THE UNIVERSITY of York

Discussion Papers in Economics

No. 1999/38

Modelling the Demand For and Supply of Elective Surgery: A Duopoly Model

by

Ana Xavier

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Modelling The Demand for and Supply of Elective Surgery: a Duopoly Model*

Ana Xavier
Department of Economics and Related Studies
University of York

Abstract

In this paper I model the demand for and supply of elective surgery using a modified Hotelling framework in which waiting time, money and distance costs are determinants of the demand for hospital care. Hospitals compete with each other by varying supply and hence their waiting times. I consider both the situation where GPs do not hold a budget (and thus Health Authorities pay for health care), and the situation where they are given budgets to buy care for their patients. Waiting time increases when production of care becomes more expensive, when the benefit obtained from treatment increases, when the unit cost of distance decreases, and when the importance given to the delay by the hospitals decreases. Moreover, the higher the money price (and if greater than the marginal cost of producing hospital care) the lower the waiting time. If the money price paid by GP fundholders is higher than that paid by HAs, fundholding patients pay a lower time price.

1 Introduction

In this paper I analyse the demand for and supply of elective surgery. The aim of this paper is to model waiting times as a price or a time cost and

*I wish to thank Prof. Hugh Gravelle, Prof. P. C. Smith, and Alberto Bacchiega, as well as those attending my session at the European Economic Association Meeting 1998, Berlin, for their helpful comments and suggestions.This research was financially supported by The Foundation for Science and Technology - Praxis XXI, Portugal, and by the Centre for Health Economics, University of York. Address for correspondence: Department of Economics, University of York, York Y01 5DD, UK. Email address: amcx100@york.ac.uk.

therefore I analyse delay as a determinant of choice between hospitals. In order to do this I make use of a duopoly model namely a modified Hotelling framework.

In the United Kingdom and in the National Health Service (NHS UK) a patient has a first appointment with her General Practitioner (GP) after which, if the GP finds it necessary, the patient may be referred to a hospital specialist or consultant. The specialist then decides whether the patient should be admitted to hospital as an inpatient. If so, a patient is either admitted for emergency treatment or put on a waiting list depending on whether waiting is critical or not. This analysis concerns those patients that do not need urgent care and are put on a waiting list for elective care.

During the first 25 years of the NHS there were about half a million people waiting for hospital treatment. The number has increased to around 1.2 million despite the fact that there has been an increase in hospital capacity and an increase in the numbers treated (Culyer and Cullis [9], Smith and Martin [30], Pope [23]). This analysis examines waiting times as a means of rationing non-emergency treatment or elective surgery.

The view that time spent waiting acts as a cost or a time price dates back to the studies of Nichols *et al.* [19], Barzel [2] and Porter [24]. When commodities are provided in a fixed quantity at a low or zero (money) price and not all the demand is satisfied, the commodity has to be allocated on a first come - first served basis. The opportunity cost of the time spent waiting acts just like a money price in that it deters demand and it clears the market. There is a time price at which quantity supplied equals quantity demanded. People are willing to join the queue as long as the value obtained is greater than or equal to the cost incurred. At the market clearing point the full price (time plus money price) equals the value to the marginal consumer who has a zero surplus. Because there are no property rights assigned (Cheung [6]) people have to queue to obtain the commodity so that rationing by queue wastes resources (scarce time).

Barzel's analysis [2] predicts that an increase in capacity leads to a decrease in the per unit waiting time and an increase in the numbers joining the queue per period of time. This theory was well accepted as it could explain why an increase in capacity could lead to an increase in the flow into the queue.

The analysis was adapted to inpatient health care by Lindsay and Feigenbaum [16]. It is argued that for certain commodities people queue in absentia on a waiting list (some examples being elective surgery and social housing). Hence, the authors assume that waiting does not impose a cost in wasted time itself but waiting acts as a cost in that it lowers the value of the commodity (in our case inpatient treatment) when received. Suppose the consumer's utility is:

$$U = be^{-gt} - C$$

where b stands for the benefit obtained from treatment (which depends on the consumer's attributes, v) when received, g is the decay rate (expressing the combined effect of market time preference and preferences due to fashion, location, circumstance, and whim which affect the value of a good), t is the delay or time waiting for treatment, and C stands for the cost of joining the queue thus including things such as examinations and medical advice. It is assumed that people will join the queue as long as the present benefit of the commodity when delivered, be^{-gt}, is greater or equal than its cost, that is, when $be^{-gt} \geqslant C$. For the marginal joiner the equality holds while the intra-marginal joiners can obtain a benefit higher than the cost.

According to Lindsay and Feigenbaum [16], one can then solve the above equation in order to find the critical waiting time, which is the maximum time the individual is willing to wait to get the treatment. It can be seen that different individuals may have different critical waiting times if the benefit from treatment obtained, b, differs, or if the decay rate, g, differs, or when their C differs. Ceteris paribus, increases in C or g reduce the critical waiting time, whereas increases in b increase it.

Lindsay and Feigenbaum [16] then build a 'joining' function (assuming the decay rate is the same for all those in the same queue) which depends on consumers attributes and on the number of people. This function has a negative slope because when waiting time increases some potential joiners will not join the queue as discounted benefits are reduced.

The supply side is not developed in detail by Lindsay and Feigenbaum [16] but it is hypothesised that the waiting time for a commodity exercises a positive influence on the quantity supplied. The supply is defined as $Q^s(t, a)$ with $\frac{\partial Q^s(t,a)}{\partial t} > 0$ and where a stands for a set of factors other than waiting time.

Equating the demand and the supply (for a constant g and C) allows us to find the equilibrium waiting time, that is, $Q^d(t) = Q^s(t) \Longrightarrow t^* = t(C, b, a)$. As argued by the authors, the wait acts very much like a price. Given that $\frac{\partial Q^d(t)}{\partial t} < 0$ and that $\frac{\partial Q^s(t)}{\partial t} > 0$ the wait will converge to the equilibrium. Therefore, waiting time clears the market by making the good less valuable (Lindsay and Feigenbaum [16]).

Changes in capacity or changes in decay rate lead to changes in the equilibrium waiting time and, according to the authors [16], this may also help explaining why there are different waiting times and different waiting lists sizes for different conditions. For diseases with a lower decay rate waiting times are higher and the list length may also be higher.

The Lindsay and Feigenbaum [16] model has been further developed by Goddard *et al.* [13] and by Smith and Martin [30] in order to consider patients' choices between public care (that implies a time price) and private care (which implies a money price but no time price) as well as to consider the case where public hospital managers care about the waiting time.

The next section presents the context for the analysis. I start by assuming that patients do not differ in terms of the benefit they receive from treatment but do differ in terms of the distance from hospital. The equilibrium supply and waiting times are determined and some comparative statics are computed. In section 3 I introduce fundholding. As an extension to section 2 I allow for the benefit to vary across patients in section 4. Section 5 contains my conclusions.

2 The Model

The aim of this paper is to model demand for and supply of elective surgery and the consumers' choice between hospitals taking into consideration time costs and distance costs. The model developed here extends the models mentioned earlier both on the demand and on the supply sides. On the demand side both time and distance costs are considered and the choice between hospitals is analysed. On the supply side, competition between two hospitals is introduced and the supply is endogenous.

In order to analyse the demand and supply of elective surgery when patients choose between hospitals I adopt a modified Hotelling [14] framework. There are two hospitals (hospital 1 and hospital 2) each located at one of the extremes of a road of length l (see figure 1). N potential elective patients are uniformly distributed along the road. Patients vary in terms of location but obtain the same benefit from treatment.

Each patient is registered with a general practitioner that he visits after developing some disease symptoms, and with whom he takes the decision of demanding health care. I assume there are two general practitioners (GPs) serving all those potential patients and it is assumed each GP has half the patients situated at each point on the road between the two hospitals.

As well as being registered with a general practitioner, a patient belongs to a Health Authority (HA). The government allocates funds to Health Authorities and these must purchase care for their patients. Hence, hospital care provided within the National Health Service is paid by Health Authorities. GPs may be allocated budgets and run these to buy care for their patients. I explore this situation in section 3. The purchasers in this case are two: a Health Authority (paying for the patients registered with the GP that does not hold a budget) and a GP budget holder paying for his patients.

Patient Utility Function It is assumed that the utility functions represent the preferences of the consumers/patients. The utility functions represent the monetary value of the benefit obtained from treatment at each hospital. Nevertheless, it is important to note that GPs are the ones generally deciding with the patient and/or on behalf of the patient whether the latter needs to visit the specialist. I assume that GPs are acting as perfect agents for the patients and therefore use their utility functions. This assumption is important and will be relaxed in section 3.

To make the analysis tractable I make use of a function that is linear on the delay:

$$U_i = b(\gamma - t_i) - C_i - kd_i \tag{1}$$

where U_i is the utility from going to hospital i, with $i = 1, 2$. b is the benefit from treatment, t_i is the delay at each hospital, C_i is the cost (e.g. examinations and legal advice) one may incur to get on a waiting list at hospital i, k is the unit cost of distance, and d_i is the distance to hospital i. γ is just a scaling parameter. Thus, the utility from being treated at hospital i, with $i = 1, 2$, depends positively on the benefit received from treatment, b, and decreases with the delay in a linear way. The utility also decreases with the costs of getting on the list as well as with the distance costs. When utility becomes smaller than zero, $U < 0$, no care is sought.

There seems to be evidence that when choosing a hospital patients are quite sensitive to distance and travel time.[1] There also seems to be some reluctance from the part of the referring physicians to have the patients seen in distant hospitals. Moreover, in the presence of third party payers (insurance or free public care) it is suggested that travel time and waiting time are likely to replace out of pocket money costs as determinants of the demand and the choice of the hospital from which to obtain care.

In figure 1 the downward sloping lines, U_1 and U_2, represent the utility (measured in money terms) a patient can obtain when going to hospital 1 or 2, respectively.

The utility obtained when going to hospital i decreases when distance to that hospital increases (which is represented on the negative slope of the

[1] There is considerable literature analysing the importance of distance in determining the choice of the health care facility using, for example, gravity models and conditional logit models. See: Acton[1], McGuirck et al.[18], Cohen et al.[8], Studinicki[32], Folland [12], Weiss et al.[33], Erickson et al.[11], Rice[26], Shannon et al.[29], Roghmann et al.[28], Staten et al.[31], Dranove et al.[10], Robinson et al.[27], CRD Report 8[7].

lines). For each point of that road, the utility will be lower the higher the delay (represented by a downwards shift of the lines).

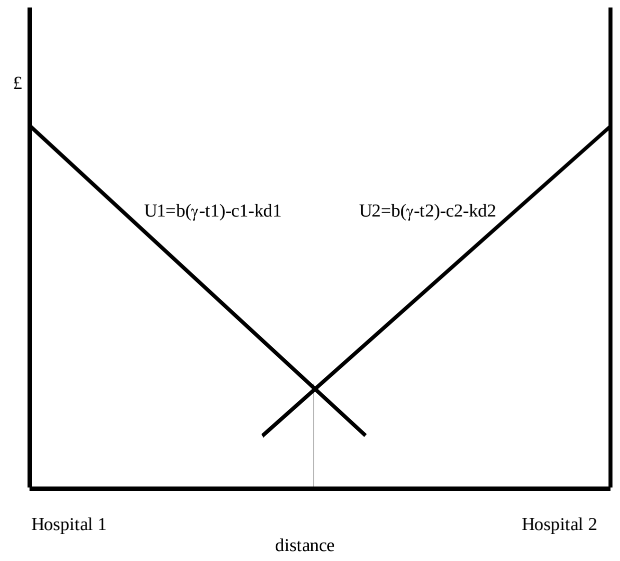

Figure 1: Hotelling Framework

Assuming that all consumers go to some hospital I obtain the demand for each hospital by equating the two utilities, $U_1 = b(\gamma - t_1) - C_1 - kd_1 = b(\gamma - t_2) - C_2 - kd_2 = U_2$, knowing that $d_1 + d_2 = l$.

Thus, one obtains a value d_1^c representing the point on the road such that patients on the left go to hospital 1 and patients to the right go to hospital 2. Since the density of consumers at any point is $\frac{N}{l}$ (patients are uniformly distributed along the road), the demand function for hospital i is as follows, with $i \neq j$, [2]

$$Q_i^d(t_i, t_j.) = \frac{N}{l}\left[\frac{b(t_j - t_i) + C_j - C_i + kl}{2k}\right] \tag{2}$$

Proposition 1 *The quantity demanded at hospital i (with $i = 1, 2$), decreases with t_i, C_i and k, and increases with t_j, and C_j ($j \neq i$).*

Note that in what follows I only look at the situation where the two lines representing the utility at each hospital cross. This corresponds to having all potential elective patients served and thus the only question one needs

[2] The demand for hospital i is calculated by integrating the population density, $Q_1^d = \int_0^{d_1^c} \frac{N}{l} dx$, with $d_1^c = \left[\frac{b\ (t_2 - t_1)\ + C_2 - C_1 + kl}{2k}\right]$.

to answer is which hospital will be chosen and what will be the average waiting time patients will face at each hospital. If the utility functions do not cross there is a group of people that will not be referred to hospital.[3] This corresponds to the demand for no care and, in that context, time and distance affect not only the choice of hospital but also the total number of referrals to hospital.

Hospital Utility Function Each hospital is assumed to maximise a utility function that depends on hospital notional revenue, on its production costs and on waiting times.[4] The utility function for hospital i is:

$$W_i = \alpha Q_i^d (t_i, .) - \beta (Q_i^s)^2 - wt_i^2 \tag{3}$$

where α can be seen as the unit price of care paid to the hospital by the purchaser (I initially assume it to be a Health Authority rather than the GP or patient), β is a production cost parameter, w is a parameter representing the weight the hospital puts on the delay experienced by its patients. Q_i^d is the quantity demanded and Q_i^s is hospital capacity or supply (as in this model the decision of what to supply in terms of elective surgery is the same as the decision on capacity).[5] An increase in α increases the utility of the hospitals whereas an increase in β or t_i decreases it.

I have chosen a duopoly situation as it is often argued that a hospital market structure is an oligopsonistic one or a differentiated oligopoly where hospitals behave as Cournot oligopolists (see Booton [3], Noether [20], Calem and Rizzo [5]).

Moreover, the literature concerning hospital market structure provides evidence that non-price competition such as competition in quality dimensions plays an important role in that market. It is argued that insurance coverage makes patients quite insensitive to money price (Pauly and Redisch [22], Joskow [15], Noether [20]) even though some forms of price competition have seen their importance increased (Robinson and Luft [27]). The current study can be seen as an example of non-price competition where the dimension along which hospitals compete is waiting time for in-patient treatment.

[3] In this case the following conditions must hold: $U_1 = 0$, $U_2 = 0$, and $d_1^c + d_2^c < l$.

[4] The setting chosen may be seen as a long run maximisation problem where capacity is not fixed but endogenous to the problem and equates to supply. A different problem would be to assume hospitals had a fixed capacity and had to determine the supply of elective surgery when capacity is limited and can be allocated between elective and emergency cases.

[5] One can also see α as the unit benefit hospital obtain from treating patients and as such it could be seen as something broader than just the money price received per patient. It would thus be some monetary measure of what hospitals gain from treating patients including a possible money price.

The price of care, α, paid by the health authority is assumed to be fixed, that is, charges for hospital services are exogenously determined by third party payers (see Calem and Rizzo [5]). α is initially assumed the same for both hospitals. Later I investigate the situation where prices, α, as well as β and w, may differ across hospitals.

The quadratic cost function, $\beta\left(Q_i^s\right)^2$, reflects the assumption of decreasing returns to scale (CRD Report 8). β is the same for both hospitals; hospitals have the same marginal and average cost of producing elective care.

Waiting time enters the hospitals utility function directly. This means that a hospital has a direct interest in waiting times per se. A possible reason why delay may be important to hospitals relates to the fact that waiting time is an important element of the contracting process in the NHS. Health authorities want to contract with the trusts offering the smallest delays. Therefore, the lower the delay offered the higher is the probability that a contract is signed with a health authority and thus the higher the hospital utility. It can also be thought that health care providers have a degree of altruism and care about the disutility waiting time imposes on patients. The preference weight, w is initially assumed to be the same for both hospitals so that both give the same importance to the delay.

I consider the hospital as a whole, as one entity with well defined preferences. One can think that the hospital objective function is the utility function of the hospitals managers, following Smith and Martin's specification [30]. In this case, hospital managers may also care about the waiting time because it is included in the set of indicators by which performance of managers is judged.

Taking the action of the other hospital as given, each hospital chooses the quantity to supply as well as the waiting time in order to solve

$$\begin{aligned}\max_{Q_i,t_i} W_i &= \alpha Q_i^d\left(t_i,.\right)-\beta\left(Q_i^s\right)^2-wt_i^2\\ \text{subject to } Q_i^d &\leq Q_i^s.\end{aligned}$$

The Lagrangian function is

$$L_i=\alpha Q_i^d\left(t_i,.\right)-\beta\left(Q_i^s\right)^2-wt_i^2+\mu\left[Q_i^s-Q_i^d\left(t_i,.\right)\right]$$

and the first order conditions of the maximisation problem are

$$\begin{aligned}\frac{\partial L_i}{\partial t_i} &= \alpha\frac{\partial Q_i^d}{\partial t_i}-\mu\frac{\partial Q_i^d}{\partial t_i}-2wt_i\leqslant 0,\ t_i\geqslant 0,\ \frac{\partial L_i}{\partial t_i}t_i=0\\ \frac{\partial L_i}{\partial Q_i^s} &= -2\beta Q_i^s+\mu\leqslant 0,\ Q_i^s\geqslant 0,\ \frac{\partial L_i}{\partial Q_i^s}Q_i^s=0\\ \frac{\partial L_i}{\partial \mu} &= Q_i^s-Q_i^d(t_i,.)\geqslant 0,\ \mu\geqslant 0,\ \frac{\partial L_i}{\partial \mu}\mu=0.\end{aligned}$$

However, one is interested only in the situation where a positive quantity of care is supplied, that is, $Q_i^s > 0$, and thus $\mu = 2\beta Q_i^s$. It can further be assumed that $\mu > 0$, implying that the third constraint is binding and $Q_i^d = Q_i^s$. The reason for having such an equality is quite intuitive. When Q_i^s has a positive marginal cost it can never be optimal to have spare capacity so the constraint $Q_i^d = Q_i^s$ always binds at the optimum. As a consequence, as long as quantity demanded is positive, quantity supplied will also be positive.

When $Q_i^s > 0$ there can be two solution types: one with $t = 0$ and one with $t > 0$. When $t > 0$, increasing supply has a direct marginal cost of $2\beta Q_i^s$ but reduces the cost of waiting by $2wt_i$. Hence, the total marginal cost lies below the curve define by $2\beta Q_i^s$ when t is positive. Marginal revenue is a step function, equal to α when increasing supply will lower the waiting time and increase demand, but equal to zero when waiting times have fallen to zero and demand cannot be increased by increasing supply.

The situation where $t = 0$ corresponds to the case where α is high. A high price of care is paid and quantity supplied equates the total demand, which can be seen in figure 2. If on the contrary, α is low such as in Figure 3 marginal revenue equates total marginal cost when this is the line below $2\beta Q_i^s$, that is, $\alpha = 2\beta Q_i^s - \frac{2wt_i}{\left(\frac{\partial Q_i^d}{\partial t_i}\right)}$.[6]

I shall consider the context where waiting times may be positive. Bearing the above in mind one can solve the above problem with respect to the delay, t_i, by substituting the constraint $Q_i^d = Q_i^s(t_i, t_j, .)$ into the hospital objective function:

$$W_i = \alpha Q_i^d\left(t_i, .\right) - \beta\left(Q_i^d\left(t_i, .\right)\right)^2 - wt_i^2$$

The first order condition is

$$\frac{\partial W_i}{\partial t_i} = \alpha\frac{\partial Q_i^d}{\partial t_i} - 2\beta Q_i^d\frac{\partial Q_i^d}{\partial t_i} - 2wt_i = 0$$

Substituting both expressions, $Q_i^d(.)$ and $\frac{\partial Q_i^d}{\partial t_i}$, into the condition above and solving with respect to t_i, with $i = 1, 2$, one can obtain the reaction function for hospital i:

$$t_i(t_j, .) = \frac{\beta b^2N^2}{4wl^2k^2 + \beta b^2N^2}t_j + \frac{\beta bN^2(C_j - C_i + kl) - bNkl\alpha}{4wl^2k^2 + \beta b^2N^2} \tag{4}$$

Solving the pair of simultaneous equations defined by (4) one can obtain the equilibrium waiting times as well as the equilibrium quantities.

[6] Even when rationing is only by delay and suppliers do not care about delay, that is, even when $w = 0$ there can a situation where equilibrium implies $t > 0$.

When C_1 equals C_2 the symmetric Nash Equilibrium can be obtained. This is:[7]

$$t_i^* = \frac{\beta b N^2 - bN\alpha}{4lkw} \text{ and } Q_i^{s*} = \frac{N}{2} = Q_i^d$$

Proposition 2 *When hospitals present symmetric characteristics, the same output is supplied (each hospital supplies half the market), and the waiting time is the same at both hospitals.*

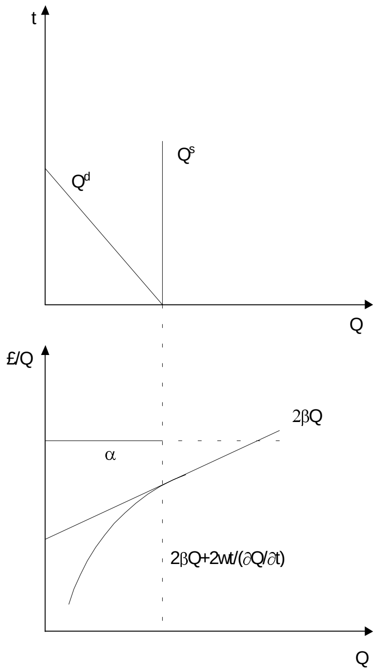

Figure 2

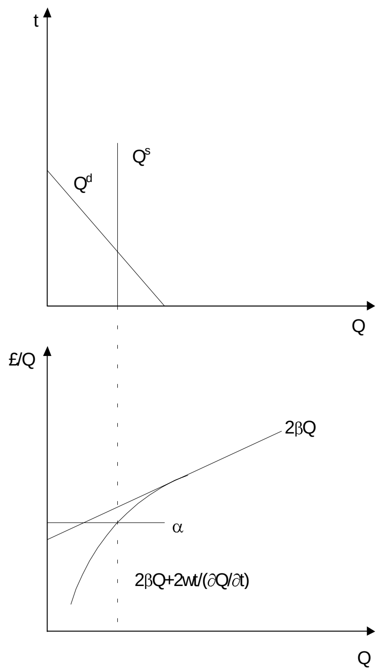

Figure 3

It can be seen that the equilibrium waiting time increases:

- when β increases (when production becomes more expensive each hospital wishes to take fewer patients),
- when b increases (and thus the maximum amount of time one is willing to wait increases),
- when k decreases (a decrease in k leads to an increase in the quantity demanded as access costs are reduced, and the critical waiting time increases),

[7]Note that the second order condition is $\frac{\partial^2 L_i}{\partial t_i^2} = -2w < 0$. Therefore, we have a maximum.

- when w decreases (that is, when hospitals care less about the waiting time they offer patients),
- when α decreases, (less is paid to each hospital)
- when N increases, (that is, more people need care).

The asymmetric equilibria ($C_1 \neq C_2$, $\alpha_1 \neq \alpha_2$, $\beta_1 \neq \beta_2$, or $w_1 \neq w_2$) are more complicated.

For example, when $C_1 \neq C_2$ one has

$$t_i^* = \frac{\beta b N^2 - bN\alpha}{4lkw} + \frac{\beta b N^2 (C_j - C_i)}{4wl^2k^2 + 2\beta b^2 N^2}$$

When $\alpha_1 \neq \alpha_2$

$$t_i^* = \frac{\beta b N^2}{4lkw} - \frac{Nb\left[\beta b^2 N^2 \alpha_1 + \left(4wl^2k^2 + 2\beta b^2 N^2\right)\alpha_2\right]}{4lkw\left(4wl^2k^2 + 2\beta b^2 N^2\right)}$$

When $\beta_1 \neq \beta_2$ the equilibrium waiting time is

$$t_i^* = \frac{\beta_i b N^2 \left(4wl^2k^2 + 2\beta_j b^2 N^2\right)}{4lkw\left[4wl^2k^2 + b^2N^2\left(\beta_i + \beta_j\right)\right]} - \frac{bN\alpha}{4lkw}$$

and finally if $w_1 \neq w_2$ one has

$$t_i^* = \frac{\beta b N^2 \left(2w_j l^2 k^2 - \beta b^2 N^2\right)}{2lk\left[4w_i w_j l^2 k^2 + \beta b^2 N^2 \left(w_i + w_j\right)\right]} - \frac{bN\alpha\left(2w_j l^2 k^2 + \beta b^2 N^2\right)}{2lk\left[4w_i w_j l^2 k^2 + \beta b^2 N^2\left(w_i + w_j\right)\right]}$$

Table 1 summarises the results when hospitals are asymmetric in their characteristics.

TABLE 1

Cases	Waiting Times	Capacities
Symmetric case	$t_1 = t_2$	$Q_1^s = Q_2^s$
$C_1 < C_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$\beta_1 < \beta_2$	$t_1 < t_2$	$Q_1^s > Q_2^s$
$\beta_1 < \beta_2$ and $C_1 > C_2$	$t_1 < t_2$	$Q_1^s > Q_2^s$
$\beta_1 > \beta_2$ and $C_1 < C_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$\beta_1 < \beta_2$ and $C_1 < C_2$	?	?
$\alpha_1 < \alpha_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$\alpha_1 < \alpha_2$ and $C_1 < C_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$\alpha_1 > \alpha_2$ and $C_1 > C_2$	$t_1 < t_2$	$Q_1^s > Q_2^s$
$\alpha_1 > \alpha_2$ and $C_1 < C_2$	?	?
$w_1 < w_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$w_1 < w_2$ and $C_1 < C_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$w_1 > w_2$ with $C_1 > C_2$	$t_1 < t_2$	$Q_1^s > Q_2^s$
$w_1 < w_2$ and $C_1 > C_2$	?	?
$\beta_1 > \beta_2$ and $C_1 < C_2$ and $w_1 < w_2$	$t_1 > t_2$	$Q_1^s < Q_2^s$
$\beta_1 > \beta_2$ and $C_1 < C_2$ and $w_1 > w_2$	?	?

Proposition 3 *The hospital that imposes a higher money cost, α, on the consumers has a lower waiting time and supplies a larger output. The hospital that puts a higher weight, w, on the delay offers a lower waiting time and a higher capacity. A hospital with a higher production cost, β, offers a higher waiting time and a smaller capacity. The hospital for which the cost of getting on the list, C, is higher has a lower waiting time.*

3 Introducing fundholding

In this section I relax the assumption that GPs act as perfect agents for the patients. A GP is allowed to hold funds with which he can buy services for their patients. Through holding budgets and buying care for their patients, GPs are made aware of health care costs and may take these into consideration when deciding whether to refer a patient to hospital. In this case they become imperfect agents. As a consequence, they maximise their own utility function although taking into consideration patients' benefits and costs. This section may help in explaining the consequences of introducing fundholding in the NHS, particularly the consequences for the patients in terms of waiting times.

It is now assumed that one of the GPs becomes a budget holder whilst the other continues not to hold any funds. As before each GP has half the patients at each point on the road of length l. The GP who becomes a budget holder has to buy the services for his patients. Hence, there are two types of utility functions, U_i^{nfh} and U_i^{fh}, with $i = 1, 2$, respectively the utility a patient obtains at hospital i if a non-fundholder patient and the utility a patient obtains at hospital i if a fundholder patient (which is the utility of a non-fundholder patient less the cost of the fundholding GP). These are:

$$U_i^{nfh} = b(\gamma - t_i) - C_i - kd_i \tag{5}$$

and

$$U_i^{fh} = b(\gamma - t_i) - C_i - kd_i - \lambda\alpha_i^{fh} \tag{6}$$

where all the parameters are as before with the exception of α and λ. α_i^{fh}, represents the cost for the GP of buying health care for her patients at hospital i and λ can be seen as the weight (or importance) a GP attributes to that cost. λ may be seen as a parameter measuring GP's altruism. If λ is zero it means that the GP holding a budget does not take into consideration the cost of care that she has to pay for the patient when deciding if a patient needs care. A $\lambda > 0$ stands for the case where budget holders do take into consideration the cost of care and the higher λ the less altruistic is the budget holder's behaviour.

I obtain the demands for each hospital given each type of GP by solving:

$$\begin{aligned} U_1^{nfh} &= b(\gamma - t_1) - C_1 - kd_1 = b(\gamma - t_2) - C_2 - kd_2 = U_2^{nfh} \\ U_1^{fh} &= b(\gamma - t_1) - C_1 - kd_1 - \lambda\alpha_1^{fh} = b(\gamma - t_2) - C_2 - kd_1 - \lambda\alpha_2^{fh} = U_2^{fh} \end{aligned}$$

to get:

$$Q_i^{dnfh}(t_i, .) = \frac{N}{2l}\left[\frac{b(t_j - t_i) + (C_j - C_i) + kl}{2k}\right] \tag{7}$$

and

$$Q_i^{dfh}(t_i, .) = \frac{N}{2l}\left[\frac{b(t_j - t_i) + (C_j - C_i) + kl + \lambda\left(\alpha_j^{fh} - \alpha_i^{fh}\right)}{2k}\right] \tag{8}$$

The hospital's utility function now depends on of two revenues (one for each type of patient), the cost of supplying care for both types of patients, and the cost of the delay for both types of patients. The utility functions of the hospitals are the following:

$$\begin{aligned} W_i &= \alpha^{fh}Q_i^{dfh}(t_i, .) + \alpha^{nfh}Q_i^{dnfh}(t_i, .) - \beta[Q_i^{sfh} + Q_i^{snfh}]^2 \\ &\quad -[w^{fh}(t_i^{fh})^2 + w^{nfh}(t_i^{nfh})^2] \end{aligned} \tag{9}$$

where α^{fh} represents the price paid by the fundholder when acquiring care for her patients and α^{nfh} is the price paid by the Health Authority. As before, α as well as β is supposed to be equal for both hospitals. α^{nfh} is assumed to be fixed because it is regulated by the government (e.g. the government may define α^{nfh} to be equal to the average cost of production). α^{fh} is not fixed chosen by the hospital.

The above cost function reflects the assumption that GP fundholding patients and non-fundholding patients are perfect substitutes. Hence, the number of beds can be exchanged between the two types of patients at the same cost. This may be a realistic assumption as the cost of supplying in-patient care is probably the same for whatever is the type of the patient (whether he is a budget holder patient or not) that enters the hospital given that cream skimming is not being considered here. When considering the "costs" imposed on the hospital managers by the delay, I assume that there could be a different weight given to each type of patient, that is, $w^{fh} \neq w^{nfh}$.[8]

Each hospital maximises its utility with respect to: the price paid by the fundholder, α^{fh}, each type of supply (type of patient), and the delay imposed on each type of patient. It is subject to two constraints: each type of supply must be equal or greater than that type of demand.

The Lagrangian function is

$$\begin{aligned} L_i &= \alpha^{fh} Q_i^{dfh}(t_i,.) + \alpha^{nfh} Q_i^{dnfh}(t_i,.) - \beta \left(Q_i^{sfh} + Q_i^{snfh} \right)^2 \\ &\quad - \left[w^{fh} \left(t_i^{fh} \right)^2 + w^{nfh} \left(t_i^{nfh} \right)^2 \right] \\ &\quad + \mu^{fh} \left[Q_i^{sfh} - Q_i^{dfh}(t_i,.) \right] + \mu^{nfh} \left[Q_i^{snfh} - Q_i^{dnfh}(t_i,.) \right] \end{aligned}$$

[8] Nevertheless, for example when analysing the cost of delay one can consider the square of the sum instead of the sum of the squares. This is just the same as considering just one weight, one w, for both types of delay. Moreover, some variations can be introduced in the cost function (the expression representing the cost of capacity). In here, the sum of the squares is the context considered instead of the square of the sum. The latter would correspond to the situation where there could be different costs in producing elective surgery depending on the type of the patient. However, and although one can play around with the expression the main results do not change significantly.

and the Kuhn-Tucker conditions are (with complementary slackness)

$$\frac{\partial L_i}{\partial \alpha^{fh}} = \alpha^{fh}\frac{\partial Q_i^{dfh}}{\partial \alpha^{fh}} + Q_i^{dfh} - \mu^{fh}\frac{\partial Q_i^{dfh}}{\partial \alpha^{fh}} \leqslant 0, \quad \alpha^{fh} \geqslant 0, \tag{10}$$

$$\frac{\partial L_i}{\partial t_i^{fh}} = \alpha^{fh}\frac{\partial Q_i^{dfh}}{\partial t_i^{fh}} - \mu^{fh}\frac{\partial Q_i^{dfh}}{\partial t_i^{fh}} - 2w^{fh}t_i^{fh} \leqslant 0, \quad t_i^{fh} \geqslant 0, \tag{11}$$

$$\frac{\partial L_i}{\partial t_i^{nfh}} = \alpha^{nfh}\frac{\partial Q_i^{dnfh}}{\partial t_i^{nfh}} - \mu^{nfh}\frac{\partial Q_i^{dnfh}}{\partial t_i^{nfh}} - 2w^{nfh}t_i^{nfh} \leqslant 0, \quad t_i^{nfh} \geqslant 0, \tag{12}$$

$$\frac{\partial L_i}{\partial Q_i^{sfh}} = -2\beta(Q_i^{sfh} + Q_i^{snfh}) + \mu^{fh} \leqslant 0, \quad Q_i^{sfh} \geqslant 0, \tag{13}$$

$$\frac{\partial L_i}{\partial Q_i^{snfh}} = -2\beta(Q_i^{sfh} + Q_i^{snfh}) + \mu^{nfh} \leqslant 0, \quad Q_i^{snfh} \geqslant 0, \tag{14}$$

$$\frac{\partial L_i}{\partial \mu^{fh}} = Q_i^{sfh} - Q_i^{dfh} \geqslant 0, \quad \mu^{fh} \geqslant 0, \tag{15}$$

$$\frac{\partial L_i}{\partial \mu^{nfh}} = Q_i^{snfh} - Q_i^{dnfh} \geqslant 0, \quad \mu^{nfh} \geqslant 0. \tag{16}$$

With respect to conditions (15) and (16) it is assumed that the constraints are binding, or in other words, both μ^{fh} and μ^{nfh} are positive which means that $Q_i^{sfh} = Q_i^{dfh}$, and that $Q_i^{snfh} = Q_i^{dnfh}$. If supply is costly there will not be excess supply.

Looking at conditions (13) and (14) and assuming that $Q_i^{sfh} > 0$ and that $Q_i^{snfh} > 0$ (which will be true as long as the conditions above hold and quantity demanded is positive), one has that $2\beta(Q_i^{sfh} + Q_i^{snfh}) = \mu^{fh}$ and $2\beta(Q_i^{sfh} + Q_i^{snfh}) = \mu^{nfh}$. This means that $\mu^{fh} = \mu^{nfh}$. Moreover, it can be seen that the are equal to marginal cost, $\mu^{fh} = \mu^{nfh} = 2\beta(Q_i^{sfh} + Q_i^{snfh})$.

Looking at condition (10), with a positive price paid by the fundholder, $\alpha^{fh} > 0$, one has that $\frac{\partial L}{\partial \alpha^{fh}} = 0$. This implies that $-\left(\alpha^{fh} - \mu^{fh}\right)\frac{\partial Q_i^{dfh}}{\partial \alpha^{fh}} = Q_i^{dfh}$, which implies that $\alpha^{fh} - \mu^{fh} > 0$ because $\frac{\partial Q_i^{dfh}}{\partial \alpha^{fh}} < 0$ and $Q_i^{dfh} > 0$ (by assumption quantity supplied is positive). Therefore, given that $\mu^{fh} = 2\beta(Q_i^{sfh} + Q_i^{snfh})$, budget holders will be paying a price for health care that exceeds the hospital marginal cost.

Condition (11) also leads to an important result. As $\alpha^{fh} - \mu^{fh} > 0$ one can see that $(\alpha^{fh} - \mu^{fh})\frac{\partial Q_i^{dfh}}{\partial t_i^{fh}} - 2w^{fh}t_i^{fh} < 0$, for $\frac{\partial Q_i^{dfh}}{\partial t_i^{fh}} < 0$ and $2w^{fh}t_i^{fh} \geqslant 0$. Consequently, $t_i^{fh} = 0$. This means, in other words, that if the price fundholders pay for care is positive and exceeds the hospital marginal cost, then the waiting time for their patients is zero. This is an important as well as a strong result. The rationale is the following: at any given supply raising the price, α, and lowering the waiting time, t, leaves production costs unchanged and increases revenue making the hospital better off.

Bearing this in mind one needs just to look at condition (12). Given that $\frac{\partial Q_i^{dnfh}}{\partial t_i^{nfh}} < 0$, that $-2w^{nfh}t_i^{nfh} < 0$, and that $\mu^{fh} = \mu^{nfh}$, one has to compare the expression, $(\alpha^{nfh} - \mu^{nfh})\frac{\partial Q_i^{dnfh}}{\partial t_i^{nfh}} - 2w^{nfh}t_i^{nfh}$ with the one given by the second condition, that is, $(\alpha^{fh} - \mu^{fh})\frac{\partial Q_i^{dfh}}{\partial t_i^{fh}} - 2w^{fh}t_i^{fh}$.

It can be concluded from the analysis of condition 2 that $(\alpha^{fh}-\mu^{fh})\frac{\partial Q_i^{dfh}}{\partial t_i^{fh}} - 2w^{fh}t_i^{fh}$ is positive. Hence, it can be seen that **if** $\alpha^{fh} > \alpha^{nfh}$ and $\alpha^{nfh} < \mu^{fh}$ a situation where $(\alpha^{nfh} - \mu^{nfh})\frac{\partial Q_i^{dnfh}}{\partial t_i^{nfh}} - 2w^{nfh}t_i^{nfh} = 0$, and $t_i^{nfh} > 0$ results. This means that as long as Health Authorities pay a price for health care (for patients with a non-budget holder) which is lower than that paid by the fundholders and lower than the hospital marginal cost then the non-fundholding patients have a positive waiting time. This may be the situation where for example $\alpha^{nfh} = \beta(Q_i^{sfh}+Q_i^{snfh}) = ac < \mu^{fh} = 2\beta(Q_i^{sfh} + Q_i^{snfh}) = mc$ or, in other words, the situation where the price paid by the health authorities equals the hospital average cost of production as defined by hospital regulation.

These results are supported by claims that fundholders are paying a higher price for their patients' health care compared with the price paid by Health Authorities, and their patients seem to be jumping the waiting lists. It is argued suggested fundholders pay a price equal to marginal cost whereas health authorities pay a price equal to average cost.

The Nash-Cournot equilibrium in the main case where α^{nfh} is less than α^{fh} and α^{fh} exceeds the marginal cost is:[9]

$$\begin{aligned}
t_i^{fh*} &= 0,\ t_i^{nfh*} = \frac{\beta N^2 b - \alpha^{nfh} N b}{8lkw^{nfh}},\ \alpha^{fh*} = \beta N + \frac{lk}{\lambda} \\
Q_1^{sfh*} &= \frac{N}{4} = Q_1^{dfh} \text{ and } Q_2^{sfh*} = \frac{N}{4} = Q_2^{dfh} \\
Q_1^{snfh*} &= \frac{N}{4} = Q_1^{dnfh} \text{ and } Q_2^{snfh*} = \frac{N}{4} = Q_2^{dnfh}.
\end{aligned}$$

In this situation the fundholding patients face a zero waiting time whereas the non-fundholding patients still face a positive waiting time.

Proposition 4 *If the price paid for care is positive and quantity supplied is positive, the price paid by the budget holders exceeds the hospital marginal*

[9]Note that for the second order conditions to hold the Hessian must be negative semi-definite which requires that $16N\lambda w^{fh}lk \geqslant N^2b^2$.

costs of production and the waiting time for the fundholding patients is zero. If Health Authorities pay a price for health care which is lower than that paid by the budget holders and lower than the hospital marginal cost, then the non-fundholding patients have a positive waiting time.

4 Patients with heterogeneous benefits from care

In this section I model the demand for hospital care dropping the assumption that all the individuals obtain the same benefit from treatment. I will also be able to allow for some patients not to choose any of the hospitals

I use the concept of random utility by McFadden [17].[10]

Define Y_i^* as

$$Y_i^* = U_i + e = b(\gamma - t_i) - C_i - kd_i + e$$

where Y_i^* is the indirect utility obtained from going to hospital i and e is the error that captures the variation in the benefit from treatment among individuals as well as possible errors of perception. Thus, the use of the random utility function allows for the individuals to vary in the net benefit.

I assume the error is the same for each alternative as it measures the benefit from treatment and as such is a characteristic of the individual and not of the alternative. e follows a uniform distribution within $[e^l, e^h]$, where e^l is the lowest value e can take and e^h is the highest one. It is assumed that the errors are positive. A further assumption that I impose is that the utility for the choice of no care is equal to zero. Hence, our indirect utility for each of the choices becomes

$$\begin{aligned} Y_1^* &= U_1 + e = b(\gamma - t_1) - C_1 - kd_1 + e \\ Y_2^* &= U_2 + e = b(\gamma - t_2) - C_2 - kd_2 + e \\ Y_3^* &= 0. \end{aligned}$$

An individual considers going to hospital i if the utility he obtains from going to hospital i is greater or at least equal to the utility obtained from any of the other two options.

[10] In the context developed by McFadden consumers are assumed to be rational in that they make choices so to maximise their perceived utility subject to constraints on expenditures. However, there are errors present in that maximisation due to imperfect perception and optimization as well as to incapacity to measure exactly all the relevant variables. Thus, utility from any choice is a random function.

Hospital 1 is preferred to hospital 2 if:

$$Y_1^* \geqslant Y_2^* \Leftrightarrow d_1 \leqslant d_{1c} = \frac{b(t_2 - t_1) + C_2 - C_1 + kl}{2k}.$$

Hospital 1 is preferred to no care if

$$b(\gamma - t_1) - C_1 - kd_1 + e \geqslant 0$$
$$or$$
$$e \geqslant e_c = -\left[b\left(\gamma - t_1\right) + C_1 + kd_1\right]$$

The demand for hospital 1 consists of those whose location is such that they prefer hospital 1 to hospital 2 (i.e. have $d_1 \leqslant d_{1c}$) and prefer hospital 1 to no care (i.e. have $e \geqslant e_c$). Hence, the demand function for hospital 1 is:

$$\begin{aligned} Q_1^d &= \int_0^{d_{1c}(t_1,.)} \int_{e_c(t_1,d_1,.)}^{e^h} f(e).\phi.de.dd_1 \\ &= \int_0^{d_{1c}(t_1,.)} \int_{e_c(t_1,d_1,.)}^{e^h} \left(\frac{1}{(e^h - e^l)}\right)\left(\frac{N}{l}\right) de.dd_1 \end{aligned}$$

where $f(e)$ and ϕ are respectively the density distribution of the error and the population density. The population density is just $\phi = \frac{N}{l}$ and by the same token $f(e) = \frac{1}{e^h - e^l}$. Hence,

$$\begin{aligned} Q_1^d &= \frac{N}{l}\left[\frac{e^h + b\left(\gamma - t_1\right) + C_1}{(e^h - e^l)}\right]\left[\frac{b(t_2 - t_1) + C_2 - C_1 + kl}{2k}\right] \\ &\quad + \frac{Nk}{2l(e^h - e^l)}\left[\frac{b(t_2 - t_1) + C_2 - C_1 + kl}{2k}\right]^2 \end{aligned}$$

It is easy to show that demand is decreasing with t:

$$\begin{aligned} \frac{\partial Q_1^d}{\partial t_1} &= -\frac{Nb}{2kl(e^h - e^l)}\left[\left(\frac{b(t_2 - t_1) + C_2 - C_1 + kl}{2k}\right) + \left(\frac{e^h + b\left(\gamma - t_1\right) + C_1}{(e^h - e^l)}\right)\right] \\ &\quad - \frac{Nb}{2l(e^h - e^l)}\left[\frac{b(t_2 - t_1) + C_2 - C_1 + kl}{2k}\right] \end{aligned}$$

In the case where waiting time is positive, Hospital 1 maximisation problem is

$$\max_{t_1} W_1 = \alpha Q_1^d(.) - \beta\left(Q_1^d(.)\right)^2 - w\left(t_1\right)^2$$

and the first order condition is

$$\frac{\partial W_1}{\partial t_1} = \alpha \frac{\partial Q_1^d(.)}{\partial t_1} - 2\beta Q_1^d(.) \frac{\partial Q_1^d(.)}{\partial t_1} - 2wt_1 = 0$$

expression that defines the equilibrium waiting time implicitly where $Q_1^d(.)$ and $\frac{\partial Q_1^d(.)}{\partial t_1}$ are defined above.

Given the complexity of the expressions the value for the equilibrium waiting times is not computed. However, it is known from the previous model that a symmetric Nash equilibrium is obtained when the costs of getting on the list, C, the money prices, α, the production cost parameters, β, and parameters w are equal. Therefore, I try to establish some comparative statics knowing that $t_1 = t_2$, when hospitals have similar characteristics.

The analysis is conducted for hospital 1 but the expressions can be generalised for t_2.

I start by analysing the impact on the equilibrium waiting time of a variation on the money price, α. Thus, one has to differentiate the expression of the first order condition. The following is obtained

$$\frac{dt_1}{d\alpha} = \frac{\frac{\partial Q_1^d(.)}{\partial t_1}}{2\beta \frac{\partial Q_1^d(.)}{\partial t_1} \frac{\partial Q_1^d(.)}{\partial t_1} + (2\beta Q_1^d(.) - \alpha) \frac{\partial^2 Q_1^d(.)}{\partial t_1^2} + 2w} < 0$$

which has a negative sign as $\frac{\partial Q_1^d(.)}{\partial t_1} < 0, \frac{\partial^2 Q_1^d(.)}{\partial t_1^2} > 0$, and $(2\beta Q_1^d(.) - \alpha) > 0$.

Secondly, one can look at the impact on the equilibrium delay due to a variation in the benefit.

$$\frac{dt_1}{db} = \frac{(\alpha - 2\beta Q_1^d(.)) \frac{\partial^2 Q_1^d(.)}{\partial t_1 \partial b} - 2\beta \frac{\partial Q_1^d(.)}{\partial t_1} \frac{\partial Q_1^d(.)}{\partial b}}{2\beta \frac{\partial Q_1^d(.)}{\partial t_1} \frac{\partial Q_1^d(.)}{\partial t_1} + (2\beta Q_1^d(.) - \alpha) \frac{\partial^2 Q_1^d(.)}{\partial t_1^2} + 2w} > 0$$

which has a positive sign as $\frac{\partial Q_1^d(.)}{\partial b} < 0$, $\frac{\partial^2 Q_1^d(.)}{\partial t_1 \partial b} < 0$.

Following the same procedure one has:

$$\frac{dt_1}{dk} < 0, \frac{dt_1}{dl} < 0, \frac{dt_1}{dw} < 0, \frac{dt_1}{d\beta} > 0, \frac{dt}{dN} > 0$$

just as previously. thus, I have shown that one does not loose information by simplifying the problem to having the people varying just in terms of the benefit.

5 Conclusions

In this paper I attempted to model the demand for and the supply of elective care making use of a duopoly model more precisely a modified Hotelling framework. The insights provided by Barzel [2], Lindsay and Feigenbaum [16], Goddard *et al.*[13] and Smith and Martin [30] where waiting time was considered a major determinant of the demand for (non-urgent) hospital care were my starting point. Throughout the analysis I was particularly interested in the situation where the equilibrium waiting time was positive.

As a result of patients and hospitals maximising their utility functions the equilibrium waiting time for each hospital (as well as the capacity supply in terms of the number of patients) could be calculated.

Solving the Hotelling's duopoly model I obtained the Nash equilibrium for the case where hospitals were equal in all the parameters and for the case when they differed. The hospital that imposed a higher money cost, α, on the consumers had a lower waiting time and supplied a larger output. The hospital that put a higher weight, w, on the delay offered a lower waiting time and a higher capacity. A hospital with a higher production cost, β, offered a higher waiting time and a smaller capacity. The hospital for which the cost of getting on the list, C, was higher had a lower waiting time.

The model predicts that equilibrium delay increases when production of care becomes more expensive, when the benefit obtained from treatment increases, when the unit cost of distance as well as the length of the road decreases, when the importance given to the delay by the hospitals decreases, and when there is a population increase. All relationships have the expected sign.

Budget holders may care about the money price that has to be paid to the hospital. The price they paid is assumed to be the result of the bargaining process with the hospitals, whereas the price paid by HAs is assumed fixed. If budget holders pay a price for care that exceeds the hospitals marginal cost of production the waiting time faced by fundholding patients is zero. Non-fundholding patients have a positive waiting time as long as Health Authorities pay a money price lower than that paid by the fundholders and lower than the marginal cost of producing elective surgery.

Finally, patients were allowed to vary in the benefit they can obtain from treatment. This was done by introducing an error term in the utility function that would capture that variation in terms of the benefit. I then computed the demand for each hospital giving the example for hospital 1. An implicit function for the equilibrium waiting time is obtained and the comparative statics computed. The results are similar to the those of the original situation.

The contribution to the research in the area may be that to model demand of and supply of elective care, more precisely consumer choice between hospitals, in an environment where hospitals compete for potential patients in terms of waiting time. Time, money and distance costs are important determinants of the demand for elective surgery and the demand functions are built using the above modified Hotelling framework. On the supply side, hospitals maximise their utility which depends directly on the waiting time for elective surgery and supply is endogenous.

References

[1] Acton, J.P. (1976), *"Demand for Health Care Among The Urban Poor: With Special Emphasis on The Role of Time"*, in The Role of Health Insurance in the Health Services SEctor, ed. R. N. Rosett, New York: National Bureau of Economic Research,

[2] Barzel, Y. (1974), *"A Theory of Rationing by Waiting"*, The Journal of Law and Economics, 17, 73-95

[3] Booton, L.A. (1985), *"Hospital Market Structure and the Return of Nursing Education"*, The Journal of Human Resources, 20, 184-196

[4] Becker, G. (1965), *"A Theory of The Allocation of Time"*, The Economic Journal, 299, 493-517

[5] Calem, P.S. and Rizzo, J.A. (1995), *"Competition and Specialization in The Hospital Industry: An Application of Hotelling's Location Model"*, Southern Economic Journal, 61, 1182-1198

[6] Cheung, S.N.S. (1974), *"A Theory of Price Control"*, The Journal of Law and Economics, 53-71

[7] CRD REPORT 8 (1997), *"Concentration and Choice in the Provision of Hospital Services"*, Report, NHS Centre for Review and Dissemination, University of York

[8] Cohen, M.A. and Lee, H.L. (1985), *"The Determinants of Spatial Distribution of Hospital Utilization in a Region"*, Medical Care, 23, 27-38

[9] Culyer, A.J. and Cullis, J.G. (1976), *"Some Economics of Hospital Waiting Lists in the NHS"*, Journal of Social Policy, 5, 239-264

[10] Dranove, D. and Shanley, M. (1989), *"A Note on The Relational Aspects of Hospital Market Definitions"*, Journal of Health Economics, 8, 473-478

[11] Erickson, G.M. and Finkler, S.A. (1985), *"Determinants of Market Share for a Hospital's Service"*, Medical Care, 23, 1003-1018

[12] Folland, S.T. (1983), *"Predicting Hospital Market Shares"*, Inquiry, 20, 34-44

[13] Goddard, J.A., Malek, M. and Tavakoli, M. (1995), *"An Economic Model of The Market for Hospital Treatment for Non-Urgent Conditions"*, Health Economics, 4, 41-55

[14] Hotelling, H. (1970), *"Stability in Competition"*, in Spatial Economic Theory, eds. Dean, Leahy, and Mckee, The Free Press, Macmillan, New York

[15] Joskow, P. (1983), *"Reimbursement Policy, Cost Containment and Non-Price Competition"*, Journal of Health Economics, 2, 167-174

[16] Lindsay, C.M. and Feigenbaum, B.(1984), *"Rationing by Waiting Lists"*, American Economic Review, 74(3), 404-417

[17] McFadden, D. (1973), *"Condtitional Logit Analysis of Qualitative Choice Behavior"*, in P. Zarembka (ed.), Frontiers in Econometrics, New York Academic

[18] McGuirk, M. A. and Porell, F.W. (1984), *"Spatial Patterns of Hospital Utilization: The Impact of Distance and Time"*, Inquiry, 21, 84-95

[19] Nichols, D., Smolenksy, E. and Tideman, T.N. (1971), *"Discrimination by Waiting Time in Merit Goods"*, American Economic Review, 61, 312-323

[20] Noether, M., (1988), *"Competition Among Hospitals"*, Journal of Health Economics , 7, 259-284

[21] Pauly, M. (1987), *"Nonprofit Firms in Medical Markets"*, American Economic Review, 77, 257-262

[22] Pauly, M. and Redisch, M. (1973), *"The Not-For-Profit Hospital as a Physicians Cooperative"*, American Economic Review, 58, 87-99

[23] Pope, C. (1991), *"Trouble in Store: Some Thoughts on the Management of Waiting Lists"*, Sociology of Health and Illness, 13, 193-212

[24] Porter, R.C. (1977), *"On the Optimal Size of Underpriced Facilities"*, American Economic Review, 67, 753-760

[25] Propper, C. (1994), *"The Disutility of Time Spent on the United Kingdom's National Health Service Waiting Lists"*, The Journal of Human Resources, 30, 677-700

[26] Rice, M.F. (1987), *"Inner-city Hospital Closures/Relocations: Race, Income Status, and Legal Issues"*, Social Science and Medicine, 24, 889-896

[27] Robinson, J.C. and Luft, H.S. (1985), *"The Impact of Hospital Market Structure on the Patient Volume, Average Lenght of Stay, and The Cost of Care"*, Journal of Health Economics, 4, 333-356

[28] Roghmann, K.J. and Zastowny, T. R. (1979), *"Proximity as a Factor in the Selaction of Health Care Providers"*, Social Science and Medicine, 13D, 61-69

[29] Shannon, G.W., Skinner, J.L. and Bashshur,R.L. (1973), *"Time and Distance: The Journey for Medical Care"*, International Journal of Health Services, 3, 237-243

[30] Smith, P. C. and Martin, S. (July 1995), *"Modelling waiting times for elective surgery"*, Discussion Paper, Centre for Health Economics and Department of Economics, University of York

[31] Staten, M. (1987), *"Market Share and The Illusion of Power"*, Journal of Health Economics, 6, 43-58

[32] Studnicki, J. (1975), *"The Minimization of Travel Effort as a Delinating Influence for Urban Hospital Service Areas"*, International Journal of Health Services, 5, 679-693

[33] Weiss, J.E. and Greenlick, M.R. (1970), *"Determinants of Medical Care Utilisation: The effect of Social Class and Distance on Contacts with the Medical Care System"*, Medical Care, 8, 456-462