THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2000/27

Waiting Times and Waiting Lists: A Model of the Market for Elective Surgery

by

Hugh Gravelle, Peter C Smith and Ana Xavier

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

Waiting times and waiting lists: a model of the market for elective surgery

Hugh Gravelle* Peter Smith† Ana Xavier‡§

Abstract

We present a simple dynamic model of the demand and supply for elective surgery in the NHS and test it using a panel of quartely data for 123 English health authorities from the second quarter of 1987 to the first quarter of 1993. We find that supply is increasing in measures of the previous period's waiting time and that demand is decreasing in the previous period waiting time. The results imply that health care systems which are rationed by waiting do respond to indicators of waiting times and waiting lists. More generally, the paper adds to the small but consistent body of research which demonstrates that health care systems respond to the publication of high profile performance data.

Keywords: Waiting times. Rationing. Elective surgery.
JEL Nos: I11, H51

1 Introduction

Large waiting lists and long waits for many procedures, have been a persistent feature of the UK National Health Service since its inception in 1948. The numbers of voters affected by NHS waiting at any time has meant that

*National Primary Care Research and Development Centre, Centre for Health Economics. Email hg8@york.ac.uk

†Corresponding author: Centre for Health Economics, University of York, Heslington, York, YO10 5DD. Email pcs1@york.ac.uk

‡Centre for Health Economics. Email amcx100@york.ac.uk

§Funding from the UK Department of Health to the National Primary Care Research and Development Centre and the Centre for Health Economics is acknowledged. Ana Xavier is also funded by the programme Praxis XXI, Fundação para a Ciência e Tecnologia, Portugal. The views expressed are not necessarily those of the funders.

the topic has always had a high political priority, and the national government has for some years made use of waiting time and waiting list data as important indicators of local performance in the management of the NHS.

It is therefore surprising that the volume of useful research into the topic is relatively modest (Hamblin et al., 1998). The purpose of this paper is to place the use of NHS waiting data as a performance measure in the context of an economic model of demand and supply for hospital care. We start by describing the institutional setting relevant to our model. We present an economic model of waiting lists in section 3. In section 4 we describe the data and estimate the demand and supply functions obtained in section 3. After giving the main empirical results in section 5 we discuss them and propose some issues for future research in the concluding section.

2 Institutional setting

2.1 Demand for elective surgery

Almost all citizens of the United Kingdom are registered with an NHS general practitioner (GP). Apart from those admitted through accident and emergency units, no patient can be seen by an NHS hospital specialist without being referred by a GP. The GP therefore has a key gatekeeping role in relation to secondary care. Demand for elective (non-emergency) surgery is affected by the health of patients and the referral practices of GPs.

Historically, although general practitioners have had no direct incentive to restrain referrals for elective procedures, there has been a general culture of restraint amongst GPs which has helped the NHS keep expenditure within strict public expenditure limits, enabling the United Kingdom to be one of the lowest spenders on health care (measured as a percentage of gross domestic product) amongst developed nations.

The NHS reforms of 1991 gave GPs the option of becoming fundholders, which meant that they held fixed budgets to purchase certain routine procedures from local hospitals. Fundholding GPs had a direct incentive to restrain referrals, as they were able to retain any budgetary surplus for use on other health services. The bulk of the evidence suggests that such restraint did not occur (Le Grand, J., Mays, N. and Mulligan, J.,1998) though there is some evidence that fundholders had fewer patients admitted (Croxson, Propper and Perkins, 2000; Gravelle, Dusheiko and Sutton, 1999). In 1999 the fundholding scheme was replaced by a system of 481 Primary Care Groups consisting of around 20 general practices, which hold a budget for the purchase of almost all health care for their patients (UK Government, 1997).

Waiting times for elective admissions are long: they averaged 101 days in 1990/91 and in 1997/98 the average was 111 days. There is evidence that increases in waiting times leads to reductions in admissions of NHS patients (Martin and Smith, 1999; Gravelle, Dusheiko and Sutton, 1999) with some potential NHS patients being deterred from seeking treatment and others diverted to private treatment.

About 11% of the British population is covered by some form of private medical insurance. Such insurance enables patients to by-pass the NHS waiting list for elective surgery and accounts for the majority of private elective procedures. In addition some 15% of private elective procedures are self-financed (Laing's, 1994). About 17% of all elective procedures in the UK are funded privately (either insurance or self-finance). The length of waits in the NHS, rather than clinical quality of care, is used as one of the most important marketing devices for private medical insurance (Besley, Hall and Preston, 1996).

2.2 The Supply Side

In the NHS elective surgical services are provided by NHS hospitals, known as NHS Trusts. Trusts are public sector bodies with management boards appointed by the Secretary of State for Health. They are financed by the contracts they negotiate with local NHS purchasers who either health authorities or PCG/Ts (or GP fundholders between 1991/2 and 1998/9). Trusts compete for contracts with each other, but are not allowed to retain any financial surpluses. The strategic objectives held by Trusts are far from clear, although retention of business from GPs and HAs and attaining government performance criteria are important considerations.

In the early years of the NHS the size of the waiting lists was the main focus of political concern (Yates, 1987). However, between 1988 and 1997 the focus shifted to waiting times. In 1992 the government published a "Patient's Charter", which included the promise that no NHS patient should have to wait more than two years, since reduced to eighteen months, for elective surgery. Annual performance reports are published which document the performance of individual Trusts against waiting time benchmarks. As a consequence, there is a general perception that Trust managers consider the control of waiting times to be a high priority.

The political importance attached to waits for elective procedures was reflected in the establishment by the government of a "Waiting List Fund" in 1986, which became the "Waiting Time Fund" in 1991. It was directed at areas with particularly large numbers of patients waiting for long times, and from 1992 was used to secure achievement of the Patient's Charter targets. Waiting lists once again became an urgent politic issue in early 1998. The government introduced a number of initiatives, including a £500 million injection of funds specifically aimed at reducing waiting lists, a £32 million "performance fund" with which to reward health authorities that do the most to reduce waiting lists, and a threat of dismissal to non-executive directors of health authorities and Trusts that do not meet waiting list targets.

The hospital surgeons who assess patients, recommend treatment and undertake procedures are employees of the Trusts. About 71% of senior NHS surgeons (consultants) also engage in private practice (Monopolies and Mergers Commission, 1993). Private practice accounts for about 10% of working time for full time NHS consultants, and Yates (1995) estimates that on average a senior NHS surgeon undertakes two operations per week in the private sector. There is some concern that the pursuit of private practice by NHS surgeons gives rise to a conflict of interest (Yates, 1995). For example, given the long waiting times for certain procedures in the NHS, an NHS consultant can advise a patient that treatment would be much quicker if undertaken as a private patient - which that consultant may be able to offer. NHS surgeons therefore may have a perverse incentive to maintain long NHS waiting times in order to make private health care appear more attractive.

A patient's wait for a first consultation in a hospital outpatient department with an NHS surgeon after referral by a GP may be considerable, and is now the subject of increased government scrutiny. The NHS Patient's Charter specifies that no patient should have to wait longer than 3 months for their first appointment. Although important, the wait for a first appointment is not the topic of this study, since there is no routinely available information on waiting times for outpatient appointments. Rather, we focus on the wait for surgery after the patient has been assessed by the surgeon. If the surgeon recommends that surgery is required, the patient will usually be added to the NHS waiting list for the relevant procedure, and it is from this point that the waiting time for surgery is conventionally measured. It is this waiting time which was used as a hospital performance indicator during the period of our study.

A clear understanding of the way in which the market operates is an essential requirement for policy: without it policy initiatives may have small or even perverse effects. As an initial step in producing an empirically tested model of the market we investigate the extent to which decision makers on the demand and supply sides of the market are influenced by waiting times and waiting lists and the way in which their decisions interact to determine waiting outcomes. To this end we formulate a simple model of waiting lists and then test it on a new panel data.

3 A model of waiting lists

3.1 Demand for care

The specification of the demand side is similar to that of Lindsey and Feigenbaum (1984). Patients who develop a non-life threatening condition which can be treated in hospital consult their general practitioner and decide, under the advice of the GP, whether to seek hospital care. Seeking hospital care is in itself costly since the patient must be seen in the outpatient department by the hospital doctor who decides whether to place the patient on the waiting list for elective care. When care is eventually provided, it will yield a benefit to the patient. The longer the time waited the smaller the present value of the procedure. The patient will decide to seek care only if the expected discounted benefit from the procedure exceeds the cost of getting onto the waiting list.[1]

The demand, D_i, for elective care in period i depends on the underlying population morbidity (which affects both the probability of developing the relevant condition and the distribution of benefits from the procedure), the cost of getting on the list and the perceived waiting time:

$$D_i = D(w_i^p, z_i^d) \tag{1}$$

where w_i^p is the waiting time perceived by patients when considering joining the waiting list in period i. The vector z_i^d comprises demand shifters such as the socio-economic and morbidity characteristics of the population. An increase in the waiting time reduces the number of people added to the waiting list in each period ($D_w < 0$) because it reduces the expected benefit from treatment.

The specification of the perceived waiting time depends on what we assume about the information available to patients and GPs and their degree of sophistication. In this initial exploration we assume that expectations are myopic: in period i patients and GPs base their expectation of the time the patient will wait for treatment on what they observed in period $i-1$.

For example, suppose the number of people L_{i-1} waiting for treatment at the end of the previous period and the number admitted S_{i-1} are known. The perception of expected waiting time for patients added to the list in period i might then be the time to clear last period's list:

$$w_i^p = \frac{L_{i-1}}{S_{i-1}} \tag{2}$$

We use this simple assumption below for illustrative purposes but test a number of alternatives in the empirical work.

[1] We ignore the possibility of opting for private care, rather than doing without. Martin and Smith (1998) and Goddard, Malek and Tavakoli (1995) show how characteristics of private treatment can be incorporated into the demand function for NHS treatment. The conclusion that increases in NHS waiting times reduce demand is unaffected.

3.2 Admissions

We assume that decisions are taken by a hospital manager with period i utility function

$$u_i = u(S_i, L_i, w_i^m; z^s), \qquad u_L < 0, u_w < 0 \tag{3}$$

where L_i is the number of people on the waiting list at the end of the period, w_i^m is the manager's perception of the waiting time or list performance measure and z^s are exogenous factors affecting the manager's utility.

We assume that the waiting list evolves as:

$$L_i = L_{i-1} + D_i - S_i \tag{4}$$

and ignore the possibility that patients leave the waiting list because they change their minds about the benefit from treatment, die or leave the area.

The supply of elective care may enter the managerial utility function for a variety of reasons. The evaluation system may reward managers who generate higher "profits" or the manager may desire higher profits because it yields more managerial perquisites. Hence the manager may care about S_i because of its impact on implicit profit. With a well behaved profit function, the marginal utility of supply will be positive at small S_i but will eventually become negative. Alternatively the manager may not care about profit but have to work harder to increase elective throughput, so that the marginal utility of additional output is negative.

The inclusion of the waiting performance measure in the managerial utility function reflects the performance evaluation system in place in the NHS. Managers who run hospitals with smaller waiting lists or shorter waiting times are rewarded, implicitly or explicitly. We consider a number of plausible alternatives specifications of the waiting time performance measure in the empirical analysis and consider the implication of one of them in more detail below in this section.

The perceived waiting measure w_i^m depends on the performance indicators in place and the manager's beliefs about how they are affected by supply decisions. The manager may be concerned about the waiting time measure at the end of period i and believe that this is best forecast as some function of last period waiting time w_{i-1}^m, the numbers waiting, the numbers added to the list in the current period and supply:

$$w_i^m = f(S_i; w_{i-1}^m, L_{i-1}, D_i) \tag{5}$$

The manager would expect w_i^m to be decreasing in current supply and increasing in last period waiting time indicator, last period's list and the number expected to be added to the list in the current period (which would depend on patient information in period $i-1$). One simple specification would be that the manager cares about the time to clear the list at the end of the period given the current rate of admissions: $w_i^m = L_i/S_i = (L_{i-1} + D_i - S_i)/S_i$.

Since the utility of the manager in period i is affected by the list inherited from the previous period, her behaviour in a period will in general depend on her willingness to trade period i utility for future utility. Managers who take account of the future consequences of current supply decisions will choose S_i to maximise

$$u(S_i, L_i, w_i^m; z^s) + \delta_i V(L_i + D_{i+1}, w_i^m, z^s) \tag{6}$$

subject to (4) and (5), where δ_i is the manager's one period ahead discount factor and V is the maximised value of discounted utility at the start of period $i+1$. V depends on the waiting list inherited from the previous period plus the demand in period $i+1$ which is determined by patients' myopic perceptions. V may also depend on the waiting measure w_i^m in the previous period, either because it affects managerial forcasts of w_{i-1}^m or because managers' performance may be judged on the change in the waiting indicator.

The first order condition is

$$u_s - u_L + u_w f_S + \delta_i \left[V_1 \left(\frac{\partial D_{i+1}}{\partial w_{i+1}^p} \frac{\partial w_{i+1}^p}{\partial S_i} - 1 \right) + V_2 f_S \right] = 0. \tag{7}$$

where S_i will affect the maximised value of future utility by reducing the list at the start of period $i+1$ and possibly by altering patients' perceptions of waiting time which is based on what they observe in period i. Optimal supply in period i is

$$S_i^* = S(L_{i-1}, w_{i-1}^m, D_i, z^s, \delta_i) = S^*(L_{i-1}, w_{i-1}^m, w_i^p, z^s, z^d, \delta_i) \tag{8}$$

where $S^*(\cdot)$ is the reduced form supply equation. In general predictions about the effect of L_{i-1} and w_{i-1}^m on current supply are ambiguous.

The extent to which current decisions reflect their future implications for waiting lists and waiting times depends on the manager's one period ahead discount factor δ_i. There is no market in the ownership rights to NHS hospitals so that there is no mechanism by which mangers can capitalise all the future value of their current behaviour. The performance of the hospitals they manage will affect their income only so long as they remain managers. Younger managers who expect to remain with the same provider will be more forward looking but as they age their behaviour (supply)will evolve as they place less weight on the future consequences (future list sizes and waiting times) of their current actions. This would be reflected in the model by a decrease in the one period ahead discount factor δ_i and supply decisions in any period would depend on the age distribution and career mobility of managers.

3.2.1 Comparative statics of supply with myopic managers

To emphasise the generally ambiguous effects of last period waiting times and list on current behaviour we adopt an extreme specification of the consequences of the NHS property rights. We assume for the purposes of comparative static analysis that managers care only about the impact of their current actions on their current utility: $\delta_i = 0$. The first order condition on S_i reduces to

$$\frac{du}{dS_i} = u_s - u_L + u_w f_S = 0 \tag{9}$$

Increase in list size. The qualitative response of current supply to an increase in the number of people waiting at the end of the previous period is given by

$$\begin{aligned}
\operatorname{sgn}\frac{\partial S_i^*}{\partial L_{i-1}} &= \operatorname{sgn}\left[\frac{\partial(du/dS_i)}{\partial L_{i-1}}\right] \\
&= \operatorname{sgn}\left[(u_{SL} - u_{LL} + u_{wL} f_S)\frac{dL_i}{dL_{i-1}}\right. \\
&\qquad \left. + (u_{Sw} - u_{Lw} + u_{ww} f_S)\frac{df}{dL_{i-1}} + u_w f_{SL_{i-1}}\right]
\end{aligned} \tag{10}$$

Even in this simple model an important comparative static property depends on fine details of preferences and perceptions. An assumption of concavity u in S is insufficient to sign (10), and the impact of waiting list size on activity is indeterminate.

Suppose that the perceived waiting time that managers care about is the time to clear the end of period waiting list

$$w_i^m = \frac{L_i}{S_i} = \frac{L_{i-1} + D\left(w_i^p, z_i^d\right) - S_i}{S_i} \tag{11}$$

If we additionally assume that preferences are additively separable in supply, the waiting list and the waiting time, the bracketed term on the right hand side of (10) reduces to

$$-\left[u_{LL} + u_{ww}\left(\frac{L_{i-1} + D_i}{S_i^3}\right) + u_w\frac{1}{S_i^2}\right]\left(1 + D_w\frac{\partial w_i^p}{\partial L_{i-1}}\right) \tag{12}$$

The concavity of u implies that the first bracketed term is negative. The second term, which is the effect of last period's list on the size of the current list, is ambiguous. An increase in last period's list directly increases the current list, holding admissions constant, but it also has an indirect effect via the perceived waiting time of patients and thus the numbers added to the list.

To sign the effect of last period's list on current supply we must make some assumption about patient perceptions and their effect on demand: further restrictions on managerial preferences (for example that u is linear in L_i and w_i^p) are not sufficient. If we assume that the number of people waiting at the end of the previous period has no effect on the demand in the current period then the last term in (12) is positive and increases in L_{i-1} increase current supply.

More plausibly, increases in the number waiting at the end of last period increase the perceived waiting time and reduce the number added to the list. If patients' perceived waiting time is the time to clear the list (2) we can rearrange the second bracketed term in (12) to get

$$\text{sgn}\frac{\partial L_i}{\partial L_{i-1}} = \text{sgn}\left(\frac{L_{i-1}}{D_i} + \eta\right) \tag{13}$$

where η is the elasticity of demand with respect to the time to clear the list.[2] The overall effect on the list depends on the elasticity of demand and the length of the period. The longer the period the more important is the indirect effect of L_{i-1} via the additions compared with its direct effect.

Increase in waiting time. The effect of an increase in a measure w_{i-1} of the waiting time in the previous period depends on

$$\begin{aligned}\text{sgn}\frac{\partial S_i}{\partial w_{i-1}} = \text{sgn}\Bigg[&(u_{SL} - u_{LL} + u_{wL} f_S)\frac{dL_i}{dw_{i-1}} \\ &+ (u_{Sw} - u_{Lw} + u_{ww} f_S)\frac{df}{dw_{i-1}} + u_w f_{Sw_{i-1}}\Bigg]\end{aligned} \tag{14}$$

which is again ambiguous. If we assume that the relevant waiting time for both patients and mangers is time to clear the list ($w^p_{i-1} = L_{i-1}/S_{i-1}, w^m_i = L_i/S_i$), that this is also the waiting time measure which has changed ($w_{i-1} = L_{i-1}/S_{i-1}$) and that managerial preferences are additively separable we have

$$\text{sgn}\frac{\partial S_i}{\partial w^m_{i-1}} = \text{sgn}\left\{-\left[u_{LL} + u_{ww}\left(\frac{L_{i-1} + D_i}{S_i^3}\right) + u_w\frac{1}{S_i^2}\right]D_w\right\} < 0 \tag{15}$$

In this case we get the counter-intuitive result that increases in the previous period's waiting time leads to a lower current supply. The reason is that given our assumptions about the manager's perceived waiting time, the change in the time to clear the list has an effect only via its effect on demand. A longer time to clear the list leads to a lower demand in the current period, making it easier to achieve any particular list or waiting time at the end of the period, and thus reduces the number of admissions which are costly for the manager.

[2] Use $w_{i-1} = L_{i-1}/S_{i-1}$ to get

$$\frac{\partial L_i}{\partial L_{i-1}} = 1 + D_w\frac{1}{S_{i-1}} = 1 + \frac{D_w}{D_i}\frac{L_{i-1}}{S_{i-1}}\frac{D_i}{L_{i-1}} = 1 + \frac{D_w w_{i-1}}{D_i}\frac{D_i}{L_{i-1}}$$

On the other hand if the manager's perceived waiting time is not the time to clear the list then the effect of an increase in the waiting time measure w_{i-1} depends on

$$\operatorname{sgn}\frac{\partial S_i}{\partial w_{i-1}} = \operatorname{sgn}\left[-u_{LL}\frac{\partial D_i}{\partial w_{i-1}} + u_{ww}\frac{\partial w_i^m}{\partial w_{i-1}}\frac{\partial w_i^m}{\partial S_i} + u_w\frac{\partial^2 w_i^m}{\partial S_i \partial w_{i-1}}\right] \tag{16}$$

The first term is again negative, assuming that increases in $w_{i-1}$are associated with increases in the waiting time measure perceived by patients, but the second and third are of ambiguous sign and increases in the waiting time last period may reduce or increase supply in the current period.

3.2.2 Disequilibrium dynamics

We have not imposed a market clearing condition that demand (additions to the waiting list) must equal supply (admissions from the waiting list) in any period. The waiting list acts as a stock which increases when additions exceed admissions and falls when they are less than admissions. The admission function (8), the additions function (1) and assumptions about the determinants of the patients' perceived waiting time w_i^p determine the evolution of the waiting list from (4).

4 Data

To examine empirically the determinants of demand and supply of elective surgery we use a panel of data for 123 geographically defined English District Health Authorities (HAs) for 24 quarters from the second quarter of 1987 to the first quarter of 1993. The data are from the KH7 returns made by HAs and have not previously been used for this type of analysis.

The raw data are at hospital level. However, definitions of hospitals were subject to change over the study period because of mergers and demergers. As a result we use the health authority as the unit of observation, as the physical configuration of hospitals within a health authority is unlikely to change markedly over the study period, notwithstanding merger activity. Thus, the data refer to all providers located within a health authority's boundaries.[3]

[3]In 1991 there were 196 DHAs which were the administrative units for secondary care in

This approach has two further advantages. First, it avoids the obvious difficulties in defining populations if the hospital is used as the unit of analysis. Second, and relatedly, the social economic data are available only in respect of areas defined by residence, rather than use of a particular provider. At the time of the study, HAs were responsible for populations of approximately 300,000.

Table 1 gives variable definitions and summary statistics. The waiting list information consists of the number of people who were waiting at the end of each period and the numbers on the list who had waited from 0 to 3 months, 3 to 6 months, 6 to 9 months, 9 to 12 months, from 12 to 24 months, and 24 months and more.

No routine data are available on the numbers of patients actually added to the waiting list in a quarter. Our demand variable (**additions**) is therefore proxied by the number of patients who had been waiting 3 months or less at the end of the quarter, divided by the health authority population. This variable will understate the numbers actually added to the list in the quarter whenever there are any patients who have been added to the list and treated within the same quarter. In a small number of HAs there were some quarters in which there were no patients who had been waiting for three months or less at the end of a quarter. We estimated the demand function with and without these cases and found that it made very little difference in terms of the estimated coefficients, although it did affect the quality of the specification of the model. The reported results exclude such observations and use 119 HAs.

We have information on those who were admitted in each quarter. The data distinguish between ordinary cases (requiring an inpatient stay) and daycases. We aggregate over day and ordinary elective cases over all specialities, and define the rate of **admissions** (supply) as the number admitted in the quarter divided by the relevant health authority population.

The main features of the data over the period were:

- The number of those waiting (the list) was fairly constant, with a slight

the NHS over this period. DHAs had average populations of around 250000. Elimination of DHAs without complete runs of data or which had substantial boundary changes reduced the sample to 123.

increase in the first half of the period and a slight decrease in the second half.

- For in-patient surgery the number of people waiting was initially around 430,000, decreasing to around 358,000 in the last period. For day case surgery the number of people waiting increased steadily from 91,000 to 185,000.
- The number of people with extremely long waits fell over the period. The percentage of those waiting who had been waiting from 12 to 24 months decreased from 15% to 7% for ordinary cases and from 11% to 3% for day cases. The percentage for those waiting more than two years has decreased from 11% to 0% for ordinary cases, and from 8% to 0% for daycases. The proportion of those waiting less then 6 months increased steadily from 50% to 69% for ordinary cases and from 60% to 80% for daycases.
- The average time waited by those on the list also fell for both day cases and for ordinary cases. The average time that ordinary cases on the list had been waiting was around 8.6 months in the first quarter of 1988 but 5.0 months in the last quarter of 1993. For day cases the average time fell from 6.9 months in the first quarter of 1988 to 3.8 months in the last quarter of 1993.
- The total number of admissions (summing day cases and ordinary cases) went up sharply during the period from 306,000 to 422,000 due mainly to the number of day cases more than doubling. The percentage of total admissions that were daycases increased from 25% up to 44%.

4.1 Explanatory Variables

Waiting measure variables enter both the demand (1) and supply (8) equations. We computed four measures of perceived waiting time as well as a measure of the waiting list (values lagged one period are indicated by **variablename_1**):

- **meanwait:** mean time that people on the list have been waiting

- **timeclear:** time to clear the waiting list defined as the ratio of the number waiting at the end of period divided by admissions in period (L_i/S_i)

- **3monthwait:** proportion of all people on the list who have been waiting more than 3 months

- **yearwait:** the proportion of people on the list who have been waiting more than a year

- **list**: number of people waiting.

The proportion of admissions which are day cases (**daycase**) can be taken as a proxy for quality or cost and convenience for patients and thus affect the demand. It can also be proxy for productivity and affect supply.

We expect demand for elective surgery to depend on the accessibility of both primary and secondary care. We measure access to primary care by a distance weighted ratio of GPs to area population: **accgps** and access to secondary care by a distance weighted ratio of hospital beds to population: **acchos** (Carr-Hill et al., 1994). We expect that easier access to secondary care will increase demand for elective procedures. The effect of easier access to primary care is less obvious. For some conditions easier access will lead to more referrals. For others primary care may be a substitute for secondary care and so easier access to primary care may lead to fewer secondary care admissions (Giuffrida, Gravelle and Roland, 1999).

To allow for differences in other area characteristics which might affect demand we include a number of social economic variables computed from the 1991 Census of Population. These ranged from the proportion of the working age population who were unemployed to the proportion of single parent families. We also include measures of population health (**smr, longill**, **permill**). Details are in Table 1.

4.2 Estimation

Because we assume that patients have myopic expectations about waiting times and do not impose market clearing, we do not need to take account of simultaneous equation problems and can separately estimate the demand (additions) and supply (admissions) functions given by (1) and 8.

We considered a variety of specifications. Our preferred specification for time effects was as yearly and seasonal dummies. We also experimented with 23 separate quarterly dummies and with seasonal dummies plus a linear trend. Alternative functional forms were tried including linear, logarithmic specifications and a Box-Cox transformation. We also included more than one lag of the waiting time variable as well as the square of the waiting time.

We estimated both fixed (FE) and random effects (RE) models. The RE model has the advantage that it provides estimates of the parameters on socio-economic and access variables which vary across areas only. However, it rests on the assumption that there is no correlation between the unobserved area effects and the other regressors. When the Hausman test indicates that this assumption is invalid we used the FE estimator (Greene, 1993; Baltagi, 1996) and then regressed the area fixed effects on the time invariant variables. The Reset test was used to check for general misspecification.

We also introduced interaction terms between the waiting time variable and the yearly dummies in order to see if the coefficient on the waiting time differed over time, but found no systematic significance or trend, so the results are not reported.

5 Results

We do not report all the possible permutations of models, in particular we restrict ourselves to the linear or log specifications with year and seasonal dummies. Alternatives produced similar results or were seriously misspecified. The results are in Tables 2 to 4.

5.1 Demand

We found that of the five waiting measures only the **3monthwait** gave results in the demand equation which were significant and did not appear to be misspecified. Table 2 shows four alternative specifications of the additions equation with this waiting indicator. The RE and FE models have very similar patterns of coefficients, though FE models have more significant coefficients on the time invariant variables which are in the middle block in the table. The RE linear model fails the Hausman test but passes it in the log model. The FE linear model passes the Reset test as well as the RE log model at 1.56%. The qualititative pattern of coefficients is similar across linear and log linear specifications except for the **ethnicity** and **alone75** variables. The Davidson-McKinnon PE tests did not indicate that one specification was preferable to the other.

The rate of additions is significantly negatively related to the lagged value of **3monthwait** in all four versions, with elasticities in our two preferred models of -0.30 (FE linear model) and -0.32 (RE log model). The daycase variable has significant positive coefficients in all four models so that it may indeed be acting as indicator of higher quality or more convenient care. Population health measures are significant only in log models and of conflicting signs in log and linear models. General population measures such mortality or longstanding illness are poor proxies for factors affecting condition specific morbidity. Thus populations at say greater risk of heart disease or cancer are more likely to have lower demand for say cataract procedures for which need is strongly age related. Access measures (**accessgp**, **accesshos**) were not significant in any of the specifications.

There positive coefficients on the year dummies (relative to the first year) across all models is clear evidencce of an upward temporal trend in demand. There is also a noticable seasonal pattern with **autumn** and **winter** have a positive and significant impact on the additions to the list, indicating that more additions are made from October to March.

5.2 Admissions

Estimates of the linear version of the reduced form supply equation $S^*(\cdot)$ are shown in Table 3 for three different waiting measures: **3monthwait**, **yearwait**, and **meanwait**. The Hausman test indicated that the RE model was satisfactory in each case and Reset test suggested that the log forms but not the linear forms were misspecified. Models estimated with **timeclear** and **list** failed the Hausman tests in RE form and the Reset test when FE estimation was used when estimated in linear or log form.

Although supply is not significantly related to **3monthwait**, the coefficients on the other two waiting measures are positive and significant. The implied elasiticities (at the means of the variables) are 0.091 for **yearwait** and 0.15 for **meanwait**.

Admissions are also positively related to **daycase** though it is unclear whether this is because increases in demand last period lead to increases in current supply and **daycase** is a proxy for quality or because it is a proxy for efficiency. The socio-economic variables are presumably also picking up indirect effects on last period demand. Fewer of them are significant than in the case of the demand equations.

Admissions, like additions, show a clear upward trend over the period. There is a marked season pattern though it differs from that for additions since in one specification **summer** is negatively associated with admissions and **autumn** is always negatively associated. Admissions are significantly higher in **winter**. It may be that admissions are lower in summer because of staff holidays. It is also possible to explain the lower rate of admissions in the authumn (October to December) as reflecting the increased number of emergency admissions, due to worse weather, flu epidemics etc, which leaves fewer beds available for elective patients. However the increase in elective admissions in the winter quarter (January to March) is puzzling.

We also experimented with a specification of supply $S(L_{i-1}, w^m_{i-1}, D_i, z^s, \delta_i)$ in which we replaced the demand shifters with the level of demand predicted by the linear FE demand model in table 2. The results are shown in table 4 for the two more successful supply models from table 3. In both cases the demand variable is positively and significantly related to supply indicating that current supply is greater the higher last period demand. The waiting time measures are again positive and significant in this specification.

6 Discussion

The aim of this study has been to investigate the extent to which a simple economic model of the market for elective surgery can explain variations in the numbers of patients added to surgical waiting lists and variations in the numbers admitted for surgery from the lists. We use both the past waiting time and the past numbers waiting (the size of the waiting list) as indicators of performance. The study examines their effect on the demand for surgery, in the form of additions to lists, and on the supply of surgical capacity, in the form of current admissions.

Our results, from a dataset that has not hitherto been used in this way, confirm the important influence of waiting time on both the demand for and the supply of elective surgery. The results are in accordance with the background theory on rationing by waiting lists and waiting times. We have been able to show that additions to the list are influenced (negatively) by waiting time measures in the predicted way and that measures of previous period waiting time and of the numbers waiting for treatment affect (positively) the rate of admissions from the list. Our results are consistent with the results of Martin and Smith (1999) and Besley et al. (1996), which were derived using entirely different methods.

Some important policy implications can be drawn from the results. Over a lengthy period, patterns of patient demand for surgery and responses by NHS providers have been shown to be strongly associated with various measures of waiting. The period investigated in this study includes only the very early stages of the Patient's Charter, which increased the policy importance attached to waiting. Thus we would surmise that - were we able to construct longer time series - we might be able to detect even stronger effects in more recent years. More generally, the results show that a high profile performance measure can have a significant impact on organizational behaviour and patients' demand. Hence, recent policy initiatives to increase the importance attached to performance data in the NHS, in the form of the Performance Assessment Framework, may lead to appreciable changes in system behaviour (Department of Health, 1997).

There are of course a certain number of issues that are left for future research. For example, we have assumed that patients have myopic expectations. An obvious future step would be to investigate the implication of adopting adaptive and rational expectation models of waiting times. Furthermore, the data set we have used has certain limitations. Although there has been major organizational change in the NHS in more recent time periods, it may be possible to extend the length of the time series. In order to enrich the model, it would also be desirable to incorporate additional time-varying explanatory variables, relating to issues such as private surgery and private health insurance. This would enable us to extend the model to include explicit treatment of the important issue of private health care within the model. It is also important to recall that a more complete model of NHS waiting would incorporate the wait for the first appointment with a specialist. For this we must await the development of more satisfactory NHS data systems. Finally, the structure of the data used in this study make it possible in principle to repeat this study for individual specialities.

Notwithstanding these inevitable open questions, we believe that this study adds significantly to the small but consistent body of research which demonstrates that health care systems do respond to the publication of high profile performance data. The daunting challenge for policy makers is to harness this responsiveness to create a system that delivers what is - in some sense - an optimal pattern of health care.

References

Baltagi, B. H. (1996). *Econometric Analysis of Panel Data*, John Wiley & Sons

Besley, T, Hall, J. and Preston, I. (1996). "The demand for private health insurance: do waiting lists matter?", *Institute for Fiscal Studies Working Paper* W96/7, London: IFS.

Carr-Hill, R., G. Hardman, S. Martin, S. Peacock, T. Sheldon and Smith, P. (1994). "A formula for distributing NHS revenues based on small area use of hospital beds", *Occasional Paper, Centre for Health Economics*, University of York.

Croxson, B., Propper, C. and Perkins, A. (2000). "Do doctors respond to financial incentives? UK family doctors and the GP fundholding scheme", *Journal of Public Economics*, forthcoming.

Department of Health (1997). *The new NHS: modern, dependable*, London: HMSO

Giuffrida, A., Gravelle, H., and Roland, M. “Measuring quality of care with routine data: performance indicators should not be confused with health outcomes”. *British Medical Journal*, 319, 10 July 1999, 94-98.

Gravelle, H., Sutton, M. and Dusheiko, M. (1999). “Exploring variations in practice admission rates for cataracts”, November, *mimeo.*

Greene, W. H. (1993). *Econometric Analysis*, Macmillan, 2nd edition.

Goddard, J. A., Malek, M. and Tavakoli, M. (1995). “An economic model of the market for hospital treatment for non-urgent conditions”, 4(1), 41-55.

Hamblin, R., Harrison, A. and Boyle, S. (1998). *Access to Elective Care: Why Waiting Lists Grow*, London: King’s Fund.

Laing’s Review of Private Health Care 1994, London: Laing & Buisson Publications.

Le Grand, J., Mays, N. and Mulligan, J. (eds.) (1998). *Learning from the NHS Internal Market*, Kings Fund, London.

Lindsay, C. M. and Feigenbaum, B. (1984). “Rationing by waiting lists”, *American Economic Review*, 74(3), 404-417.

Martin, S. and Smith, P. C. (1999). “Rationing by waiting lists: an empirical investigation”, *Journal of Public Economics*, 71, 141-164.

Monopolies and Mergers Commission (1993), *Private Medical Services: a Report on Agreements and Practices Relating to Charges for the Supply of Private Medical Services*, London: HMSO.

Yates, J. (1987), *Why are We Waiting?*, Oxford: Oxford University Press.

Yates, J. (1995), *Private Eye, Heart and Hip*, London: Churchill Livingstone

Variable definitions and Descriptive statistics

Variable	Definition	Mean	Std. Dev.	Min	Max
admissions	admissions per 1000 DHA population (per quarter)	13.765	6.910	0.000	66.595
additions	number waiting less than 3 months per 1000 DHA population	7.510	3.416	0.000	29.771
list_1	number waiting per 1000 DHA pop	18.468	9.686	0.000	78.682
3monthwait_1	proportion waiting more than 3 months	0.577	0.122	0.000	0.908
yearwait_1	proportion waiting more than a year	0.171	0.105	0.000	0.597
meanwait_1	mean time (months) waited by those still waiting	6.714	2.156	1.500	15.221
timeclear_1	time (quarters) to clear the list: list divided by admissions	1.508	1.563	0.000	42.325
daycases_1	proportion of admissions that were daycases	0.351	0.134	0.000	1.000
accessgp	GPs per person (distance weighted)	0.523	0.081	0.360	0.760
accesshos	hospital beds per person (distance weighted)	2.388	0.697	1.210	4.100
smr	standardised mortality ratio	101.155	17.977	66.510	180.620
longill	proportion with long term illness	0.128	0.023	0.080	0.190
permsick	proportion of those of working age permanently sick	0.038	0.015	0.020	0.080
ethnicity	proportion from ethnic minorities	4.970	6.500	0.440	34.940
oneparent	proportion of one parent families	13.365	4.368	8.410	29.900
alone75	proportion over 75 living alone	3.092	0.582	2.210	5.520
crowded	proportion living in crowed accommodation	0.044	0.025	0.020	0.160
pensioner	proportion of pensionable age living alone	0.337	0.031	0.290	0.470
familysize	average family size	2.477	0.116	1.992	2.659
density	population density	0.063	0.028	0.010	0.200
rent	proportion living in rented accommodation	0.057	0.033	0.020	0.240
student	proportion of 17 year olds who are students	0.407	0.072	0.291	0.668
dependant	proportion of dependants with no carer	0.148	0.031	0.084	0.219
manual	proportion economically active in manual social class	0.463	0.091	0.210	0.670
unemployed	proportion of the economically active unemployed	0.091	0.035	0.050	0.220
age65	proportion of the population over 65	0.158	0.026	0.120	0.250
female	proportion population who are female	0.511	0.007	0.490	0.540

Note 1: Rates are per 1000 inhabitants.
Note 2: The variables above are described in table 1
Note 3: 2829 quarterly observations (123 DHAs and 23 quarters)
Note 4: _1 indicates that the variable is lagged one quarter

DEMAND - ADDITIONS

Dependent variable: the rate of additions per thousand inhabitants

	linear Fixed Effects	linear Random Effects	log Fixed Effects	log Random Effects
additions	Coef.	Coef.	Coef.	Coef.
3monthwait_1	-4.1173**	-3.9095**	-0.2292**	-0.207**
daycases_1	1.7420**	1.7377**	0.0562**	0.0574**
smr	-0.0109	-0.0097	-0.3502**	-0.3141**
longill	1.6784	2.6422	1.069**	0.8513
permsick	5.7642	20.9015	-0.306**	-0.3012
accessgp	-1.8242	-4.2246	-0.0867	-0.2226
accesshos	-0.1083	0.1006	-0.2056**	-0.2032
ethnicity	-0.0970**	-0.0838	0.0745**	0.0609
oneparent	0.4437**	0.4129*	0.5494**	0.6948
alone75	0.8978*	1.6405	-0.3868**	0.2911
crowded	-11.0472	-3.8160	-0.1718**	-0.1952
pensioner	-0.0438	-13.4402	0.4153	-0.1277
familysize	8.9445**	0.6519	-2.683**	-2.1735
density	26.5173**	26.0904	0.2022**	0.2081*
rent	40.3483**	29.4589	0.0829**	0.0626
student	-13.9717**	-12.9457	-0.2429**	-0.2905
dependant	-12.8618	-24.1555	-0.3798*	-0.2296
manual	-8.0268**	-6.0366	0.1683	0.0961
unemployed	20.8143**	18.8168	0.1875**	0.2333
age65	8.0308	-8.2291	-0.18	-0.6819
female	-48.6279**	-96.2108	-4.0659**	-6.052
year2	0.2650*	0.2670**	0.0516**	0.0519**
year3	0.8987**	0.9036**	0.1442**	0.1449**
year4	1.1793**	1.1866**	0.1912**	0.1921
year5	1.5945**	1.6062**	0.2529**	0.2546
year6	1.5431**	1.5647**	0.2196**	0.2228
summer	0.1085	0.1069	0.0108	0.0105
autum	0.5549**	0.5543**	0.074**	0.0738
winter	0.6112**	0.6136**	0.0794**	0.0798
_cons	9.1204**	59.7423	-3.626**	-8.5546
Hausman Test: Prob > Chi2	Chi2 (11) = 25.37 =0.0080 FE			Chi2 (11) = 18.60 =0.0686 RE
RESET Test: Prob > F Prob > Chi2	(^)F(2,2605) = 2.34 =0.0969 passes			Chi2 (3) = 50.55 =0.0159 passes

(^) - fourth moment droppped *5% **1% significance level

SUPPLY - ADMISSIONS

Dependent variable: the rate of admissions per thousand inhab itants

	Random Effects	Random Effects	Random Effects
admissions	Coef.	Coef.	Coef.
3monthwait_1	0.8984		
yearwait_1		4.0016**	
meanwait_1			0.1695**
daycases_1	11.8250**	11.9229**	11.9240**
smr	0.0036	0.0048	0.0047
longill	75.1534	81.5524	79.9083
permsick	-48.1666	-49.9708	-49.5099
accessgp	5.1891	6.0998	5.9909
accesshos	0.3280	0.2913	0.3152
ethnicity	-0.2604	-0.2478	-0.2521
oneparent	0.8100*	0.8124*	0.8135*
alone75	4.7686	4.7660	4.7762
crowded	-18.4128	-24.8915	-23.5472
pensioner	-48.9928	-46.4830	-46.4547
familysize	-4.4411	-3.6304	-3.4402
density	5.7408	6.4039	6.3423
rent	33.2038	32.6077	33.0812
student	-13.4716	-11.8916	-12.1089
dependant	-26.7457	-28.6058	-28.1085
manual	-8.6071	-7.4637	-7.6528
unemployed	21.7483	19.9865	20.1620
age65	-79.8040	-81.4089	-80.5873
female	-240.2913*	-236.2881*	-237.1502*
year2	0.8533**	0.8661**	0.8724**
year3	1.1982**	1.2039**	1.2171**
year4	0.9986**	1.0772**	1.0720**
year5	1.1859**	1.3986**	1.3589**
year6	-0.2532	0.1941	0.0973
summer	-0.3727*	-0.3653	-0.3691
autum	-0.4746*	-0.4561*	-0.4615*
winter	0.6110**	0.6604**	0.6488**
_cons	145.5290	138.7225	138.3113
Hausman Test: Prob > Chi2	Chi2 (11) = 3.55 =0.9812 RE	Chi2 (11) = 3.05 =0.9901 RE	Chi2 (11) = 3.49 =0.9826 RE
RESET Test: Prob > Chi2	Chi2 (3) = 3.87 =0.2757 passes	Chi2 (3) = 3.02 =0.3888 passes	Chi2 (3) = 3.06 =0.3817 passes

*5% **1% significance level

<u>SUPPLY - ADMISSIONS</u>

Dependent variable: the rate of admissions per thousand inhabitants
Using predicted demand

	Random Effects	Random Effects
admissions	Coef.	Coef.
predicted D	1.6102**	2.2013**
yearwait_1	6.3536**	
meanwait_1		0.4396**
daycases_1	3.5661**	2.5694*
smr	0.0217	0.0278*
longill	74.4121	75.3465
permsick	-43.7033	-43.9655
ethnicity	-0.2065	-0.203
oneparent	0.7196	0.7159
alone75	3.9681	3.9489
crowded	-19.0383	-20.1602
pensioner	-43.5007	-42.8967
familysize	-7.3607	-7.2686
density	8.0551	8.1687
rent	29.3605	29.8579
student	-13.1029	-13.1071
dependant	-24.4338	-24.9898
manual	-10.9207	-10.8499
unemployed	34.0586	34.2539
age65	-76.3127	-75.974
female	-207.8017	-206.7873
year2	0.603**	0.4612
year3	0.0161	-0.5053
year4	-0.462	-1.1416*
year5	-0.5703	-1.4823*
year6	-1.2901*	-2.1535**
summer	-0.4238*	-0.4822**
autum	-1.1933**	-1.5123**
winter	-0.2986	-0.6566*
_cons	129.9595	123.1322
Hausman Test: Prob > Chi2	Chi2 (12) = 2.81 =0.9968 RE	Chi2 (12) = 3.09 =0.9948 RE
RESET Test: Prob > Chi2	Chi2 (3) = 3.77 =0.2871 passes	Chi2 (3) = 4.07 =0.2540 passes

*5% **1% significance level