Journal of Health Economics 28 (2009) 771–780

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Do waiting times reduce hospital costs?

Luigi Siciliani[a,b], Anderson Stanciole[c,*,1], Rowena Jacobs[d]

[a] Department of Economics and Related Studies, and Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK
[b] C.E.P.R., 90-98 Goswell Street, London EC1V 7DB, UK
[c] Department of Health Systems Financing, World Health Organization, Avenue Appia, 20, Geneva 1211, Switzerland
[d] Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 18 December 2007
Received in revised form 2 April 2009
Accepted 3 April 2009
Available online 10 April 2009

JEL classification:
I11
I18

Keywords:
Waiting times
Costs
Hospitals

ABSTRACT

Using a sample of 137 hospitals over the period 1998–2002 in the English National Health Service, we estimate the elasticity of hospital costs with respect to waiting times. Our cross-sectional and panel-data results suggest that at the sample mean (103 days), waiting times have no significant effect on hospitals' costs or, at most, a positive one. If significant, the elasticity of cost with respect to waiting time from our cross-sectional estimates is in the range 0.4–1. The elasticity is still positive but lower in our fixed-effects specifications (0.2–0.4). In all specifications, the effect of waiting time on cost is non-linear, suggesting a U-shaped relationship between hospital costs and waiting times. However, the level of waiting time which minimises total costs is always below ten days.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Waiting times are a major health policy issue in many OECD countries. Average waiting times range between four and eight months for common procedures like cataract and hip replacement. Waiting times act as a rationing mechanism that help to bring into equilibrium the demand for and the supply of health care (Lindsay and Feigenbaum, 1984; Martin and Smith, 1999; Cullis et al., 2000): in the absence of price rationing and if benefit from treatment is to some extent unobservable, waiting times may deter patients with small benefit from asking for treatment.

On the benefit side, waiting times postpone and therefore reduce patients' benefits. Moreover, they may deteriorate the health status of the patient, prolong suffering, and generate loss of utility and uncertainty. On the cost side, the relationship between waiting times and providers' costs is not necessarily monotonic. It has been argued that for low waiting times, a marginal increase in waiting times may reduce the cost of provision of elective surgery. When demand is stochastic, waiting times may reduce idle capacity, therefore inducing a more efficient use of resources (Iversen, 1993, 1997; Barros and Olivella, 2005). Hospital costs reduce with waiting times as a consequence of the lower excess capacity. However, as suggested by Iversen (1993), there might be a point over which higher waiting times increase costs, which may be due to the higher costs of managing the waiting list. For example when waiting times are very long, there might be an increase in the resources needed for repeated examinations of patients (since their status might change during the course of the waiting), an increase in treatment costs and in length of stay (if severity deteriorates while waiting), and an increase in cancellation rates. There is therefore, at least theoretically, a level of waiting time which minimises total costs. Above this level, higher waiting times increase hospital costs.

The purpose of this paper is to empirically estimate the elasticity of hospital costs with respect to waiting times. We use a sample of 137 acute hospitals over the period 1998–2002 in the English National Health Service (NHS). Our cross-sectional and panel-data results suggest that at the sample mean (103 days), waiting times have no significant effect on hospitals' costs or, at most, a positive one. If significant, the elasticity of cost with respect to waiting time in our cross-sectional estimates is in the range 0.4–1. The elasticity is still positive but lower in our fixed-effects specifications (0.2–0.4). In all specifications the effect of waiting time on cost is

* Corresponding author at: Department of Health Systems Financing, World Health Organization, Avenue Appia, 20, Geneva 1211, Switzerland.
Tel.: +41 22 791 1489.
E-mail addresses: ls24@york.ac.uk (L. Siciliani), stanciolea@who.int (A. Stanciole), rj3@york.ac.uk (R. Jacobs).
[1] Work undertaken while at the University of York.

0167-6296/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2009.04.002

non-linear, suggesting potentially a U-shaped relationship between hospital costs and waiting times, which is consistent with Iversen's (1993) model.

Our results therefore suggest that the level of waiting times observed in our sample is above the one which minimises total costs. If healthcare providers could ration the demand by dumping or neglecting treatment to patients with low expected benefit (explicit rationing), we should not observe providers with waiting times held above the cost-minimising level. However, since waiting times also have a rationing role, then waiting times might as well be above the cost-minimising level (Gravelle and Siciliani, 2008, Section 3.4). If the waiting time which brings in equilibrium the demand for and supply of health care is sufficiently high, waiting times generate a negative spillover effect on production costs.

There might be several reasons why explicit rationing might not be feasible for the providers. First, the patients' benefit might be at least to some extent unobservable to the provider. Second, even if benefit is perfectly observable, patients with low expected benefit might feel entitled to treatment in the NHS: clinicians might therefore prefer to add patients on the waiting list, rather than taking responsibility for explicitly declining treatment to patients. Finally, there might be political constraints which prevent the adoption of explicit rationing.

1.1. Related literature

Our study relates to the literature which estimates cost functions in the hospital sector. One main concern of this extensive literature has been to estimate the degree of scale and scope economies in the hospital sector (Cowing and Holtmann, 1983; Vitaliano, 1987; Breyer, 1987; Vita, 1990; Grannemann et al., 1986). An important distinction in this literature is whether a short-run or a long-run cost function is estimated. If all inputs are variable, then a long-run cost function can be estimated. However, some capital inputs, like the number of beds, are typically fixed in the short run. The inclusion of beds as an additional regressor, if positive and statistically significant, usually confirms that some inputs are fixed in the short run. A survey by Aletras et al. (1997) suggests that economies of scale are likely to appear below 100 beds, while diseconomies of scale appear above 300 beds.[2]

The above literature assumes that demand is deterministic. Joskow (1980) argues that at least to some extent the demand for health care is uncertain (because of emergency treatments). Hospitals need therefore to provide standby capacity in order to minimise the risk that demand exceeds supply. The optimal amount of standby capacity is calculated such that the probability of excess demand is equal to a certain target level. High probability of excess demand implies poorer quality of services for the patients as some patients risk to be turned down. The empirical evidence suggests that uncertain demand matters and increases costs (Joskow, 1980; Friedman and Pauly, 1981; Mulligan, 1985; Joskow, 1985; Gaynor and Anderson, 1995; Keeler and Ying, 1996; McGuire and Hughes, 2003). In other words, higher uncertainty, measured through higher unexpected demand, leads to higher hospital costs (controlling for actual activity). Unexpected demand can be calculated as the difference between actual demand and expected demand, where demand is specified as an auto-regressive process (Keeler and Ying, 1996; McGuire and Hughes, 2003).

Smet (2004) suggests that an alternative measure of unexpected demand, or standby capacity, is an index based on expected waiting times. The author argues, similarly to Iversen (1997), that higher waiting times imply lower standby capacity and should therefore imply lower costs. He finds that such proxy of waiting time is negatively correlated with costs.

We depart from Smet (2004) by using actual waiting time, recorded from administrative data, rather than using a proxy of waiting times based on arrival rates. Moreover, the above literature refers either to hospitals in the US (Joskow, 1980; Friedman and Pauly, 1981; Gaynor and Anderson, 1995; Keeler and Ying, 1996) or to hospitals from Belgium (Smet, 2004), where waiting times are known to be low (Siciliani and Hurst, 2005). The study by McGuire and Hughes (2003) includes hospitals from the UK but they do not focus on waiting times.

In the UK, waiting times have been accurately recorded at patient level since the 1980s. Differently from the above-mentioned literature, we test the effect of waiting times on costs. Moreover, in the UK waiting times are significantly larger than in the US and Belgium. Therefore, higher waiting times may not imply lower standby capacity. Indeed as argued by Iversen (1997) there is a point where higher waiting times may increase costs, rather than reducing them. Our empirical results support this latter scenario.

The study is organised as follows. Section 2 presents the methods. Section 3 describes the institutional setting. Section 4 describes the data. Section 5 presents the results. Section 6 concludes.

2. Methods

Define C as the total cost of a representative hospital, w as the waiting time of the patients admitted for treatment, and y as the number of patients treated. Following Iversen (1993, 1997), the cost function of a hospital can be represented by

$$C = C(w, y) \tag{1}$$

with $C_y > 0$, higher activity increases costs; $C_w < 0$ if $w < \tilde{w}$, $C_w = 0$ if $w = \tilde{w}$ and $C_w > 0$ if $w > \tilde{w}$. The relationship between waiting times and costs is U-shaped: waiting times reduce costs for low levels of waiting times, while waiting times increase costs for high levels of waiting times. Iversen (1993, 1997) argues that for low waiting times, higher waiting times reduce hospital costs, as a consequence of lower excess capacity. If the demand for health care is stochastic, higher waiting times reduce the probability that the system has idle capacity and therefore reduce costs (for a formal model with a stochastic demand function and the effect of waiting times on idle capacity, see also Goddard et al. (1995); Olivella (2003) also assumes that waiting times reduce costs because waiting times allow for a more efficient use of hospital equipment).

Iversen (1993) suggests that there is a level of waiting times over which higher waiting times increase costs. For high waiting times, the reductions in costs from a marginal increase in waiting, in terms of lower probability of idle capacity, become negligible. In contrast, for high waiting times, a marginal increase in waiting may increase the costs of managing the waiting list. For example, providers might need to spend more resources in repeated examination of patients. This might be the case if the health status of the patients deteriorates in the course of the waiting, and new examinations are needed before the treatment is provided. Treatment costs might also increase due to higher cancellations rates. The longer the wait, the higher the chance that in the meanwhile patients scheduled for treatment have found treatment somewhere else, or have considered alternative treatments. Moreover, the provider will have to spend more resources to keep the register of the patients on the list up to date. Therefore overall prioritisation costs will be higher when waiting times are higher.

[2] A more recent strand of this literature focuses on hospital cost frontiers and estimates efficiency scores for each hospital in the sample under consideration (see Jacobs et al., 2006, for a review).

There is then, at least theoretically, a level of waiting time which minimises total costs. The purpose of this paper is to estimate empirically the relationship between hospital costs and waiting times. We estimate three types of regressions: pooled OLS, panel fixed effects and panel random effects. The pooled OLS model is given by:

$$C_{it} = \alpha + \gamma_1 w_{it} + \gamma_2 (w_{it})^2 + \mathbf{y}'_{\mathbf{it}}\boldsymbol{\beta}_1 + \mathbf{x}'_{\mathbf{it}}\boldsymbol{\beta}_2 + \mathbf{d}'_{\mathbf{t}}\boldsymbol{\beta}_3 + u_{it}, \quad (2)$$

where C_{it} is the cost of hospital i at year t, w_{it} is waiting time, $\mathbf{y_{it}}$ is a vector of outputs, $\mathbf{x_{it}}$ is a vector of control variables, $\mathbf{d_t}$ is a vector of time dummies, and u_{it} is the idiosyncratic error. According to the theoretical literature discussed above, we should expect $\gamma_1 < 0$ and $\gamma_2 > 0$.

An alternative approach is to assume that individual effects are specific to each observation. This leads to the fixed-effects model:

$$C_{it} = \alpha_i + \gamma_1 w_{it} + \gamma_2 (w_{it})^2 + \mathbf{y}'_{\mathbf{it}}\boldsymbol{\beta}_1 + \mathbf{x}'_{\mathbf{it}}\boldsymbol{\beta}_2 + \mathbf{d}'_{\mathbf{t}}\boldsymbol{\beta}_3 + u_{it} \quad (3)$$

The hospital-specific fixed-effects α_i capture individual unobserved heterogeneity. An alternative to the fixed-effects model is the random-effects model, where $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$ and $u_{it} \sim N(0, \sigma_\varepsilon^2)$. In this formulation the individual effects are randomly *iid* distributed over the population of hospitals. Fixed-effects and random-effects models can be compared by the Hausman test, which tests for systematic differences in coefficients between the two models (Cameron and Trivedi, 2005).

It might be argued that the relationship between costs and waiting times is endogenous. If a hospital has high costs, it is more likely to have longer waiting times. There are several channels through which this may happen.

First, more inefficient hospitals have higher costs (due for example to poor management): if higher inefficiency also implies higher inefficiency in the management of the waiting list, then inefficient hospitals may have both higher costs and higher waiting times (a positive correlation). If the researcher has no access to variables correlated with inefficiency, then the OLS estimates of Eq. (2) will be biased upwards. We use at least two control variables that might be correlated with inefficiency: length of stay and proportion of day cases.[3] Keeping other factors constant, more inefficient providers have a higher length of stay and a smaller proportion of day cases.

Second, hospitals with higher quality might have a higher cost and at the same time attract a higher number of patients (Pope, 1989; Ma, 1994; Chalkley and Malcomson, 1998), which leads to a higher waiting time (again, a positive correlation). We use at least two control variables that might be correlated with quality: length of stay and (age and gender adjusted) readmission rates.[4] Keeping other factors constant, providers with higher quality should have a higher length of stay and lower readmission rates.

We use length of stay as a variable potentially being correlated with inefficiency or quality. However, it may also be argued that length of stay is endogenous and (partly) influenced by the management of the hospital. A hospital can to a certain extent control length of stay to maximise hospital revenues or it can adjust it in order to optimally use capacity. Regardless of the exact interpretation given, it is important to control for length of stay as it is likely to be positively correlated with costs.

If there is some residual unobserved efficiency and quality, the OLS might still be biased. However, by estimating a fixed-effects model, all unobserved inefficiency and unobserved quality will be captured by the individual fixed effects, as long as quality and inefficiency are time invariant, which seems plausible over short intervals of time.

In Section 5 we provide estimates for both the OLS and fixed-effects specifications. OLS works well if there are no omitted variables. Given that we have an extensive number of control variables, our estimates will be unbiased. The fixed-effects model works well only if there is sufficient variation in waiting times and costs over the short period of time. Since we have only four years and waiting times do not vary to a great extent over time (see within and between sample standard deviations in the data section), fixed effects are likely to provide inefficient estimates (i.e. to suggest an insignificant effect even if there is one).

The random-effects model is more efficient compared to the fixed-effects model. Ideally we would prefer to estimate the random-effects model rather than the fixed effects. The gains in efficiency would be particularly helpful in an application where there is little variation over time, and the time dimension and i dimensions are not large ($i = 137$ and $t = 4$). However, the random-effects model relies on the assumption of no correlation between unobservables and the regressors. This seems unlikely in our model, as we can think unobservable quality or efficiency dimensions to be correlated for example with length of stay, activity, waiting times, and quality. Therefore, there is a real trade-off between using the fixed-effects versus random-effects, i.e. unbiased but inefficient estimates versus biased and efficient estimates.

3. Institutional setting

Most health care in the UK is supplied by the National Health Service (NHS) and is free at the point of consumption. Family doctors (general practitioners) act as gatekeepers to secondary care. Therefore, a patient needs to visit a family doctor to access elective hospital treatment. Hospitals are publicly owned and operate in a pseudo-market. They have to compete for contracts from public purchasers. Moreover, they are subject to a stringent monitoring based on several performance indicators, such as waiting times, expenditure control and patient safety (Smith, 2002; Le Grand, 2002).

While patients can choose treatment in the private sector, the ability to choose between alternative NHS providers was initially constrained. However, in 1991 an internal market was introduced and the purchasers (Health Authorities) and providers (NHS hospitals) were split. The majority of hospital care was commissioned by Health Authorities (HAs). Each HA was responsible for a geographically defined market. HAs negotiated block contracts with NHS hospitals.

As part of the internal market reforms, larger general practices could volunteer to become fundholders and be allocated an annual budget by their local HA to purchase some elective procedures. By the time fundholding was abolished in April 1999, about 50%

[3] Several authors have used length of stay as a proxy of efficiency (see for example Martin and Smith, 1996; Fenn and Davies, 1990; Vita, 1990). The idea is that providers will be able to treat a higher number of patients if they reduce the length of stay of each patient, for a given number of beds (or capacity). Hauck and Street (2007), Martin et al. (2007), and Siciliani and Martin (2007) use the proportion of day cases as a proxy of efficiency.

[4] Several studies testing the effect of variations in hospital prices on quality use length of stay as a proxy of quality or the intensity of services received by the patients (see for example Dafny, 2005; Gilman, 2000; Ellis and McGuire, 1996). The idea is that a longer length of stay implies a higher intensity of care which ultimately implies a higher quality. In other words, length of stay is a proxy for the amount of care received by the patient during the stay in the hospital. A short length of stay might imply an early discharge with the patient left on its own in case of complications. Notice that we use length of stay as a proxy of both quality and inefficiency as higher length of stay might reflect both. For the purposes of this study, it does not matter whether variations in length of stay reflects variations in inefficiency or quality. The important point is that both inefficiency and quality increase costs, and these might be captured by a higher length of stay. Readmission rates as a proxy to quality have been utilised by Evans and Beomsoo (2006), Ho and Hamilton (2000) and Arocena and García-Prado (2007).

of practices had volunteered to join the scheme (Dusheiko et al., 2006). In April 1999 GP fundholding was abolished and 481 primary care groups (PCGs) were established. Initially some PCGs merely supported their local Health Authority in its commissioning role while others took responsibility for a proportion of the commissioning budget. Gradually, PCGs took responsibility for a larger proportion of the commissioning budget so that, by April 2002, all 481 PCGs had converted to 303 primary care trusts (PCTs). These were now the principal purchasers of secondary care. With the loss of their commissioning role, the Health authorities were merged into 28 Strategic Health Authorities (SHAs) with the role of monitoring the performance of PCTs and NHS Trusts.

Our study period covers the four-year period 1998/1999–2001/2002. HAs/PCTs had contracts with one or more hospital Trusts. Dusheiko et al. (2006) report that, in 1999/2000, two-thirds of all inpatient admissions went to the HA's/PCT's main provider but that on average each HA/PCT employed 14 providers: less than five providers accounted for most admissions with the remaining admissions being handled by a much larger number of providers. Throughout the study period (1998/1999–2001/2002), patients were usually treated in hospitals with whom their HA/PCT had a contract although it was possible for patients to be referred to and treated at other hospitals.

Before April 1999 these extra contractual referrals (ECRs) were paid for by the patient's HA/PCT at a price to be agreed with the provider (such referrals constituted a small proportion of all admissions). In April 1999 ECRs were replaced by other forms of commissioning in order to reduce transaction costs (not least the need to agree a price for each treatment, raise invoices and monitor payments). HAs and NHS Trusts were required to negotiate contracts (Service Agreements) where there was a predictable flow of work and, for the remaining treatments, out of area treatment (OAT) arrangements were introduced.

3.1. Competition, costs and waiting times

When assessing the potential effect of waiting times on costs, it may be important to control for the degree of competition among public hospitals, as competition may be correlated with both waiting times and costs. Omitting competition may potentially generate biased estimates.

The theoretical literature on the effect of competition (or choice) on waiting times is mixed. Xavier (2003) and Siciliani (2005) show that more competition always increases waiting time. Brekke et al. (2008) show that more competition increases waiting times only if the proportion of high-benefit patients is sufficiently large. Dawson et al. (2007) suggest that competition reduces waiting times if the demand is inelastic. Propper et al. (2008) and Dranove and Satterthwaite (2000) argue that in more competitive environments providers may reduce the quality dimensions which are more difficult to observe, like certain dimensions of clinical quality, while they may increase the quality dimensions which are more easily observable (like waiting times). The empirical evidence suggests that more competition is likely to reduce waiting times (Siciliani and Martin, 2007; Dawson et al., 2007; Propper et al., 2008). The theoretical literature also suggests that more competition should be associated with higher cost-containment effort and therefore lower costs (Pope, 1989), as confirmed by some empirical evidence (Zwanziger and Melnick, 1988).

The main implication for our empirical work is to control for the degree of competition. As we will show below, we find that competition has no effect on hospitals' costs. This may be due to the fact that hospitals are already working at reasonably high levels of efficiency, so the effect of competition is not significant.

A related issue is the one of cream-skimming and skimping (Ellis, 1998). If hospitals in more competitive areas have a stronger incentive to cream the low-severity patients and skimp the high severity patients, then hospitals in more competitive areas might have lower costs and lower case-mix. If the lower case-mix is associated with higher waiting times, because for example of waiting-time prioritisation (Gravelle and Siciliani, 2008), then biased estimates might arise. In our analysis below we control for the degree of competition, and for the case-mix of the patients as captured by the average HRG weight of the hospital.

4. Data

The sample comprises 137 English NHS *acute* hospitals observed annually between 1998/1999 and 2001/2002, making an unbalanced panel of 440 observations. To reduce the degree of heterogeneity across hospitals, we have therefore excluded from our sample teaching hospitals, multi-service hospitals and hospitals specialising in orthopaedics and ophthalmology.

The data were collected from several sources, including the Hospital Episodes Statistics (HES), the Department of Health (DoH), the National Health Service Information Authority (NHSIA) and Dr Foster, the independent organization that provides information on the quality of health services.

Our dependent variable is "total hospital cost", measured in thousands of Pounds Sterling. It was compiled from the Department of Health and was transformed into real values for 2002 using the GDP deflator provided by HM Treasury. Ideally, we would like as dependent variable a measure of total costs for elective activity only. However, this information is not available as hospitals' costs include all the costs for both elective and emergency activity, and outpatient activity. Nevertheless, we are able to control for both emergency activity and outpatient activity (see description of the variables below). Therefore, the variable "total hospital costs" *conditional* on the volume of emergency and outpatient activity provides a proxy for the costs of operating expenses for elective surgery.[5]

Our measure of waiting times is the average wait for elective admissions, which was provided by HES. It measures the average number of days between the decision of being admitted to the waiting list and the actual admission for treatment. Notice that the average waiting time is calculated across all elective (non-emergency) admissions and across all HRGs. It includes both daycases and inpatient admissions. Only emergency patients are excluded.

HES defines as 'elective' all admission episodes where the admission was from a waiting list (codes 11 and 12), in contrast to 'emergency' admissions (codes 21–24, and 28). Elective and emergency activity refers to all the (more than 400) HRG episodes (Health Care Resource Groups, the equivalent of the Medicare DRGs) treated within the hospital. The proportion of 'elective' admissions can vary substantially across HRGs. For example in year 2001/2002, the proportion of elective patients was more than 90% for procedures like cataract (HRG B02, B03), 87% for Coronary Bypass (HRG E04), 48% for major procedures at the Oesophagus (HRG F03), 2% for Pulmonary Disease or Bronchitis (HRG D20), less than 1% for Acute Myocardial Infarction (HRG E11, E2). For a full

[5] To check the robustness of our results, we have also re-estimated the analysis using "total operating expenses" as an alternative dependent variable. "Total operating expenses" includes total expenditures on salaries and wages, clinical and general supplies, establishment expenditures and premises and fixed plant expenditures. The correlation between the two variables is 0.97. The main results are qualitatively very similar.

Table 1
Description of variables.

Variable name	Description	Source[a]
(a) Hospital cost		
Total cost	Total hospital cost (pounds 000) (2002 real values using Treasury GDP deflator)	DoH
Variable cost	Hospital operating expenses (pounds 000) (2002 real values using Treasury GDP deflator)	DoH
(b) Waiting times		
Waiting time	Mean waiting time (days)	HES
(c) Measures of activity		
Inpatient spells	Total inpatient spells (000)	HES
Outpatient spells	Total outpatient attendances (000)	DoH
(d) Case mix		
Emergency admissions	Emergency admissions as % of total inpatient spells	HES
HRG case mix	HRG casemix index based on Reference Costs	NHSIA
(e) Efficiency on use of resources		
Day cases	Number of day cases as % of elective admissions	HES
Length of stay	Average length of stay	HES
(f) Capital inputs		
Beds	Number of available beds	DoH
(g) Quality of services		
Readmission	% of emergency re-admissions within 28 days (standardised by age and gender)	DoH
(h) Competition		
Competition	Number of hospitals within a 20 km radius	HES
(i) Dummy variables		
London	Trust is in London	CIPFA

[a] DoH: Department of Health; HES: Hospital Episodes Statistics; NHSIA: National Health Service Information Authority; CIPFA: The Chartered Institute of Public Finance and Accountancy.

Table 2
Descriptive statistics.

Variable	Observations	Mean	Median	S.D.	Min	Max
Total cost	440	104,731.5	94,147.6	42,563.7	26,096.4	260,045.0
Variable cost	440	90,170.5	80,773.2	37,360.2	22,029.5	234,260.2
Waiting time	440	102.9	101.0	29.9	17.0	219.0
Inpatient spells	440	56.0	51.1	22.4	4.2	131.4
Outpatient spells	440	215.6	195.9	87.8	35.5	609.3
Emergency admissions	440	36.1	35.7	5.1	4.5	65.0
HRG case mix	440	93.7	93.3	7.1	75.5	163.4
Day cases	439	50.1	49.8	8.0	0.0	77.2
Length of stay	440	5.3	5.2	1.7	2.5	28.9
Readmission	385	5.8	5.7	0.9	3.7	10.2
Beds	440	681.8	626.5	255.2	166.0	1,574.7
Competition	359	4.5	2.0	4.8	1.0	19.0
London	440	0.1	0.0	0.4	0.0	1.0
Year 1998	440	0.3	0.0	0.4	0.0	1.0
Year 1999	440	0.3	0.0	0.4	0.0	1.0
Year 2000	440	0.2	0.0	0.4	0.0	1.0
Year 2001	440	0.2	0.0	0.4	0.0	1.0

description by HRG see HES website (www.hesonline.nhs.uk). In our sample, around 36% of inpatient spells are originated as emergency admissions, while the remaining are elective.[6]

Table 1 provides a description of the variables employed in the analysis and corresponding sources of data. We divide the explanatory variables into six groups. Hospital activity is measured by the total number of inpatient spells and the total number of outpatient attendances. Both variables are measured in 1000 cases. A second group of variables captures the severity of cases treated by the hospital and the demand on resources. It includes emergency admissions as a proportion of total spells and an HRG (Healthcare Resource Group) casemix index based on reference costs (this is equivalent to the case-mix adjustments based on DRGs in other countries, like the Medicare Programme in the US or Italy). Controlling for hospital's severity might be important, if hospitals with longer waiting times increase the severity of the patients (i.e. the delay in treatment makes on average the condition worse).

Hospital costs also depend on the efficiency in the use of resources. We control for the number of day cases as a proportion of elective surgeries and the average length of stay. More efficient hospitals are expected to have a higher proportion of surgeries carried out on a day case basis, and a lower average length of stay. The capital stock is proxied by the number of available beds (Vita, 1990; Jacobs et al., 2006, p. 31).

[6] We think that our waiting-time measures from HES should not be subject to strong incentives to manipulation. Waiting-times manipulation is perhaps more likely for another measure of waiting time, i.e. 'the waiting time of the patients on the list'. This may be manipulated by inflating the number of patients on the list and including also patients that will never receive treatment or have already received treatment. Moreover, while the 'waiting time of the patients on the list' is used as a major performance indicator for managers working in the hospital, the waiting time of the patients admitted for treatment is not.

Table 3
Unbalanced pooled OLS regressions of total hospital cost. Dependent variable: log(totcost).

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(Waiting time)	−1.4 (0.842)	−.93* (0.483)	−.93* (0.479)	−.76* (0.43)	−.73* (0.433)	−0.28 (0.378)	−0.28 (0.354)
[ln(Waiting time)]^2	0.29 (0.184)	.21* (0.107)	.21** (0.106)	.17* (0.095)	.16* (0.096)	0.07 (0.085)	0.07 (0.082)
ln(Inpatient spells)	.55*** (0.061)	.3*** (0.063)	.27*** (0.065)	.28*** (0.065)	.34*** (0.083)	.62*** (0.105)	.65*** (0.127)
ln(Outpatient spells)	.38*** (0.06)	.28*** (0.05)	.28*** (0.051)	.28*** (0.049)	.29*** (0.049)	.23*** (0.048)	.24*** (0.057)
ln(HRG case mix)	1.2*** (0.196)	.81*** (0.145)	.83*** (0.148)	.8*** (0.146)	.77*** (0.143)	.73*** (0.168)	.72*** (0.213)
ln(Beds)		.4*** (0.056)	.43*** (0.062)	.42*** (0.06)	.35*** (0.08)	0.14 (0.098)	0.1 (0.116)
Emergency admissions			−0.003 (0.002)	−0.003 (0.002)	−0.003 (0.002)	−0.002 (0.003)	−0.001 (0.004)
Day cases				−0.002 (0.001)	−.0027* (0.002)	−.0037** (0.002)	−.0034* (0.002)
ln(Length of stay)					0.08 (0.073)	.25*** (0.086)	.26** (0.113)
Readmission						0.002 (0.014)	0.004 (0.02)
ln(Competition)							−0.002 (0.019)
London dummy	.19*** (0.044)	.18*** (0.04)	.17*** (0.041)	.16*** (0.042)	.16*** (0.044)	.15*** (0.046)	.15** (0.061)
Year 1999 dummy	.13*** (0.013)	.13*** (0.012)	.13*** (0.012)	.13*** (0.012)	.13*** (0.013)	.14*** (0.011)	.15*** (0.013)
Year 2000 dummy	.19*** (0.018)	.17*** (0.014)	.17*** (0.014)	.17*** (0.015)	.21*** (0.032)	.28*** (0.037)	.28*** (0.046)
Year 2001 dummy	.14*** (0.018)	.14*** (0.015)	.15*** (0.016)	.15*** (0.017)	.15*** (0.017)	.15*** (0.017)	.14*** (0.02)
Constant	4.7** (1.93)	4.5*** (1.258)	4.4*** (1.25)	4.3*** (1.149)	4.4*** (1.154)	3.9*** (1.105)	3.9*** (1.147)
R^2	0.86	0.89	0.89	0.89	0.89	0.90	0.89
RESET	0.11	0.49	0.49	0.99	1.2	2	1.8
Elasticity (at mean)	1.3 (0.9)	1** (0.5)	1** (0.5)	0.81* (0.5)	0.78* (0.5)	0.32 (0.4)	0.35 (0.4)
Elasticity (at 5% percentile)	1 (0.7)	0.79** (0.4)	0.79** (0.4)	0.64* (0.4)	0.62* (0.4)	0.26 (0.3)	0.29 (0.3)
Elasticity (at 25% percentile)	1.2 (0.8)	0.91** (0.5)	0.91** (0.5)	0.74* (0.4)	0.71* (0.4)	0.29 (0.4)	0.32 (0.4)
Elasticity (at median)	1.3 (0.9)	1** (0.5)	1** (0.5)	0.82* (0.5)	0.79* (0.5)	0.33 (0.4)	0.35 (0.4)
Elasticity (at 75% percentile)	1.4 (0.9)	1.1** (0.5)	1.1** (0.5)	0.88* (0.5)	0.85* (0.5)	0.35 (0.4)	0.38 (0.4)
Elasticity (at 95% percentile)	1.6 (1)	1.2** (0.6)	1.2** (0.6)	0.98* (0.5)	0.93* (0.5)	0.39 (0.5)	0.42 (0.5)
Joint significance[a]	1.4	2	2.1	1.7	1.6	0.35	0.59
Endogeneity[b]	2.3	4.3	4.4	3.5	3.4	3.6	2.9
Over identification[c]	1.7	3.3	3.6	3.1	3	2.2	1.6
N	440	440	440	439	439	384	319
N clusters	137	137	137	137	137	109	88

Robust standard errors in parentheses.
[a] Test for joint significance of ln(Waiting time) and ln(Waiting time)^2.
[b] Test for exogeneity of endogenous regressors (endogenous regressors can actually be treated as exogenous).
[c] Test for validity of instruments (orthogonal to error process and correctly excluded from the main equation).
* $p < .1$.
** $p < .05$.
*** $p < .01$.

The quality of services is proxied by the percentage of emergency readmissions within 28 days from treatment. This variable is standardised by age and gender. Finally, the degree of competition in the geographical market is measured through the number of hospitals within a 20 km radius (Propper et al., 2004; Siciliani and Martin, 2007).

We do not include salaries and other input prices (like supplies and consumables) because information is not readily available. Also, salaries are nationally agreed and therefore there is little variation in salary expenditure across hospitals (similarly, for other input prices).[7]

Table 2 presents some descriptive statistics. For the 440 hospitals observed over the four-year period, the sample average total cost is just below £105 million per year and the average waiting time for elective surgery is about 103 days. The variation across hospitals accounts for most of the variation in these two variables, while there is relatively less variation over time. The within and between sample standard deviations are, respectively, 10,033 and 43,422 for total hospital cost and 10.4 and 29.0 for waiting times. The average hospital in the sample provides around 215,600 outpatient attendances and 56,000 inpatient spells, 36% (20,000) of which are originated as emergency attendances and 64% (36,000) as elective attendances. On average, the HRG casemix index is at 93.7 (with a higher index indicating a more complex mix of cases) and each hospital faces the competition of 4.5 other hospitals in a 20 km radius. With respect to the efficiency of resource use, each hospital admits on average 50% of the elective patients as day cases, with an average length of stay of 5.3 days. The proportion of emergency readmissions within 28 days is around 6%.

With the exception of emergency admissions, readmissions and day cases, all the other continuous variables (including total cost and waiting times) are included in the log scale, which reduces skewness and allows the interpretation of coefficients as elasticities. Emergency admissions, readmissions and day cases are kept in levels. Since they are measured as percentages, the associated coefficients can also be interpreted as elasticities. After the log transformation, the mean total cost in the sample is equal to the median.

5. Results

The results of the regression analysis for pooled OLS and fixed effects are reported in Tables 3 and 4. The dependent variable in both regressions is the log of total hospital cost in real values of year 2002.

Table 3 shows the OLS results for seven different specifications. We add regressors progressively in order to test the stability of results. The basic regression (column (1)) includes mean waiting times (linear and quadratic effect) and activity indicators (inpatient spells and outpatient attendances), and controls for the HRG index, London effect and year.[8] We then progressively add controls

[7] As pointed out by Smet (2002), if factor prices are excluded, we either impose the restriction that input prices are identical across hospitals or that hospital technology is characterised by zero input substitution.

[8] We do not include the interaction term between activity and waiting times, because this variable is never significant. As pointed out by Iversen (1997), waiting

Table 4
Unbalanced fixed effects regressions of total hospital cost. Dependent variable: log(totcost).

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(Waiting time)	−0.2 (0.21)	−0.2 (0.19)	−0.15 (0.18)	−0.16 (0.18)	−0.15 (0.18)	−0.17 (0.18)
[ln(Waiting time)]^2	0.06 (0.051)	0.06 (0.047)	0.05 (0.045)	0.05 (0.045)	0.05 (0.045)	0.05 (0.045)
ln(Inpatient spells)	.15* (0.09)	0.12 (0.11)	.16* (0.092)	.16* (0.092)	.17* (0.093)	−0.02 (0.099)
ln(Outpatient spells)	.12* (0.06)	.11* (0.061)	.12* (0.067)	.11* (0.064)	.12* (0.065)	0.1 (0.061)
ln(HRG case mix)	−0.09 (0.19)	−0.2 (0.18)	−.29* (0.17)	−.3* (0.17)	−.37** (0.18)	−.46** (0.19)
ln(Beds)		.25*** (0.073)	.19*** (0.069)	.18*** (0.069)	.14** (0.072)	.23*** (0.068)
Emergency admissions			.0079** (0.003)	.0075** (0.003)	.007** (0.003)	0.004 (0.004)
Day cases				−0.001 (0.001)	−0.002 (0.001)	−0.002 (0.002)
ln(Length of stay)					.087* (0.053)	.097* (0.054)
Readmission						−0.01 (0.012)
Year 1999 dummy	.13*** (0.009)	.13*** (0.009)	.13*** (0.009)	.13*** (0.01)	.13*** (0.01)	.13*** (0.011)
Year 2000 dummy	.15*** (0.013)	.13*** (0.012)	.12*** (0.012)	.13*** (0.012)	.16*** (0.025)	.17*** (0.026)
Year 2001 dummy	.21*** (0.013)	.21*** (0.012)	.2*** (0.012)	.21*** (0.013)	.2*** (0.012)	.22*** (0.012)
Constant	11*** (1.1)	9.9*** (1.1)	10*** (1)	10*** (1.1)	11*** (1.1)	12*** (1.1)
R^2 within	0.67	0.69	0.70	0.70	0.70	0.70
R^2 between	0.69	0.80	0.73	0.74	0.75	0.78
R^2 overall	0.66	0.80	0.74	0.75	0.74	0.66
$corr(\alpha_i, Xb)$	0.62	0.67	0.59	0.61	0.63	0.63
σ	0.30	0.24	0.26	0.26	0.26	0.29
σ_u	0.30	0.24	0.25	0.25	0.26	0.29
σ_e	0.07	0.06	0.06	0.06	0.06	0.06
ρ	0.95	0.93	0.94	0.94	0.94	0.96
Hausman	122***	80***	98***	95***	94***	82***
Breusch–Pagan	186***	211***	206***	202***	197***	195***
Elasticity (at mean)	0.39 (0.3)	0.36 (0.2)	0.28 (0.2)	0.28 (0.2)	0.26 (0.2)	0.25 (0.2)
Elasticity (at 5% percentile)	0.32 (0.2)	0.3 (0.2)	0.24 (0.2)	0.24 (0.2)	0.22 (0.2)	0.2 (0.2)
Elasticity (at 25% percentile)	0.36 (0.2)	0.34 (0.2)	0.26 (0.2)	0.26 (0.2)	0.24 (0.2)	0.23 (0.2)
Elasticity (at median)	0.39 (0.3)	0.37 (0.3)	0.28 (0.2)	0.29 (0.2)	0.27 (0.2)	0.25 (0.2)
Elasticity (at 75% percentile)	0.41 (0.3)	0.39 (0.3)	0.3 (0.3)	0.3 (0.3)	0.28 (0.3)	0.27 (0.3)
Elasticity (at 95% percentile)	0.45 (0.3)	0.42 (0.3)	0.33 (0.3)	0.33 (0.3)	0.31 (0.3)	0.29 (0.3)
Joint significance[a]	5.3***	3.1**	2.6*	2.3	2.2	0.75
N	440	440	440	439	439	384
N clusters	137	137	137	137	137	109

Robust standard errors in parentheses.
[a] Test for joint significance of ln(Waiting time) and ln(Waiting time)^2.
* $p < .1$.
** $p < .05$.
*** $p < .01$.

for capital stock (available beds (2)), demand on resources (emergency admissions (3)), efficiency on use of resources (daycases (4) and average length of stay (5)), quality of service (emergency readmissions (6)) and competition (number of competitors in a 20 km radius (7)). Given the limited coverage of our sample, the inclusion of the quality and competition indicators reduces significantly the number of observations. In Table 4 there are only six specifications for panel regressions because the competition indicator, the HRG index and the London dummy either do not vary or vary little over time, which prevents fixed-effects estimation.

We initially estimated the regressions using a translog specification, which is a second-order Taylor approximation adding squared terms for the activity indicators. However, since the square and cross terms were not significant (apart from the squared waiting-time effect), we decided to exclude them from the final specification. The OLS regressions were estimated using standard errors robust to both heteroskedasticity and the serial correlation among observations of the same hospital over the years. Thus we report both the total number of observations (*N*) and the number of clusters (*N* clusters).

Table 3 reports pooled cross-section estimates using unbalanced samples. By 'unbalanced' we mean that as additional regressors are added the sample size decreases, falling from 440 observations in the basic regression to 319 observations in the regression with all independent variables.

All the regressions have been estimated with both linear and quadratic effects for waiting times, which allows us to control for nonlinearities in the hospital cost response to waiting times (notice that this also implies that the elasticity of costs to waiting times is not constant). Two reasons guided the choice of this functional form. First, it gives a direct test of Iversen's suggestion of a nonlinear effect of waiting times. As explained in Section 2, we should expect to find a negative coefficient for low levels of waiting times and a positive coefficient for high levels. Second, the inclusion of a quadratic effect of waiting times eliminates misspecification problems. In all the specifications in Table 3 the RESET test is not significant, which suggests that the functional form is correctly specified.

We have also tested for potential endogeneity of waiting times using Hausman endogeneity tests, which compare the results of instrumental variables (IV) estimates with the OLS estimates. We have estimated IV regressions using lagged values of waiting times and waiting times squared at lag-1 and lag-2 as instruments for current waiting times and waiting times squared. In the vast majority of cases the Hausman endogeneity test is unable to reject the null hypothesis that the supposed endogenous regressors (waiting times and waiting times squared) can actually be treated as exogenous. In addition, we have also implemented over-identifying restrictions tests for exogeneity of the instruments (Sargan tests).

times might reduce not only total costs but also the marginal cost of activity, so that $Cyw < 0$. However, we find no evidence for this effect, which might be due to multicollinearity between the variables.

In this case the tests results suggest that the excluded instruments are valid, meaning that it is possible to assume that the instruments used are orthogonal to error process and correctly excluded from the main equation. Since waiting times are found to be exogenous, we choose to report only the OLS regression results, since this method is more efficient compared to IV estimation.

Let us now focus on the coefficients estimated by the regressions, starting with waiting times. In all regressions the coefficient for the linear component is negative, while the quadratic is positive. This implies that waiting times have an initial negative impact on costs. However, after some point the effect is reversed and waiting times start to increase costs. Therefore, in principle there is an optimal level of waiting times that minimises total costs (this optimal level is calculated below).

Although the effect of waiting times is consistent with the theory, the estimated coefficients are not always statistically significant. In the basic regression (column (1) of Table 3), the estimated effect of waiting times is not significant, either jointly or separately. However, the other variables display significant effects. As expected, both inpatient and outpatient activity increase cost, as does the HRG index. On average, hospitals costs in London are approximately 20% higher than in the rest of the country. Real costs increased significantly between 1998 and 1999, possibly due to nation-wide salary increases from 1999/2000 onwards.

Adding available beds (column (2)) affects the results, and the coefficients of waiting times become significant. The effect of available beds, our proxy for capital, is positive and significant in all regressions where it is included. Notice that by controlling for beds (as a proxy of capital), we are implicitly estimating a short-run cost function. This seems plausible for publicly funded hospitals, which have limited discretion over capital decisions. Since the coefficient is significant we can also exclude that hospitals are operating in the long-run equilibrium. Moreover, a positive coefficient implies over-investment in capital (Cowing and Holtmann, 1983; Smet, 2002; Vita, 1990; Grannemann et al., 1986).

The introduction of emergency admissions or day cases does not affect the results significantly (see columns 3 and 4). Emergency admissions have no effect on hospital costs. Day cases have a negative and significant coefficient, suggesting that hospitals with a higher proportion of elective admissions treated as day cases have lower costs. In column (5), we add average length of stay, which has a positive and significant effect on hospital costs, as expected. The effect of waiting times is not altered and the RESET test is still not significant.

Next we include readmission rates (column (6)), which has a positive but not statistically significant effect on hospital cost. In sharp contrast with previous specifications, the effects of waiting times and available beds cease to be significant. One possible explanation for this result is that adding readmission rates causes a sizable reduction in the sample, making the effect of waiting time insignificant.[9] Another explanation is that lower readmission rates (higher quality) generate both higher costs and higher waiting times (through lower demand). But column (7) suggests that the effect of readmission has no significant effect on costs. Also the positive coefficient on readmission rates suggests that lower readmission would reduce costs (in contrast with what we would expect). We therefore favour the first explanation.

Finally, we evaluate the effect of local competition from other hospitals (column (7)). The estimated effect of competition is negative, although not significant.

From Table 3 it is not immediately possible to infer whether the effect of waiting time on costs is positive or negative when evaluated at the sample mean. The regressions include both waiting time and its squared term. Therefore the estimated cost function has non-constant elasticity with respect to waiting time. Recall that at the sample mean the waiting time is 102.9 days. Differentiating the equation in column (3), we obtain $\varepsilon_w^C = \partial \log C_{it}/\partial \log w_{it} = -0.93 + 2 * 0.21 * \log(102.9) = 1.02$. Therefore, the elasticity of cost with respect to waiting time is markedly positive. Using similar computations, we can show that the elasticity is smaller for column (5), $\varepsilon_w^C = 0.75$, and even smaller for column (7), $\varepsilon_w^C = 0.37$. Table 3 also reports the elasticity of hospitals' cost to waiting times at different quantiles of the waiting time distribution (5%, 25%, 50%, 75%, 95%). As expected, the elasticity is higher at higher quantiles throughout the different specifications.

It is also of interest to calculate the level of waiting time which minimises total costs. By setting $\partial \log C_{it}/\partial \log w_{it} = 0$ from columns (3), (5) and (7) we obtain a waiting time respectively equal to 9.2, 9.8 and 7.4 days. Therefore, if waiting times reduce costs for low levels of waiting times, the effect vanishes after waiting time has reached less than ten days. However, it is important to point out that since the average waiting time in the sample is about 100 days, estimates of the effect of waiting time on costs for waiting times less than ten days will be imprecisely estimated.

In addition to the pooled cross-sectional analysis we also estimate fixed- and random-effects panel regressions. Results from the fixed-effects estimations are reported in Table 4. Notice that the time-invariant regressors (like London dummy and number of competitors) are excluded from the fixed-effects regressions. Fixed effects can also assist in controlling for endogeneity in explanatory variables. For instance, it is possible that average length of stay may by endogenous in the sense that a hospital can to a certain extent control average length of stay to maximise revenues or to manage capacity, and fixed effects estimation can be useful to control for this effect.

The effect of waiting times estimated by fixed-effects is qualitatively similar to the pooled OLS case, with a negative coefficient for the linear component and a positive coefficient for the quadratic one. However, the coefficients are not significant (although jointly significant in columns (1) and (2)). This might be due to the inefficiency of the fixed-effects estimator. A more efficient model is the random-effects estimator but this will give unbiased estimates only if the individual-specific effects are not correlated with the independent variables. The Hausman test rejects the random effects model: individual-specific effects are therefore correlated with independent variables. Nevertheless, the random-effects model might provide an idea on the degree of inefficiency of the fixed-effects model. We therefore report in Table 5 also the random effects estimations. In most specifications the effect of waiting times is significant and in accordance with the hypothesis proposed by Iversen (1993). The linear coefficient of waiting times is negative, but the quadratic one is positive, suggesting that increasing waiting times up to a certain level decreases costs, but past this level the effect is reversed. The elasticity of cost with respect to waiting time ($\varepsilon_w^C = \partial \log C_{it}/\partial \log w_{it}$) is always positive at the sample mean for both the fixed and the random effects models, respectively in the range 0.21–0.37 for the fixed effects and 0.31–0.85 for the random effects. Again, the level of waiting time which minimises total costs is below ten days. As expected, the elasticity is higher at higher quantiles throughout the different specifications

[9] We have compared the sample used in the main analysis with the sample which includes only providers for which re-admissions rates are available. The two samples exhibit some differences. The sample used when re-admission rates are included exhibits on average higher costs, lower waiting times, higher inpatient spells and total inpatient attendances, more available beds, lower case-mix and lower length of stay. There are no statistically significant differences in the percentage of total inpatient spells initiated as emergency admissions and the number of day cases as proportion of elective admissions. The results in column (6) need therefore to be interpreted with some caution.

Table 5
Unbalanced random effects regressions of total hospital cost. Dependent variable: log(totcost).

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(Waiting time)	−0.71 (0.55)	−0.54 (0.33)	−0.52 (0.33)	−0.48 (0.3)	−0.46 (0.29)	−0.16 (0.17)
[ln(Waiting time)]^2	0.17 (0.12)	.13* (0.073)	.12* (0.072)	.11* (0.066)	.11* (0.064)	0.05 (0.041)
ln(Inpatient spells)	.44*** (0.058)	.23*** (0.055)	.24*** (0.055)	.25*** (0.052)	.26*** (0.056)	.4*** (0.077)
ln(Outpatient spells)	.31*** (0.047)	.21*** (0.041)	.21*** (0.041)	.21*** (0.04)	.22*** (0.039)	.21*** (0.041)
ln(HRG case mix)	.59*** (0.22)	0.21 (0.16)	0.19 (0.16)	0.16 (0.16)	0.13 (0.16)	0.08 (0.17)
ln(Beds)		.46*** (0.049)	.45*** (0.051)	.43*** (0.048)	.41*** (0.052)	.35*** (0.062)
Emergency admissions			0.002 (0.002)	0.001 (0.002)	0.001 (0.002)	0.0004 (0.003)
Day cases				−0.002 (0.001)	−0.002 (0.001)	−.0033** (0.001)
ln(Length of stay)					0.04 (0.049)	0.08 (0.052)
Readmission						−0.01 (0.011)
Year 1999 dummy	.15*** (0.01)	.17*** (0.01)	.17*** (0.01)	.17*** (0.01)	.16*** (0.01)	.14*** (0.011)
Year 2000 dummy	.13*** (0.015)	.13*** (0.011)	.13*** (0.012)	.13*** (0.012)	.13*** (0.024)	.14*** (0.024)
Year 2001 dummy	.16*** (0.013)	.14*** (0.011)	.14*** (0.011)	.14*** (0.012)	.16*** (0.012)	.18*** (0.013)
Constant	.16*** (1.4)	.17*** (0.95)	.17*** (0.95)	.18*** (0.94)	.18*** (0.95)	.18*** (0.81)
R^2 within	0.60	0.65	0.65	0.65	0.65	0.65
R^2 between	0.84	0.88	0.88	0.88	0.88	0.92
R^2 overall	0.85	0.88	0.88	0.88	0.88	0.89
σ	0.15	0.14	0.14	0.14	0.14	0.12
σ_u	0.14	0.13	0.13	0.12	0.13	0.10
σ_e	0.07	0.06	0.06	0.06	0.06	0.06
ρ	0.82	0.80	0.80	0.80	0.80	0.72
Elasticity (at mean)	0.81 (0.5)	0.63 (0.3)	0.61 (0.3)	0.55 (0.3)	0.53 (0.3)	0.26 (0.2)
Elasticity (at 5% percentile)	0.64 (0.4)	0.51 (0.3)	0.49 (0.3)	0.44 (0.2)	0.43 (0.2)	0.21 (0.2)
Elasticity (at 25% percentile)	0.74 (0.5)	0.58 (0.3)	0.56 (0.3)	0.51 (0.3)	0.49 (0.3)	0.24 (0.2)
Elasticity (at median)	0.82 (0.5)	0.64 (0.3)	0.62 (0.3)	0.56 (0.3)	0.54 (0.3)	0.26 (0.2)
Elasticity (at 75% percentile)	0.87 (0.6)	0.69 (0.4)	0.66 (0.4)	0.6 (0.3)	0.58 (0.3)	0.28 (0.2)
Elasticity (at 95% percentile)	0.96 (0.7)	0.75 (0.4)	0.73 (0.4)	0.66 (0.4)	0.64 (0.4)	0.3 (0.3)
Joint significance[a]	5.1*	5.5*	5*	4.3	4.2	4.5
N	440	440	440	439	439	384
N clusters	137	137	137	137	137	109

Robust standard errors in parentheses.
[a] Test for joint significance of ln(Waiting time) and ln(Waiting time)^2.
* $p < .1$.
** $p < .05$.
*** $p < .01$.

(but always in the range 0.2–0.45 for the fixed-effects model and in the range 0.21–0.96 for the random-effects model).

6. Conclusions and policy implications

Waiting times are a significant feature of several healthcare systems. This paper has investigated the association between waiting times and hospital costs. Iversen (1993) has argued that for low waiting times, higher waiting times reduce costs due to lower idle capacity, but there might be a point over which higher waiting times increase costs, due to the higher costs of managing the waiting list. Using a sample of 137 acute hospitals over the period 1998–2002 in the English National Health Service (NHS) we have tested empirically the relationship between hospital costs and waiting times. Our cross-sectional and panel-data results suggest that at the sample mean (103 days), waiting times have no significant impact on hospitals' costs or, at most, a positive one. If demand could be rationed explicitly, waiting times should be below or at most equal to the cost minimising level. However, since waiting times play a rationing role, waiting times are so high that at the margin an increase in waiting time may increase hospitals' cost.

An alternative possible explanation for our results is that waiting times are actually a proxy for hospital unobserved, fixed inefficiency. Indeed, although the OLS estimates of the impact of waiting times on costs are often positive and statistically significant, the coefficient is generally not significant if a control for hospital fixed effects is allowed.

Although waiting times play a significant rationing role in many health care systems, other forms of rationing may increase welfare. As pointed out for example by Barzel (1974), waiting times generate a loss to patients but do not generate benefits for the providers (at least if waiting time is weakly above the cost-minimising level). If expected benefit was perfectly observable by the provider, an ideal rationing mechanism would provide swift treatment to patients with high expected benefit, refuse treatment to patients with low expected benefit, and set a waiting time which is strictly below the cost minimising level. Recent policies that focus on the development of explicit prioritisation criteria (Siciliani and Hurst, 2005) might encourage clinicians in the future to rely more on explicit rationing and less on waiting-time rationing.

Acknowledgements

We thank two anonymous referees, Jan Erik Askildsen, Tor Iversen, Andrew Jones, Sverre Kittelsen, Michael Hoel, Giorgia Marini, Nigel Rice, Eric Nævdal, Peter C. Smith and other seminar participants at the University of Oslo and at the University of York for helpful comments and suggestions. Anderson Stanciole gratefully acknowledges financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior of the Ministry of Education of Brazil (Grant 147402-2).

References

Aletras, V., Jones, A., Sheldon, T., 1997. Economies of scale and scope. In: Ferguson, B., Sheldon, T., Posnett, J. (Eds.), Concentration and Choice in Health Care. Royal Society of Medicine Press, Glasgow.
Arocena, P., García-Prado, A., 2007. Accounting for quality in the measurement of hospital performance: evidence from Costa Rica. Health Economics 16 (7), 667–685.

Barros, P., Olivella, P., 2005. Waiting lists and patient selection. Journal of Economics and Management Strategy 14 (3), 623–646.
Barzel, Y., 1974. A theory of rationing. Journal of Law and Economics 17, 73–95.
Brekke, K., Siciliani, L., Straume, O.R., 2008. Competition and waiting times in hospital markets. Journal of Public Economics 92 (7), 1607–1628.
Breyer, F., 1987. The specification of a hospital cost function: a comment on the recent literature. Journal of Health Economics 6 (2), 147–157.
Cameron, A.C., Trivedi, P., 2005. Microeconometrics—Methods and Applications. Cambridge University Press, New York.
Chalkley, M., Malcomson, J.M., 1998. Contracting for health services with unmonitored quality. The Economic Journal 108, 1093–1110.
Cowing, T.G., Holtmann, A.G., 1983. Multiproduct shortrun hospital cost functions: empirical evidence and policy implications from cross-section data. Southern Economic Journal 49 (3), 637–653.
Cullis, J., Jones, P., Propper, C., 2000. Waiting lists and medical treatment: analysis and policies. In: Culyer, A., Newhouse, J. (Eds.), Handbook of Health Economics. Elsevier, Amsterdam (Chapter 23).
Dafny, L.S., 2005. How do hospitals respond to price changes? American Economic Review 95 (5), 1525–1547.
Dawson, D., Gravelle, H., Jacobs, R., Martin, S., Smith, P.C., 2007. The effects of expanding patient choice of provider on waiting times: evidence from a policy experiment. Health Economics 16, 113–128.
Dranove, D., Satterthwaite, M., 2000. The industrial organisation of health care markets. In: Culyer, A., Newhouse, J. (Eds.), Handbook of Health Economics. Elsevier, Amsterdam (Chapter 20).
Dusheiko, M., Goddard, M., Gravelle, H., Jacobs, R., 2006. Trends in health care commissioning in the English NHS: an empirical analysis. Centre for Health Economics, University of York; CHE Research Paper 11.
Ellis, R.P., McGuire, T.G., 1996. Hospital response to prospective payment: moral hazard, selection, and practice-style effects. Journal of Health Economics 15 (3), 257–277.
Ellis, R.P., 1998. Creaming, skimping and dumping: provider competition on the intensive and the extensive margins. Journal of Health Economics 17 (5), 537–555.
Evans, W.N., Beomsoo, K., 2006. Patient outcomes when hospitals experience a surge in admissions. Journal of Health Economics 25 (2), 365–388.
Fenn, P., Davies, P., 1990. Variations in length of stay: a conditional likelihood approach. Journal of Health Economics 9 (2), 223–234.
Friedman, B., Pauly, M.V., 1981. Cost functions for a service firm with variable quality and stochastic demand: The case of hospitals. Review of Economics and Statistics 63, 610–624.
Gaynor, M., Anderson, G., 1995. Uncertain demand, the structure of hospital costs and the cost of an empty bed. Journal of Health Economics 14, 291–317.
Gilman, B.H., 2000. Hospital response to DRG refinements: the impact of multiple reimbursement incentives on inpatient length of stay. Health Economics 9, 277–294.
Goddard, J., Malek, M., Tavakoli, M., 1995. An economic model of the market for hospital treatment for non-urgent conditions. Health Economics 4 (1), 41–55.
Grannemann, T.W., Brown, R.S., Pauly, M.V., 1986. Estimating hospital costs. A multiple-output analysis. Journal of Health Economics 5 (2), 107–127.
Gravelle, H., Siciliani, L., 2008. Is waiting-time prioritisation welfare improving? Health Economics 17 (2), 167–184.
Hauck, K., Street, A., 2007. Do targets matter? A comparison of English and Welsh National Health priorities. Health Economics 16 (3), 275–290.
Ho, V., Hamilton, B.H., 2000. Hospital mergers and acquisitions: does market consolidation harm patients? Journal of Health Economics 19 (5), 767–791.
Iversen, T., 1993. A theory of hospital waiting lists. Journal of Health Economics 12 (1), 55–71.
Iversen, T., 1997. The effect of a private sector on the waiting time in a national health service. Journal of Health Economics 16 (4), 381–396.
Jacobs, R., Smith, P.C., Street, A., 2006. Measuring Efficiency in Health Care. Cambridge University Press.
Joskow, P.L., 1980. The effects of competition and regulation on hospital bed supplies and the reservation quality of the hospital. Bell Journal of Economics, Autumn, 421–447.
Joskow, P.L., 1985. The stochastic determinants of hospital-bed supply: reply. Journal of Health Economics 4 (2), 183–185.
Keeler, T.E., Ying, J.S., 1996. Hospital costs and excess bed capacity: a statistical analysis. Review of Economics and Statistics 78, 470–481.
Le Grand, J., 2002. The Labour Government and the national health service. Oxford Review of Economic Policy 18 (2), 137–153.
Lindsay, C., Feigenbaum, B., 1984. Rationing by waiting lists. The American Economic Review 74 (3), 404–417.
Ma, C.A., 1994. Health care payment systems: cost and quality incentives. Journal of Economics and Management Strategy 3 (1), 93–112.
Martin, S., Smith, P.C., 1996. Explaining variations in inpatient length of stay in the National Health Service. Journal of Health Economics 15 (3), 279–304.
Martin, S., Smith, P.C., 1999. Rationing by waiting lists: an empirical investigation. Journal of Public Economics 71 (1), 141–164.
Martin, S., Rice, N., Jacobs, R., Smith, P.C., 2007. The market for elective surgery: joint estimation of supply and demand. Journal of Health Economics 26 (2 (1)), 263–285.
McGuire, A., Hughes, D., 2003. Stochastic demand, production responses and hospital costs. Journal of Health Economics 22 (6), 999–1010.
Mulligan, J.G., 1985. The stochastic determinants of hospital-bed supply. Journal of Health Economics 4 (2), 177–181.
Olivella, P., 2003. Shifting public-health-sector waiting lists to the private sector. European Journal of Political Economy 19 (1), 103–132.
Pope, G., 1989. Hospital nonprice competition and medicare reimbursement policy. Journal of Health Economics 8 (2), 147–172.
Propper, C., Burgess, S., Green, K., 2004. Does competition between hospitals improve the quality of care? Hospital death rates and the NHS internal market. Journal of Public Economics 88 (7–8), 1247–1272.
Propper, C., Burgess, S., Gossage, D., 2008. Competition and quality: evidence from the NHS Internal Market 1991–1999. The Economic Journal 118, 138–170.
Siciliani, L., 2005. Does more choice reduce waiting times? Health Economics 14 (1), 17–23.
Siciliani, L., Hurst, J., 2005. Tackling excessive waiting times for elective surgery: a comparison of policies in twelve OECD countries. Health Policy 72, 201–215.
Siciliani, L., Martin, S., 2007. An empirical analysis of the impact of choice on waiting times. Health Economics 16 (8), 763–779.
Smet, M., 2002. Cost characteristics of hospitals. Social Science and Medicine 55 (6), 895–906.
Smet, M., 2004. Multi-product costs and standby capacity derived from queuing theory: the case of Belgian hospitals. Applied Economics 36 (13), 1475–1487.
Smith, P., 2002. Performance management in British health care: will it deliver? Health Affairs 21 (3), 103–115.
Vita, M.G., 1990. Exploring hospital production relationships with flexible functional forms. Journal of Health Economics 9 (1), 1–21.
Vitaliano, D.F., 1987. On the estimation of hospital cost functions. Journal of Health Economics 6 (4), 305–318.
Xavier, A., 2003. Hospital competition, GP fundholders and waiting times in the UK internal market: the case of elective surgery. International Journal of Health Care Finance and Economics 3, 25–51.
Zwanziger, J., Melnick, G.A., 1988. The effects of hospital competition and the Medicare PPS program on hospital cost behavior in California. Journal of Health Economics 7 (4), 301–320.