in [0,1]. N_{op} is the number of operations performed in a week, and N_{capac} is the number of operations it could perform at full capacity. The parameters w_1 and w_2 represent the relative weights placed on the two expressions which together make up the cost functions. In the solutions of the model described below, w_1 is set equal to 0.2, and w_2 is drawn at random for each hospital from the uniform distribution on [0.6, 0.7]. The interpretation of w_2 is that this represents the structural efficiency of the hospital.

The three types of hospital differ both in the amounts of information which they consider in making decisions on quality, effort, and capacity, and in their motivations.

The ***Satisficer*** hospitals try to minimise effort subject to the constraint to break even. They pay little attention to market conditions. However, they will increase quality if they can do so at minimum effort without incurring costs. If they make a profit in any given year, in the first instance they simply slacken off, and their effort is reduced. Effort is reduced to the level at which they predict that they will not make a profit but will break even in the next year.

The costs they will incur is specified in the cost function above, but to calculate the break even level of effort, they need to predict N_{op}, the number of operations performed in a week. They also need to provide the level of capacity, N_{capac} . They (and all other types of hospital) aim to operate at full capacity, so they adjust capacity so that it is equal to the expected number of customers. The capacity is based upon simple extrapolation of the change in demand over the previous year,.

Where a Satisficer is still predicting that a profit will be made in the next year effortlessly, quality is increased until break-even is predicted. However, the fact that an increase in quality will, ceteris paribus, increase the number of consumers choosing this hospital is not taken into account, and the simple extrapolation rule for revenue is used, described in the previous paragraph. If Satisficer hospitals make a loss, in the first instance they increase effort to the level at which break-even is predicted for the next year.

If they increase effort to its maximum value of 1 and still predict that a loss will be made, they do at last take some account of market conditions, and examine their waiting lists. Each hospital is allocated a target waiting list as part of the initial conditions of the model. If the waiting list is less than the target, the hospital will reduce its capacity. If the waiting list is above target, the hospital will reduce quality in an attempt to drive people away.

The motivation of ***Not for Profit*** is to maximise their objective function, which is given by

$N_{op}*(1 - e)*q$

In other words, they aspire to increase the number of customers they serve, increase their quality, and reduce effort.

They make a projection of the number of customers over the next year as described above, and they invest in capacity [i.e. incur costs in the coming year] in order to provide for this number. They then choose q and e to maximise utility, given this capacity and subject to the need to break even financially. (If a loss has been made, this constraint implies that they target a profit in the coming year sufficient to offset the loss).

The new entrants from the ***Profit maximiser*** producers aim to make profit. This is of course a simplification. In reality, there will be profit maximiser hospitals which act more like not-for-profits or satisficers and vice versa. The model is designed to explore differences in behavioural types rather than institutional or ownership regimes.

It must also be stressed that profit maximisation is by no means an unambiguous concept in this model. For example, at the point of potential entry, hospitals already in the market are making and implementing decisions about their quality and capacity over the coming year. The decision by an entrant will obviously depend upon whether it can be presumed to have knowledge of these decisions, or whether it has to rely on knowledge which is in the public domain, namely the quality and capacity of existing hospitals in the previous period. As it happens, it seems more realistic to assume this latter rather than the former, which is what in fact we do. The decisions of profit maximiser hospitals are made with reference to the profits which they might make in the immediate year ahead.

At all times, these hospitals set effort equal to its maximum value of 1, which of course helps to minimise costs.

Provided that a profit can be made at the location, a new entrant from the profit maximiser sector must replace a failed hospital in the model immediately. In addition to replacing any hospitals which exit the market at any given time, at the end of each year one new entrant considers whether to enter the market and compete at the same location as an existing one. Again, the decision to enter is based upon whether or not a profit can be made. But potential entrants which are considering competing directly at the same location with an existing one are required to expect a minimum profit of 5 per cent of total revenue in the first year before they decide to enter.

The entrant, whether replacing a failed hospital or whether competing with an existing hospital, needs to decide its quality and capacity, its effort being assumed to be always equal to 1 in order to minimise costs. However, the assessment of the levels of quality and capacity to set is by no means straightforward.

The potential entrant is assumed to have knowledge of consumer behaviour, and uses this as a basis for forecasting demand. If it is replacing a failed hospital, it estimates the demand for its services at this location. If it is considering whether to compete with an existing hospital, it examines the location at which quality is the lowest which is on offer, and considers whether it is profitable to enter there. The entrants know that they have information on the quality and capacity of other hospitals only for the period which has just ended. They do not have information on the new levels of quality and capacity which are being set at that point in time, and they know that they do not know this.

The entrant knows the consumer demand functions and the percentages of consumers allocated to the four types of function, although it does not know the precise geographical location of the individual consumers by their demand types. The entrant calculates the number of customers it expects, based on the demand it would receive from all sites, making full use of its knowledge of patients' utility functions. It therefore takes into consideration the potential demand from other locations. It is aware that in each location there is a fixed probability, σ_i, that the consumers will switch from their existing local hospital if another offers better value. It is also aware of the fact that consumers realise they are uncertain about the quality of all hospitals apart from their initial, local one.

Initially, the new entrant chooses its level of quality. It examines the quality provided by any existing hospital at a location, and the qualities of the two hospitals which are nearest to the location. It decides to set a level of quality which is higher than the maximum level of quality provided by this set of hospitals. The entrant makes this choice on the grounds that a reasonably high level of quality is necessary both to establish its reputation and to assist its prospects of survival in the short term.

Given this level of quality, the entrant then decides the level of capacity at which it will make most profit, in the light of its knowledge of consumer demand subject to them having the existing average waiting list elsewhere in the system.

Once they have entered, the profit maximiser hospitals continue to try to make as much profit as possible in the year ahead. The first action which they take at the end of each year is to set quality as low as possible, subject to it being higher than the maximum of their geographically nearest competitors, and subject to the profit constraint. If they have made a loss, their forecasts for the next year lead them to cut capacity.

Finally, a rule specifies when a hospital is deemed to have exited the market. Profits and losses are cumulated from one period to the next. In this version of the model, a hospital is deemed to exit the market for the particular operation if it incurs a cumulative loss of 8 per cent of total cost in any given year.

3. Results with the base version of the model

We start the model with 10 hospitals, 5 of which are allocated at random to behave as Satisficers, and the remaining 5 to behave as Not for Profit. The number of consumers coming forward each week in each area is drawn from a uniform distribution on [10,100]. The average number of consumers is therefore 550. As noted above, each one of these can be thought of as being representative of whatever number of actual consumers is realistic.

The capacity of each hospital is set at the same number as the number of consumers coming forward each week, so that each hospital is initially operating at full capacity. The same length of waiting list is allocated to each hospital, at 19 weeks.

The initial quality of the Satisficers is set at 0.225 and of the Not for Profits at 0.5.

The consumer demand functions are based on stated preference consumer research carried out by the Department of Health (2004) and are as follows:

8 per cent are Type A, and their utility is given by

$U_A = quality^{0.8}*(1\text{-}wait)^{0.1}*(1\text{-}distance)^{0.1}$

23 per cent are Type B, and their utility is given by

$U_B = quality^{0.1}*(1\text{-}wait)^{0.8}*(1\text{-}distance)^{0.1}$

45 per cent are Type C, and their utility is given by

$U_C = quality^{0.2}*(1\text{-}wait)^{0.2}*(1\text{-}distance)^{0.6}$

24 per cent are Type D, and their utility is given by

$U_D = quality^{0.2}*(1\text{-}wait)^{0.2}*(1\text{-}2*distance)^{0.6}$

where in the above equations all variables are scaled in [0,1] .

All consumers coming forward each week in a given area are allocated a value of σ, the propensity to switch from the local provider if a higher level of utility could be obtained elsewhere. The values of σ are fixed throughout all the periods of each individual solution.

The model advances in a series of steps, or periods. In each period, the consumers coming forward for the operation choose their hospital. The model progresses for 52 weeks, when all hospitals in the market either exit or alter their offers in accordance with the above rules. At the same point in time, a profit maximiser hospital considers whether or not to enter at a location where a hospital has just exited the market. And another profit maximiser hospital considers whether or not to enter at a location where a hospital continues to exist. We run the model for 10 successive batches of 52 periods. In other words, for 520 weeks or 10 years.

A hospital is deemed to exit the market for the particular operation if it incurs a cumulative loss of 8 per cent of total revenue in any given year.

We describe the results obtained with 500 separate solutions of the model. This is sufficient to describe the range of results which the model generates. The model is programmed in Matlab, and a single solution on a standard PC takes around 1 minute to solve for 10 hospitals over a period of 520 weeks.

In general, we find that:

- average quality improves
- average waiting times fall
- consumer utility increases
- capacity utilisation falls

In each case, averages are the appropriate weighted average by number of consumers.

Figure 1 summarises the key findings for consumers in the simulations. This plots results from the 500 solutions at the end of year 10 in each of the solutions.

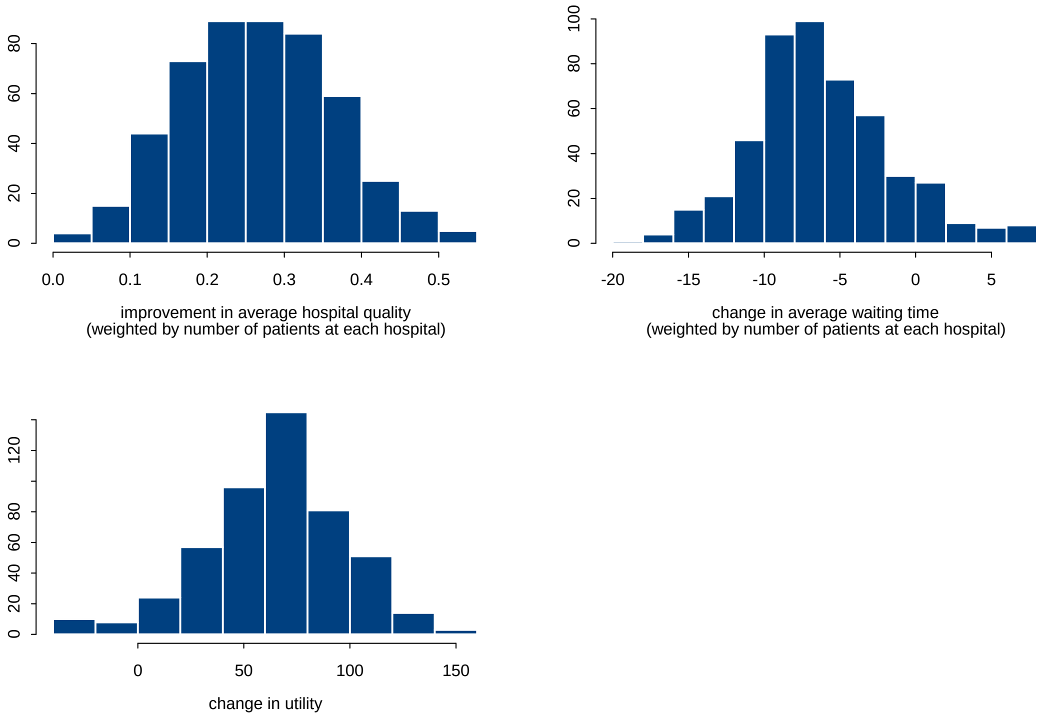

Figure 1 *Changes in average quality, average waiting times and total consumer utility. Summary of results from 500 separate solutions of the base model, at year 10*

In order to put the results into perspective, we need to recall the scales over which the various variables are measured. At the outset of each solution, the average weighted quality of hospital is 0.3625. The 5 Satisficer hospitals have quality of 0.225 and the Not for Profit 0.5. The number of consumers allocated to each one is chosen at random, so the actual weighted initial quality will vary from solution to solution, but on average it will be 0.3625. Quality is scaled within the interval [0,1], so the increases in quality which are observed are substantial with respect to the initial level, the average being 0.26.

In the consumer demand functions, waiting times are also scaled in [0,1], but here we show them in terms of actual weeks. The initial level is 19 weeks for each hospital. 90 percent of solutions show average waiting times to fall. The average (weighted) fall of almost 6 weeks is again non-trivial compared to the initial 19 weeks' wait.

We can examine in a little more detail the outcomes on quality, waiting times and utility.

Increase in quality: OLS regression

```
Coefficients:
                  Value Std. Error  t value Pr(>|t|)
   (Intercept)   0.4538   0.0230    19.7225   0.0000
final.num.hosp  -0.0307   0.0018   -16.9107   0.0000
  remaining.PM   0.0308   0.0019    15.9512   0.0000

Residual standard error: 0.06806 on 497 degrees of freedom
Multiple R-Squared: 0.5128
```

where final.num.hosp is the number of hospitals at the end of year 10 and remaining.PM is the number of profit maximiser hospitals. The number of the other two types is not statistically significant.

Figure 2 shows what happens to the number of providers in the various sectors.

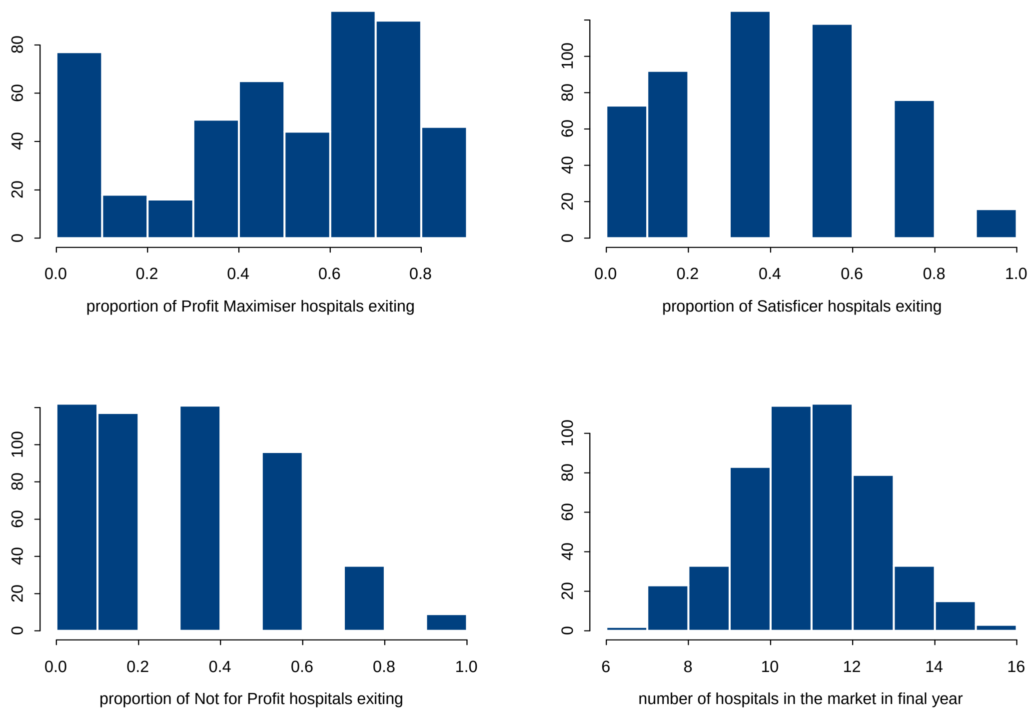

Figure 2 *Proportions of Hospitals Exiting and Number of Hospitals in the market after 10 years*

The exit rate for the new profit maximiser entrants is high, the average being 50 percent across the 500 solutions, but this is entirely typical of what happens when markets are opened up to competition. The incumbents have a survival advantage because of consumer familiarity with them, and new entrants have relatively high death rates. Land line telecommunications and the domestic supply of energy are examples of such phenomena.

4. Alternative scenarios

In terms of a more formal analysis of the sensitivity of results in this particular model, we now consider three scenarios in which, in turn:

- the exit rule is made much more lax. Specifically, a hospital now exits only if its losses exceed 20 per cent of turnover in any given period
- the tariff per patient is increased by 10 per cent from 1 to 1.1
- the probability of consumers switching to the utility maximising provider [σ_j] is drawn from [0.5, 1] rather than [0, 1] as in the base solution

The results of these three variants are summarised in Table 1, in each case taking the average of 500 separate solutions of the model and comparing them with the base case described above.

Table 1

	Change in average waiting time, weeks	**Increase in quality**	**Increase in utility %**
Base case	-5.92	0.265	17.1
Lax exit rule	-5.08	0.210	16.3
Higher tariff	-6.36	0.466	28.0
Higher propensity of consumers to switch	-8.45	0.349	18.4

The reasons for the differences in average outcomes from those of the base case are essentially as follows. The lax exit rule makes it easier for Satisficer and Not for

Profit hospitals to survive. The higher tariff makes it easier for Satisficer and Not for Profit hospitals to survive. But it also makes it easier for profit maximiser hospitals to enter. In particular, the requirement on entrants competing at the same location with an existing provider to make 5 per cent profit, is more easily satisfied. The higher propensity to switch (σ) makes entry easier for the profit maximiser. Note, however, that the increase in consumer utility is not much above that of the base case, despite the fact that average wait times are considerably less and the increase in quality definitely higher. But consumers of Types C and D, who attach high weights to distance, suffer in solutions in which locations appear in which no hospital can survive.

5. Calibrating to actual data

We gathered an initial set of data for the 13 hospitals in the Birmingham and Black Country Strategic Health Authority area which carry out the surgical procedure defined formally as HRG H02, known more commonly as elective primary hip replacement. This was done partly at a meeting with representatives of Primary Care Trusts from the area, and partly by the Health Economics Unit at Birmingham University. These results have been presented to the Primary Care Trusts involved in providing the data and were judged to be plausible.

We carried out 500 separate solutions of the model over a 10 year period. Table 2 sets out for hospitals with zero probability of exit, the probability of a hospital exiting from the provision of primary hip replacement surgery, its profit margin in year 10 if it survives, and its capacity growth over the 10 years. In the presentation of these results, although the total set of hospitals is identified, at the request of the data providers the names of the individual hospitals in the outcomes are removed.

Table 2

Provider	Probability of exit / takeover	Profit margin if survives (%)	Capacity growth (%)
A	0	0.5	-43
B	0	11.5	486
C	0	16.7	195

D	0	10.7	190
E	0	9.7	214
F	0	13.5	202

The capacity growth is large in percentage terms, but these providers in general start with low levels of capacity, so the absolute numbers are not large. Hospital A evolves a strategy for survival which involves cutting capacity quite sharply, and relying on the fact that its predominantly working class customer base will only have a low probability of switching away from their local supplier.

Table 3 shows the results for those hospitals with a low probability of exit. We also include information on the growth in capacity when the hospital is driven out of providing this operation, and its ability to do so is taken over on the site by a private provider. This is to judge whether or not there are inherent problems of survival on any particular site, such as might occur if a hospital had two large, high quality providers as near neighbours. In this instance, hospital J in general has to reduce capacity regardless of whether it survives or whether its procedures are taken over by a private sector provider.

Table 3

	Probability of exit / takeover	**Profit margin if survives (%)**	**Capacity growth if survives (%)**	**Capacity growth if replaced (%)**
G	9	0.1	-20	14
H	14	0.3	-31	53
I	17	-1.8	17	85
J	18	-0.6	-45	-19

Finally, we identify those hospitals at most risk of having to exit the market.

Table 4

	Probability of exit /	**Profit margin if**	**Capacity growth if**	**Capacity growth if**

	takeover	survives (%)	survives (%)	replaced (%)
K	82	1.1	-67	9
L	38	-2.4	9	12
M	33	-2.1	25	-1

Hospital K fares by far the worst. Its initial quality is very low, and it has a large waiting list, well over the Strategic Health Authority 6 month target. It is also based in one of the most affluent areas of Birmingham and the Black Country, so its customer base is likely, once given the choice, to be very willing to switch to better alternatives. Essentially, it is a race as to whether provider K can cut its capacity (and hence costs) sufficiently quickly to offset its loss of customers before it is deemed to exit because of its cumulative losses.

6. Conclusion

We illustrate in this paper a model developed for the economics unit in the UK Department of Health as an input into the advice which they give to policy makers on the planned creation of consumer choice in the provision of elective surgery in the National Health Service. The model can be used to examine a range of policy questions, such as the effects of different financial exit rules, different levels of tariffs, different consumer behaviour and so on.

Importantly, we calibrated the model to actual data on the surgical procedure HSG H02, primary hip replacement, carried out at present in 13 hospitals in the Birmingham and Black Country Strategic Health Authority. The data was partly provided by NHS professionals in the area, and the results were subject to scrutiny by them. Whilst these results cannot be deemed to be completely definitive and would need further examination before being used in policy decisions in the area, they were judged in general to be realistic by the practitioners to whom they were made available.

In this market, the price available to producers is fixed. Consumers have imperfect information and do not necessarily take their optimal decision even with respect to the limited degree of information which they have. Producers adopt relatively myopic decisions rules. Nevertheless, in general the introduction of consumer choice results in an increase in the quality of the product, a fall in waiting times and a rise in consumer welfare. A very similar result was obtained by Maroulis and Wilensky (2005), who developed an agent-based model to simulate the effect of introducing parent choice into the Chicago public school system.

References

Appleby J, Harrison A and Devlin N, 2003, *What is the real cost of more patient choice?*, King's Fund London

Department of Health, 2004, http://www.londonchoice.nhs.uk/documents/Patient-Experience-Study-LPC-(RAND).pdf

Maroulis S. and Wilenksy U., 'School Districts as Complex Adaptive Systems: A Simulation of Market-based Reform', paper presented at the 3rd Lake Arrowhead Conference on Human Complex Systems, UCLA, May 2005

Webb, S. 'Proposed reforms in healthcare will not lift standards', *Financial Times*, 14 September 2005