Competition and Waiting Times in Hospital Markets

Kurt R. Brekke[*] Luigi Siciliani[†] Odd Rune Straume[‡]

January 15, 2007

(INCOMPLETE AND PRELIMINARY)

Abstract

This paper studies the impact of hospital competition on waiting times. We use a Salop-type model, with hospitals that differ in (geographical) location and, potentially, waiting time, and two types of patients; high-benefit patients who choose between neighbouring hospitals (competitive segment), and low-benefit patients who decide whether or not to demand treatment from the closest hospital (monopoly segment). Compared with a benchmark case of regulated monopolies, we find that hospital competition leads to longer waiting times in equilibrium if the competitive segment is sufficiently high. Given a policy regime of hospital competition, the effect of *increased competition* depends on the parameter of measurement: Lower travelling costs increase waiting times, higher hospital density reduces waiting times, while the effect of a larger competitive segment is ambiguous. We also show that, if the competitive segment is large, hospital competition is socially preferrable to regulated monopolies only if the (regulated) treatment price is sufficiently high.

Keywords: Hospitals; Competition; Waiting times

JEL Classification: H42; I11; I18; L13

[*]Department of Economics, Health Economics Bergen (HEB), Norwegian School of Economics and Business Administration. E-mail: kurt.brekke@nhh.no.

[†]Department of Economics and Related Studies, University of York. E-mail: ls24@york.ac.uk

[‡]Department of Economics and NIPE, University of Minho, and Health Economics Bergen (HEB). E-mail: o.r.straume@eeg.uminho.pt

1 Introduction

Waiting times are a major health policy concern in many OECD countries. Mean waiting times for non-emergency care are above three months in several countries and maximum waiting times can stretch into years. Policymakers often argue that more competition and patient choice can reduce waiting times by encouraging hospitals to compete for patients and revenues (Siciliani and Hurst, 2005). The mechanisms of how this may work are, however, not very clear. Why would hospitals that operate at full capacity and face excessive demand have an incentive to compete for even more patients? The main purpose of this paper is to contribute to the understanding of the relationship between competition and waiting times in hospital markets.

We develop a model of hospital competition within a Salop framework, where hospitals differ in terms of (geographical) location and, possibly, waiting times. We assume that there are two types of patients who differ in expected benefit ("high" and "low") from hospital treatment. Hospitals compete on the segment of demand with high benefit, while they are local monopolists on the demand segment with low benefit. By comparing with a benchmark case of regulated monopolies, we analyse how the introduction of competition in the hospital market affects waiting time and activity in equilibrium. Given a policy regime of hospital competition, we also examine the effects of increasing the *degree of competition*, based on three different measures: (i) patients' travelling costs, (ii) the size of the competitive relative to the monopolistic demand segment, and (iii) hospital density. We also derive the socially optimal waiting time and assess the welfare implications of hospital competition.

Most of the existing literature assumes that hospitals are local monopolists (Lindsay and Feigenbaum, 1984; Iversen, 1993, 1997; Smith, 1999; Olivella, 2002; Barros and Olivella, 2005; see Cullis, Jones and Propper, 2000, for a review of the literature). Two exceptions are Xavier (2003) and Siciliani (2005) who model competition within a Hotelling framework and in a Cournot-type model with differentiated products, respectively. In these models, competition takes the form of duopoly, with the degree of competition being measured by the substitutability between treatments at the two hospitals, and both find that increased competition (or increased patient choice) leads to *longer* waiting times in equilibrium. An arguable limitation of both these studies is that the analysis of a potential competition effect is confined to a single competition measure that leaves considerable room for interpretation, not least because measures of the degree of competition also tend to have a direct demand effect. Furthermore, the lack of a welfare analysis leaves the more fundamental question of whether hospital competition is desirable in the first place, unanswered.

In the present paper, we complement and extend these studies in several different ways. First, we isolate a pure competition effect by considering regulated monopolies versus competition, something which has not been done in the previous literature on hospital competition and waiting times. Second, the richness of our model allows us to use several different measures of the degree of hospital competition, something that turns out to have a crucial impact with respect to both waiting times and activity levels. Third, we include a welfare analysis where we analyse the question of whether hospital competition is socially desirable within a context of third-party funding and waiting times. We also deviate from the above mentioned studies by explicitly modelling (partly) altruistic health care providers.

We find that *introducing competition*, by allowing previously regulated monopolies to compete for patients (equivalently, to introduce free patient choice), leads to an increase in equilibrium waiting times (with a corresponding reduction in hospital activity) if the competitive demand segment is sufficiently large relative to the monopoly segment, and vice versa. Thus, our result contrasts the previously derived result in the literature about the relationship between competition and waiting times. Also, given a competition regime, we find that *increasing the degree of competition* has ambiguous effects on waiting times, depending on the measure of competition. Lower travelling costs for patients increase waiting times, a larger competitive segment has an indeterminate effect, while higher hospital density reduces waiting times.[1]

[1] Siciliani and Martin (2007) provide empirical evidence supporting this relationship between hospital

Furthermore, the relationship between competition and hospital activity is often counter-intuitive. For example, lower travelling costs, which – all else equal – increase demand for hospital treatment, lead in equilibrium to lower hospital activity due to the corresponding increase in waiting time. Similarly, higher hospital density, which – all else equal – reduce demand for hospital treatment, lead in equilibrium to higher per hospital activity due to the corresponding reduction in waiting time.

Regarding social welfare, we show that, if the competitive demand segment is relatively large, hospital competition is socially desirable, compared with regulated monopolies, only if the (regulated) price per treatment is sufficiently high. For a small competitive demand segment, the result is turned upside down; in this case, competition is desirable only if the treatment price is sufficiently low. However, the socially optimal waiting time is attainable through optimal price setting, regardless of market regime.

The rest of the paper is organised as follows. The model is presented in Section 2, while, in Section 3, we derive and characterise the equilibrium waiting time. The effects on waiting time and hospital activity of, first, introducing competition, and, second, increasing the degree of competition, are analysed in Section 4. In Section 5 we derive and characterise the socially optimal waiting time and assess the social desirability of introducing competition in a hospital market. Finally, Section 6 concludes the paper.

2 Model

Consider a market for elective hospital treatment where n hospitals are equidistantly located on a circle with circumference equal to 1. There are two patient types – L and H – differing with respect to the gross valuation of treatment. Both types are uniformly distributed on the circle. A patient demands either one treatment from the most preferred hospital, or no treatment at all.

The utility of an H-type patient who is located at x and seeking treatment at hospital

density and waiting times.

i, located at z_i, is given by[2]

$$U^H(x, z_i) = V - t|x - z_i| - w_i, \tag{1}$$

where w_i is the waiting time at hospital i and t is a travelling cost parameter.

Equivalently, the utility of a L-type patient who is located at x and seeking treatment at hospital i, located at z_i, is given by

$$U^L(x, z_i) = v - t|x - z_i| - w_i, \tag{2}$$

where $V > v$. We concentrate on cases where the H-segment is always covered, while the L-segment is only partially covered. That is, some L-patients will not seek treatment in equilibrium. We assume that the H-segment constitute a share λ of the total number of patients, which is normalised to 1.

Since the distance between hospitals is equal to $1/n$, the H-patient who is indifferent between seeking treatment at hospital i and hospital j is located at x_i^H, given by

$$V - tx_i^H - w_i = V - t\left(\frac{1}{n} - x_i^H\right) - w_j,$$

yielding

$$x_i^H = \frac{1}{2t}\left(w_j - w_i + \frac{t}{n}\right). \tag{3}$$

Total demand for hospital i from the H-segment is given by $X_i^H = 2x_i^H$.

L-patients seek treatment only at the nearest hospital, if at all. The L-patient who is indifferent between treatment at hospital i and no treatment is located at x_i^L, given by

$$v - tx_i^L - w_i = 0,$$

[2]This formulation is consistent with Lindsay and Feigenbaum (1984) where patients have to afford a fixed cost to obtain health care.

yielding

$$x_i^L = \frac{v - w_i}{t}. \tag{4}$$

Total demand for hospital i from the L-segment is given by $X_i^L = 2x_i^L$. Total demand facing hospital i from both segments is thus given by

$$X_i^D = \lambda X_i^H + (1-\lambda) X_i^L = \frac{2(1-\lambda)v - w_i(2-\lambda) + \lambda w_j}{t} + \frac{\lambda}{n}, \tag{5}$$

where $\lambda \in (0,1)$. Notice that $X_i^D \in \left(\frac{\lambda}{n}, \frac{1}{n}\right)$, while total demand is given by $X^D := \sum_{i=1}^n X_i^D \in (\lambda, 1)$. To gain a better understanding of the mechanisms of the model, it is useful to see how demand reacts to changes in waiting times at the hospital level. From (5) we see that

$$\frac{\partial X_i^D}{\partial w_i} = -\frac{2-\lambda}{t} < 0. \tag{6}$$

Notice that lower travelling costs makes it less costly to for patients to demand treatment, or to switch between hospitals; this increases the demand responsiveness to changes in waiting times. However, since the demand loss due to increased waiting time is larger in the L-segment, a larger competitive segment (i.e., and increase in λ) will reduce the demand responsiveness to changes in waiting times.

Hospitals are prospectively financed by a public payer offering a lump-sum transfer T and a per-treatment price p. The objective function of hospital i is assumed to be given by

$$\pi_i = T + pX_i^S + \alpha B_i(w_i, w_j) - C\left(X_i^S\right) - F, \tag{7}$$

where X_i^S is the supply of hospital treatments. Apart from fixed hospital costs, F, the cost of supplying hospital treatments is given by an increasing and strictly convex cost function $C(\cdot)$. The convexity of the cost function captures a presumably important feature in the context of waiting times, namely that hospitals face come capacity constraints.[3] The function $B_i(\cdot)$ gives the benefit of the patients from receiving treatment at hospital i, while

[3] A convex variable cost function is also supported by evidence suggesting that economies of scale are quite rapidly exhausted in the hospital sector (see, e.g., Ferguson et al., 1999, and Folland et al., 2004, for literature surveys).

the parameter $\alpha \in [0,1]$ captures the degree of altruism of the provider.[4] More explicitly, the surplus to patients treated at hospital i is given by

$$\begin{aligned} B_i\left(w_i, w_j\right) &= 2\lambda \int_0^{\frac{1}{2t}\left(w_j - w_i + \frac{t}{n}\right)} \left(V - w_i - tx\right) dx \\ &\quad + 2\left(1-\lambda\right) \int_0^{\frac{v-w_i}{t}} \left(v - w_i - tx\right) dx, \end{aligned} \tag{8}$$

where the first term is the surplus to H-type patients, and the second term is the surplus to the L-type patients.

Differentiating (8), we obtain

$$\frac{\partial B_i\left(w_i, w_j\right)}{\partial w_i} = -X_i^D - \frac{\lambda}{t}\left(V - \frac{w_i + w_j}{2} - \frac{t}{2n}\right) < 0. \tag{9}$$

A marginal reduction in the waiting time of hospital i has two effects. First, it reduces the waiting time, and thus increases utility, for all existing patients at hospital i. This is represented by the first term in (8). Second, it increases demand for treatment at hospital i. At the margin, the increased demand from the L-segment represents a zero utility contribution. However, in the H-segment, there is an inflow of patients with a strictly positive net utility of hospital treatment. This is represented by the second term in (8). Obviously, the magnitude of this second effect depends on the size of the competitive segment, λ. Notice also that patient surplus at hospital i is a convex function of w_i (implying that the altruistic disutility of waiting is concave in w_i).[5]

3 Equilibrium waiting times

In deriving the equilibrium, we assume, as is commonly done, that waiting time acts as a re-equilibrating mechanism between demand and supply, i.e., $X^D\left(w_i, w_j\right) = X^S$.[6] This

[4] This formulation is consistent with Chalkley and Malcomson (1998).

[5] From (9) we derive $\frac{\partial^2 B_i\left(w_i, w_j\right)}{\partial w_i^2} = \frac{4-\lambda}{2t} > 0$.

[6] See Lindsay and Feigenbaum (1984), Gravelle, Smith and Xavier (2003), Iversen (1993, 1997), Martin and Smith (1999), Siciliani (2005).

implies that it is equivalent whether we maximise the hospital objective function with respect to supply or waiting time. For analytical purposes, we use the latter approach.

Thus, the hospitals simultaneously and independently choose announced waiting times, in order to maximise their objective functions. We assume that the hospitals are not able/allowed to discriminate between different patient types with respect to waiting times. We also assume that hospitals cannot turn down patients seeking treatment. This latter assumption implies that we do not allow for explicit rationing.

Substituting (5) into (7) and maximising (7) with respect to waiting time yields the following first-order condition for hospital i,

$$\frac{\partial \pi_i}{\partial w_i} = \left[p - C'\left(X_i\left(w_i, w_j\right)\right)\right] \frac{\partial X_i\left(w_i, w_j\right)}{\partial w_i} + \alpha \frac{\partial B_i\left(w_i, w_j\right)}{\partial w_i} = 0, \tag{10}$$

which implicitly defines a best response function $w_i\left(w_j\right)$. Notice that we have skipped the superscript on the demand function.[7]

Differentiating (10), we see that waiting times are strategic complements:[8]

$$\frac{dw_i}{dw_j} = -\frac{\partial^2 \pi_i / \partial w_j \partial w_i}{\partial^2 \pi_i / \partial w_i^2} = \frac{\left(C''\left(\cdot\right) \frac{2-\lambda}{t} - \alpha\right) \frac{\lambda}{t} + \alpha \frac{\lambda}{2t}}{\left(C''\left(\cdot\right) \frac{2-\lambda}{t} - \alpha\right) \frac{2-\lambda}{t} - \alpha \frac{\lambda}{2t}} > 0 \tag{11}$$

If, say, firm j increases its waiting time, some (H-type) consumers switch to hospital i, which now faces a higher demand. To meet this increase in demand, hospital i has to increase its supply, but this would increase the marginal costs, making the first term in (10) more positive, implying that $\partial \pi_i / \partial w_i > 0$. Since the price is fixed, we see from the first-order condition that the optimal response for hospital i to a higher w_j, is to reduce demand by increasing its waiting time, w_i, until the level where $\partial \pi_i / \partial w_i = 0$. Thus, waiting times are strategic complements for competing hospitals.

In a symmetric equilibrium, $w_j = w_i = w^*$. Using (5) and (6), the equilibrium waiting

[7] The second-order condition is $\partial^2 \pi_i / \partial w_i^2 = -\left[\left(C''\left(\cdot\right) \frac{2-\lambda}{t} - \alpha\right) \frac{2-\lambda}{t} - \alpha \frac{\lambda}{2t}\right] < 0$, which is always satisfied for sufficiently convex cost function; also, $\partial^2 \pi_i / \partial w_j \partial w_i = \left(C''\left(\cdot\right) \frac{2-\lambda}{t} - \alpha\right) \frac{\lambda}{t} + \alpha \frac{\lambda}{2t} > 0$, which is always positive whenever $\partial^2 \pi_i / \partial w_i^2 < 0$.

[8] The denominator is positive by the second-order condition. The numerator is also positive as $C''\left(\cdot\right) \frac{2-\lambda}{t} - \alpha > 0$ is required for the second-order condition to be satisfied.

time is given by the pair of equations

$$-\frac{(2-\lambda)}{t}\left[p-C'\left(X_i\left(w^*\right)\right)\right]=\alpha\left[X_i\left(w^*\right)+\frac{\lambda}{t}\left(V-w^*-\frac{t}{2n}\right)\right], \quad i=1,2, \tag{12}$$

where

$$X_i\left(w^*\right)=2\left(1-\lambda\right)\left(\frac{v-w^*}{t}\right)+\frac{\lambda}{n}, \tag{13}$$

and $w^*=w^*\left(v,t,\lambda,p,n\right)$.[9] Since the right-hand side of (12) is positive, the expression in the square brackets on the left hand side of (12) must be negative in an interior solution. Thus, the equilibrium waiting time is such that the (regulated) price is lower than the marginal treatment cost. In other words, the marginal patient is financially unprofitable to treat for the hospital.

We want to focus on equilibria with strictly positive waiting times. This requires that the cost of treating the last patient who demands treatment at $w=0$ is larger than the treatment price p. This requirement will be met if the supply cost function is sufficiently convex. Furthermore, we restrict attention to interior solutions with a partially covered L-segment in equilibrium, i.e., $x_i^L\in\left(0,\frac{1}{2n}\right)$.

Proposition 1 *Assume that the degree of altruism is sufficiently small. Then there exists an equilibrium waiting time, implicitly defined by (12), which is positive and involves a partially covered L-segment, if* $p\in S:=\left(\underline{p},\min\left\{\overline{p}_1,\overline{p}_2\right\}\right)$, *where* $\underline{p}$ *and* $\overline{p}_1$ *are implicitly defined by*

$$\underline{p}=C'\left(\frac{\lambda}{n}\right)-\frac{\alpha t}{2-\lambda}\left[\frac{\lambda}{n}+\frac{\lambda}{t}\left(V-w^*(\underline{p})-\frac{t}{2n}\right)\right]$$

and

$$\overline{p}_1=C'\left(\frac{1}{n}\right)-\frac{\alpha t}{2-\lambda}\left[\frac{1}{n}+\frac{\lambda}{t}\left(V-w^*(\overline{p}_1)-\frac{t}{2n}\right)\right],$$

[9] Uniqueness and stability of the equilibrium is confirmed by the positive sign of the Jacobian:

$$\Delta:=\begin{vmatrix}\frac{\partial^2\pi_i}{\partial w_i^2} & \frac{\partial^2\pi_i}{\partial w_j\partial w_i}\\ \frac{\partial^2\pi_j}{\partial w_i\partial w_j} & \frac{\partial^2\pi_j}{\partial w_j^2}\end{vmatrix}=\frac{4}{t}\left(C''\left(\cdot\right)\frac{2-\lambda}{t}-\alpha\right)\left[\left(C''\left(\cdot\right)\frac{2-\lambda}{t}-\alpha\right)\frac{1-\lambda}{t}-\alpha\frac{\lambda}{2t}\right]>0,$$

where the expression in the square brackets is positive whenever the second-order condition is satisfied.

while $\overline{p}_2$ is given by

$$\overline{p}_2 = C'\left(2\left(1-\lambda\right)\frac{v}{t}+\frac{\lambda}{n}\right)-\frac{\alpha t}{2-\lambda}\left[2\left(1-\lambda\right)\frac{v}{t}+\frac{\lambda}{2n}+\frac{\lambda}{t}V\right].$$

The equilibrium waiting time is monotonically decreasing in the treatment price p.

Proof. We start by confirming the last part of the Proposition. By total differentiation of the first-order conditions, we obtain[10]

$$\frac{\partial w^*}{\partial p} = -\frac{(2-\lambda)/t}{2\left[\left(C''\left(\cdot\right)\frac{2-\lambda}{t}-\alpha\right)\frac{(1-\lambda)}{t}-\alpha\frac{\lambda}{2t}\right]} < 0$$

An interior solution with positive equilibrium waiting times requires that the following conditions are met: $w^* > 0$ and $x^L \in \left(0, \frac{1}{2n}\right)$. Assume $x^L = 0$, which implies $X\left(w^*\right) = \frac{\lambda}{n}$. Inserting this into the first-order condition for hospital i, and rearranging, we get

$$p = C'\left(\frac{\lambda}{n}\right)-\frac{\alpha t}{2-\lambda}\left[\frac{\lambda}{n}+\frac{\lambda}{t}\left(V-w^*(p)-\frac{t}{2n}\right)\right]$$

Denote the price that solves this equation by $\underline{p}$. Since $\partial w^*/\partial p < 0$ and $\partial x^L/\partial w < 0$ we know that $x^L > 0$ if $p > \underline{p}$. Now assume $x^L = \frac{1}{2n}$, which implies $X\left(w^*\right) = \frac{1}{n}$. Inserting this into the first-order condition yields

$$p = C'\left(\frac{1}{n}\right)-\frac{\alpha t}{2-\lambda}\left[\frac{1}{n}+\frac{\lambda}{t}\left(V-w^*(p)-\frac{t}{2n}\right)\right].$$

Denote the price that solves this equation by $\overline{p}_1$. Again, since $\partial w^*/\partial p < 0$ and $\partial x^L/\partial w < 0$ we know that $x^L < \frac{1}{2n}$ if $p < \overline{p}_1$. Finally, assume $w^* = 0$, which implies $X\left(0\right) = 2\left(1-\lambda\right)\frac{v}{t}+\frac{\lambda}{n}$. The first-order condition is then given by

$$p = C'\left(2\left(1-\lambda\right)\frac{v}{t}+\frac{\lambda}{n}\right)-\frac{\alpha t}{2-\lambda}\left[2\left(1-\lambda\right)\frac{v}{t}+\frac{\lambda}{2n}+\frac{\lambda}{t}V\right]$$

[10] $\frac{\partial w^*}{\partial p} = -\frac{\begin{vmatrix} \partial^2\pi_i/\partial w_i\partial p & \partial^2\pi_i/\partial w_i\partial w_j \\ \partial^2\pi_j/\partial w_j\partial p & \partial^2\pi_j/\partial w_j^2 \end{vmatrix}}{\Delta}$. Notice that $\partial^2\pi_i/\partial w_i\partial p = \partial^2\pi_j/\partial w_j\partial p$, so that $\frac{\partial w^*}{\partial p} = -\frac{1}{\Delta}\left(\partial^2\pi_i/\partial w_i\partial p\right)\left[\partial^2\pi_j/\partial w_j^2-\partial^2\pi_i/\partial w_i\partial w_j\right] = -\frac{\partial^2\pi_i/\partial w_i\partial p}{\partial^2\pi_j/\partial w_j^2+\partial^2\pi_i/\partial w_i\partial w_j}$.

Denote this price by $\overline{p}_2$. By a similar argument as above, $w^* > 0$ if $p < \overline{p}_2$. Since $\frac{\lambda}{n} < \min\left\{\frac{1}{n}, 2(1-\lambda)\frac{v}{t} + \frac{\lambda}{n}\right\}$, it is straightforward to see that $\underline{p} < \min\{\overline{p}_1, \overline{p}_2\}$, implying that S is non-empty, if α is sufficiently small. ■

The inverse relationship between equilibrium waiting times and the treatment price is easily explained. A higher price simply means that the marginal patient becomes less unprofitable to treat, which dampens the incentive to use waiting time as an instrument to shift demand from unprofitable patients towards neighbouring hospitals.

Notice also that, since positive equilibrium waiting times imply that the marginal patient is unprofitable for the hospitals to treat, the equilibrium is "undercutting proof", in the sense that it is never profitable for a hospital to deviate from the equilibrium by reducing waiting times in order to drive neighbouring hospitals out of the market.

4 The impact of competition on waiting times and activity

We will now use the model to analyse if and how competition in hospital markets affects waiting times and hospital activity in equilibrium. The analysis is done in two steps. We start out by considering the effect of *introducing competition* in a hospital market characterised by regulated monopolies. Subsequently, we consider the effects of different measures to *increase the degree of competition* in a hospital market where there is competition to begin with.

4.1 Introducing competition

Assume that the hospital market described in the previous section consists of regulated monopolies, where patients are allocated to hospitals purely according to geographical distance. If a patient decides to visit a hospital to undergo treatment, she has to attend the nearest hospital. In our model, this means that hospital i's demand from the H-segment is exogenously given by $X_i^H = \frac{1}{n}$. Total demand for hospital i is thus given by

$$X_i^D = \frac{\lambda}{n} + (1-\lambda)\frac{2(v-w_i)}{t}. \tag{14}$$

The surplus to patients treated at hospital i is then given by

$$B_i(w_i) = \lambda 2 \int_0^{\frac{1}{2n}} (V - w_i - tx)\,dx + (1-\lambda)\,2 \int_0^{\frac{v-w_i}{t}} (v - w_i - tx)\,dx, \tag{15}$$

where the first term is the surplus to H-type patients, and the second term is the surplus to the L-type patients. Differentiating (15), we obtain

$$\frac{\partial B_i(w_i, w_j)}{\partial w_i} = -X_i^D. \tag{16}$$

In the absence of competition, notice how the marginal reduction in patient surplus from waiting is lower in absolute value (see (9)).

Inserting (14) into the first-order condition, (10), and applying symmetry, the equilibrium waiting time in a market with regulated monopolies, w^m, is given by[11]

$$-\frac{2(1-\lambda)}{t}\left[p - C'(X_i(w^m))\right] = \alpha X_i(w^m), \quad i = 1, 2, \tag{17}$$

where

$$X_i(w^m) = 2(1-\lambda)\frac{(v - w^m)}{t} + \frac{\lambda}{n}. \tag{18}$$

Comparing (12) and (17) we see that, for $w^* = w^m$, both the left-hand side and the right-hand side of (17) are smaller than the left-hand side and right-hand side of (12). This means that $w^m \lesseqgtr w^*$. A closer scrutiny of the two first-order conditions enables us to derive the following result:

Proposition 2 *Introducing competition in a hospital market with regulated monopolies leads to longer (shorter) waiting times and lower (higher) activity in equilibrium if*

$$\lambda > (<)\, 1 - \frac{t}{2n(V - v)}.$$

[11] The second-order condition is given by $\partial^2 \pi_i / \partial w_i^2 = -\left(C''(\cdot)\,\frac{2(1-\lambda)}{t} - \alpha\right)\frac{2(1-\lambda)}{t} < 0$.

Proof. Subtracting (12) from (17) yields

$$\frac{2}{\alpha}\left[C'\left(X_i\left(w^*\right)\right)-C'\left(X_i\left(w^m\right)\right)\right]-2\left(w^m-w^*\right)=\frac{2\lambda\left(1-\lambda\right)\left(V-v\right)-t\lambda}{n\left(1-\lambda\right)\left(2-\lambda\right)}.$$

Let us first confirm that the left-hand side (LHS) of this equation is monotonic in w^m and w^*. Using (5) and (14), we have that $\partial\left(LHS\right)/\partial w^* = -\frac{2}{\alpha}C''\left(X_i\right)\frac{2-\lambda}{t}+2$ and $\partial\left(LHS\right)/\partial w^m = \frac{2}{\alpha}C''\left(X_i\right)\frac{2(1-\lambda)}{t}-2$. Applying the second-order conditions, it is straightforward to verify that $\partial\left(LHS\right)/\partial w^* < 0$ and $\partial\left(LHS\right)/\partial w^m > 0$. Since $LHS = 0$ if $w^* = w^m$, it follows that $w^* > (<)\, w^m$ if the right-hand side of the equation is negative (positive), which is the case if $\lambda > (<)\, 1-\frac{t}{2n(V-v)}$. Since (13) and (18) are identical for a given waiting time, $w^m < w^*$ implies that $X_i\left(w^m\right) > X_i\left(w^*\right)$ and vice versa. ■

There are two counteracting effects that contribute to this result. First, the demand responsiveness to waiting times, $\partial X_i/\partial w_i$, is affected by the introduction of competition. From (5) and (14) we clearly see that $\partial X_i/\partial w_i$ increases in absolute value with the introduction of competition. Thus, introducing competition means that demand at each hospital becomes more responsive to changes in the waiting time announced by the hospital, and the magnitude of this effect is increasing in λ. This is intuitive, since, without competition, only patients in the L-segment respond to waiting times. So how does the magnitude of $|\partial X_i/\partial w_i|$ affect equilibrium waiting times? Remember that, with a hospital disutility of positive waiting times (due to altruism), the marginal patient is unprofitable to treat. In equilibrium, this financial loss is optimally weighed against the disutility of increasing waiting times. When the demand responsiveness to waiting times increases, each hospital has a stronger incentive to increase the waiting time, since this now becomes a more effective instrument for shifting unprofitable patients to neighbouring hospitals.

However, there is also another effect, related to the altruistic preferences of the hospitals, that works in the opposite direction. Comparing (9) and (16) we see that the utility gain of reduced waiting times is higher under hospital competition. With free patient choice, a reduction in waiting times by hospital i attracts patients from neighbouring hospitals who, due to altruism, contribute positively to the hospital objective function. All else equal, this gives the hospitals incentives to reduce waiting times with the introduction of competition.

Both of the above described effects get stronger when the relative size of the competitive segment increases. However this relationship is more pronounced for the first effect (mainly due to the convexity of treatment costs), which consequently dominates when the competitive segment is sufficiently large. Thus, competition leads to longer waiting times in equilibrium if $\lambda > 1 - \frac{t}{2n(V-v)}$. Furthermore, we see that an increase in t and/or a reduction of n increase the parameter space for which competition leads to longer waiting times. The reason is that higher travelling costs and/or lower hospital density reduce the (altruistic) utility gain of reducing waiting times under competition, as can be seen from (9).

It should be noted that the ambiguous nature of the competition effect on equilibrium waiting times is crucially dependent on the way altruism is modelled, where hospitals are (partly) altruistic only toward their own patients. If instead hospitals cared equally about *all* patients in the market, competition would not influence the effect of waiting time changes on the altruistic component in the hospital objective function.[12] In this case, competition would unambiguously increase waiting times.

Finally, it is important to notice that the introduction of competition does not affect demand *per se*; thus, changes in equilibrium waiting times are driven solely by strategic competition effects.

4.2 Increasing the degree of competition

Depending on interpretation, the effect of increased competition (or increased patient choice) on waiting times and activity can work through three different parameters in the model: t, λ and n. First, a reduction in travelling costs, t, will intensify competition between hospitals in the competitive segment of the market. Second, competition will

[12] Under both competition and regulated monopolies, the effect of a waiting time increase on total patient utility is given by

$$\frac{\partial \left(\sum_{k=1}^{n} B_k\right)}{\partial w_i} = -X_i^D.$$

also naturally increase if a larger share of the total market becomes competitive, i.e., if λ increases. One possible (outside-the-model) interpretation is a reduction in fixed costs of undergoing hospital treatment for some patients, implying that a larger share of patients find themselves in the competitive demand segment. Finally, the number of hospitals in the market, n, is a standard measure of the degree of competition. Below we present the comparative statics results with respect to the different competition measures on both waiting time and activity levels, obtained by total differentiation of (12), applying Cramer's rule.

4.2.1 Lower travelling costs

$$\frac{\partial w^*}{\partial t} = \frac{1}{2}\frac{\left(\frac{2-\lambda}{t}C''(\cdot) - \alpha\right)\frac{\partial X}{\partial t} + \frac{1}{t}\left[(p - C'(\cdot))\frac{2-\lambda}{t} + \alpha\frac{\lambda}{t}(V - w^*)\right]}{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{1-\lambda}{t} - \alpha\frac{\lambda}{2t}} < 0, \tag{19}$$

$$\begin{aligned} \frac{dX(w^*)}{dt} &= \underset{-}{\frac{\partial X}{\partial t}} + \underset{-}{\frac{\partial X}{\partial w}}\underset{-}{\frac{\partial w^*}{\partial t}} \\ &= \frac{-\left[(2-\lambda)(p - C'(\cdot)) + \alpha\lambda(V - w^*)\right] + \alpha\lambda(v - w^*)}{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{2(1-\lambda)}{t} - \alpha\frac{\lambda}{t}}\frac{2(1-\lambda)}{t^3} > 0, \end{aligned} \tag{20}$$

where $\frac{\partial X}{\partial t} = -\frac{2(1-\lambda)(v-w)}{t^2} < 0$.[13,14] Lower travelling costs have two different effects on the hospitals' optimal choice of waiting times. First, there is a direct demand effect, as more patients in the L-segment will seek treatment. Each hospital will meet this demand increase partly by increasing waiting times, and the strength of this response depends on the additional costs of treating more patients relative to the altruistic disutility of longer waiting times. Notice here that a higher level of demand also implies that the utility loss of increasing the waiting time is larger, since there are more patients that need to wait for treatment at hospital i. However, due to the convexity of treatment costs, the net effect

[13] Notice that the first-order condition ensures that the expression in the square bracket of the numerator of $\partial w^*/\partial t$ is negative.

[14] $\frac{\partial w^*}{\partial t} = -\frac{\begin{vmatrix} \partial^2\pi_i/\partial w_i\partial t & \partial^2\pi_i/\partial w_i\partial w_j \\ \partial^2\pi_j/\partial w_j\partial t & \partial^2\pi_j/\partial w_j^2 \end{vmatrix}}{\Delta}$. Notice that $\partial^2\pi_i/\partial w_i\partial t = \partial^2\pi_j/\partial w_j\partial t$, so that $\frac{\partial w^*}{\partial t} = -\frac{1}{\Delta}\left(\partial^2\pi_i/\partial w_i\partial t\right)\left[\partial^2\pi_j/\partial w_j^2 - \partial^2\pi_i/\partial w_i\partial w_j\right] = -\frac{\partial^2\pi_i/\partial w_i\partial t}{\partial^2\pi_j/\partial w_j^2 + \partial^2\pi_i/\partial w_i\partial w_j}$.

is still positive with respect to waiting time. Second, lower travelling costs imply that demand facing each hospital becomes more sensitive to changes in waiting times (see (6)), which means that it becomes more effective to use waiting times as an instrument to shift unprofitable demand to neighbouring hospitals. Thus, both effects contribute to increase equilibrium waiting times as a result of lower travelling costs.

The effect of lower travelling costs on equilibrium hospital activity is given by the sum of a direct positive demand effect and an indirect negative effect through the increase in equilibrium waiting time. We see from (20) that the total effect is negative. It is perhaps surprising that lower travelling costs can actually lead to reduced activity in equilibrium. This can be explained in the following way: since treatment costs are strictly convex, while the disutility of waiting (due to altruism) is concave in w_i, it is more costly for hospitals to meet increased demand by increasing activity, relative to waiting times. Consequently, the hospitals will increase waiting times until the level where the demand increase is completely offset. However, there is a second effect of lower travelling costs, as explained above. The effect on the responsiveness of demand to waiting times implies that the hospitals have incentives to increase demand even beoynd the level where the initial demand increase is nulled out. Thus, a reduction of travelling costs, which initially causes an increase in demand for hospital treatments, will actually lead to lower activity in equilibrium, due to the equilibrium response in waiting times.

4.2.2 A larger competitive segment

$$\frac{\partial w^*}{\partial \lambda} = \frac{1}{2}\frac{\left(\frac{2-\lambda}{t}C''(\cdot) - \alpha\right)\frac{\partial X}{\partial \lambda} + \frac{p-C'(\cdot)}{t} - \frac{\alpha}{t}\left(V - w^* - \frac{t}{2n}\right)}{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{1-\lambda}{t} - \alpha\frac{\lambda}{2t}} \gtrless 0. \tag{21}$$

$$\begin{aligned} \frac{dX(w^*)}{d\lambda} &= \underset{+}{\frac{\partial X}{\partial \lambda}} + \underset{-}{\frac{\partial X}{\partial w}}\underset{-/+}{\frac{\partial w^*}{\partial \lambda}} \qquad (22) \\ &= \frac{1}{2}\frac{-\left(p - C'(\cdot)\right)\frac{2(1-\lambda)}{t^2} + \frac{2\alpha}{t^2}\left[(1-\lambda)\left(V - w^* - \frac{t}{2n}\right) + \lambda(v - w^* - \frac{t}{2n})\right]}{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{1-\lambda}{t} - \alpha\frac{\lambda}{2t}}, \end{aligned}$$

where $\frac{\partial X}{\partial \lambda} = 2(\frac{1}{2n} - \frac{v-w}{t}) > 0$ since, in equilibrium, $x^H = 1/2n$ and $x^L = (v-w)/t$, and, by assumption, $x^L < x^H$.[15]

The first term in the numerator of $\partial w^*/\partial\lambda$ is positive while the second and the third are negative. Notice that even for a low degree of altruism, the effect of λ on waiting time is indeterminate. There are two offsetting effects that contribute to this ambiguity. Since demand is higher from the competitive segment, a higher λ will increase total demand, which – all else equal – contributes to longer waiting times. However, a larger H-segment implies that demand becomes less responsive to changes in waiting times, as seen from (6). This means that it becomes less effective to use waiting times to shift unprofitable patients to neighbouring hospitals, which – all else equal – reduces equilibrium waiting times. The sum of these two effects is indeterminate.

The effect of a larger competitive segment on equilibrium activity is also indeterminate, although clearly positive for sufficiently low values of λ. The first term in the numerator of $dX(w^*)/d\lambda$ is always positive. The second term is given by a weighted average of the utility of a H-type patient and a L-type patient when receiving treatment and located at $x = 1/2n$ (by assumption this utility is positive for the H-type and negative for the L-type). This term is consequently also positive if λ is sufficiently low.

4.2.3 Increased hospital density

$$\frac{\partial w^*}{\partial n} = -\frac{1}{2}\frac{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{\lambda}{n^2} + \alpha\frac{\lambda}{2n^2}}{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{1-\lambda}{t} - \alpha\frac{\lambda}{2t}} < 0 \tag{23}$$

$$\frac{dX(w^*)}{dn} = \underset{-}{\frac{\partial X}{\partial n}} + \underset{-}{\frac{\partial X}{\partial w}}\,\underset{-}{\frac{\partial w^*}{\partial n}} = \frac{1}{2n^2 t}\frac{\alpha\lambda}{\left(C''(\cdot)\frac{2-\lambda}{t} - \alpha\right)\frac{1-\lambda}{t} - \alpha\frac{\lambda}{2t}} > 0 \tag{24}$$

$$\frac{d\left[nX(w^*)\right]}{dn} = X + n\frac{dX}{dn} > 0. \tag{25}$$

Notice that the signs of (23) and (24) are determined by applying the second-order condition.[16]

---

[15] $\frac{\partial w^*}{\partial \lambda} = -\frac{\begin{vmatrix} \partial^2\pi_i/\partial w_i\partial\lambda & \partial^2\pi_i/\partial w_i\partial w_j \\ \partial^2\pi_j/\partial w_j\partial\lambda & \partial^2\pi_j/\partial w_j^2 \end{vmatrix}}{\Delta}$. Notice that $\partial^2\pi_i/\partial w_i\partial\lambda = \partial^2\pi_j/\partial w_j\partial\lambda$, so that $\frac{\partial w^*}{\partial\lambda} = -\frac{1}{\Delta}\left(\partial^2\pi_i/\partial w_i\partial\lambda\right)\left[\partial^2\pi_j/\partial w_j^2 - \partial^2\pi_i/\partial w_i\partial w_j\right] = -\frac{\partial^2\pi_i/\partial w_i\partial\lambda}{\partial^2\pi_j/\partial w_j^2 + \partial^2\pi_i/\partial w_i\partial w_j}$.

[16] $\frac{\partial w^*}{\partial n} = -\frac{\begin{vmatrix} \partial^2\pi_i/\partial w_i\partial n & \partial^2\pi_i/\partial w_i\partial w_j \\ \partial^2\pi_j/\partial w_j\partial n & \partial^2\pi_j/\partial w_j^2 \end{vmatrix}}{\Delta}$. Notice that $\partial^2\pi_i/\partial w_i\partial n = \partial^2\pi_j/\partial w_j\partial n$, so that $\frac{\partial w^*}{\partial n} =$

Increased hospital density unambiguously reduces waiting times in equilibrium. The intuition is quite simple. An increase in n means that – all else equal – each hospital faces a lower demand from the competitive segment. This means, due to the convexity of treatment costs, that the marginal treatment cost (for the last patient) is lower at each hospital. Consequently, the marginal patient becomes less unprofitable to treat and the hospitals will respond by reducing waiting times. Note that increased capacity, in itself, is not enough to reduce waiting times, since the effect on waiting times comes only through the competitive segment, where increased capacity means lower demand for each hospital. This can easily be confirmed by observing that $\partial w^*/\partial n = 0$ if $\lambda = 0$.

There are two effects – one direct and one indirect – of an increase in n on the equilibrium activity at the hospital level. Increased hospital density in the market means that the number of patients treated per hospital from the competitive segment goes down. However, there is an indirect "spillover" effect from the competitive to the monopoly demand segment. Due to the demand effect in the competitive segment, resulting in shorter waiting times, demand increases from the hospitals' monopoly segments. Equation (24) shows that the net effect on demand is positive. In this case, the reduction in waiting times fully compensates for the initial drop in demand. *Total activity* clearly increases with hospital density, given that activity per hospital increases.

The effects of increased hospital competition on waiting times and activity can be summarised as follows:

Proposition 3 *(i) Lower travelling costs increase waiting times and decrease hospital activity.*

(ii) A larger competitive market segment has an indeterminate effect on waiting times and hospital activity. In general, the effect on waiting times is positive if the degree of altruism is sufficiently low, while the effect on activity is positive if the competitive segment is sufficiently small.

(iii) Increased hospital density reduces waiting times and increases activity per hospital,

---

$-\frac{1}{\Delta}\left(\partial^2\pi_i/\partial w_i\partial n\right)\left[\partial^2\pi_j/\partial w_j^2 - \partial^2\pi_i/\partial w_i\partial w_j\right] = -\frac{\partial^2\pi_i/\partial w_i\partial n}{\partial^2\pi_j/\partial w_j^2+\partial^2\pi_i/\partial w_i\partial w_j}$.

as well as total activity in the market.

5 Hospital competition and welfare

Having derived and characterised the equilibrium waiting time, we want to explore the issue of whether competition leads to excessive or suboptimal levels of waiting time from a social welfare perspective. To answer this question, we first need to specify the welfare function. We use the conventional measure of welfare as an unweighted sum of consumers' and producers' surplus. The welfare analysis is conducted at the hospital level; for total welfare just multiply by n.

Since the model is symmetric, the socially optimal waiting time must be uniform across hospitals. Setting $w_i = w_j = w$, the surplus to patients treated at a particular hospital is then given by

$$B(w) = \lambda 2 \int_0^{\frac{1}{2n}} (V - w - tx)\, dx + (1-\lambda)\, 2 \int_0^{\frac{v-w}{t}} (v - w - tx)\, dx, \tag{26}$$

where the first term is the surplus to H-type patients, and the second term is the surplus to the L-type patients. Notice that we are assuming, as we did for the hospitals, that the regulator cannot discriminate between patient types in terms of waiting time. The patient surplus function can be written as

$$B(w) = \frac{\lambda}{n}\left(V - w - \frac{t}{4n}\right) + \frac{(1-\lambda)}{t}(v-w)^2. \tag{27}$$

Not very surprisingly, we see that the consumer surplus is always maximised at zero waiting time.

Writing the social welfare function as the sum of consumers' and producers' surplus net of third-party payments, welfare at the hospital level is given by[17]

$$W(w) = B(w) + T + pX(w) - C(X(w)) - F - (1+\gamma)\left[pX(w) + T\right], \tag{28}$$

[17] The expression for aggregate welfare is obtained by simply multiplying (28) with the number of hospitals, n. It can easily be verified that this does not affect the optimal waiting time.

where $\gamma > 0$ is a positive constant denoting the opportunity cost of public funds.[18] Since it is costly for the regulator to fund hospital care, we assume that the lump-sum transfer T is set such that the hospital's participation constraint is binding. Adding the (realistic) assumption that the provider also has a limited liability constraint, the transfer is set so that $pX + T = C(X) + F$. The social welfare function then simplifies to

$$W(w) = B(w) - (1+\gamma)[C(X) + F], \tag{29}$$

5.1 The socially optimal waiting time

The socially optimal waiting time is obtained by maximising welfare with respect to waiting time, yielding the following first-order condition[19]

$$\frac{\partial W}{\partial w} = \frac{\partial B(w)}{\partial w} - (1+\gamma) C'(\cdot) \frac{\partial X(w)}{\partial w} = 0, \tag{30}$$

which states that waiting time is socially optimised at a level where the utility loss to patients from a marginal increase in waiting time is equal to the corresponding reduction of treatment costs.

Using (13) and (27), and rearranging (30), we can write the expression for the socially optimal waiting time, denoted by w^s, as follows:

$$(1+\gamma) C'(X(w^s)) = \frac{-X(w^s)}{\frac{\partial X(w^s)}{\partial w}}, \tag{31}$$

[18] The altruistic component αB is not included in the welfare function as this would lead to double-counting. As argued by Chalkley and Malcomson (1998), "There is a strong case for excluding this benevolent component from social welfare on the grounds that benevolence represents a desire to do what is in the social interest and, as such, should have no role in determining what the social interest is." See also Hammond (1987) for further discussion.

[19] The second-order condition is given by

$$\frac{\partial^2 W}{\partial w^2} = -\frac{2(1-\lambda)}{t}(1+\gamma)\left[C''(\cdot)\frac{2(1-\lambda)}{t} - \frac{1}{1+\gamma}\right] < 0.$$

Thus, the supply cost function must be sufficiently convex for the condition to be fulfilled, i.e.,

$$C''(\cdot) > \frac{t}{2(1-\lambda)(1+\gamma)}.$$

where

$$X(w^s) = 2(1-\lambda)\left(\frac{v-w^s}{t}\right) + \frac{\lambda}{n}, \tag{32}$$

$$\frac{\partial X(w^s)}{\partial w} = -\frac{2(1-\lambda)}{t}. \tag{33}$$

and $w^s = w^s(v, t, \lambda, n)$.

Equation (31) defines an interior solution for the socially optimal waiting time with a partially covered L-segment, i.e., $w^s > 0$ and $x^L \in \left(0, \frac{1}{2n}\right)$. Proposition 4 below provides the exact conditions needed to support this equilibrium:

Proposition 4 *There exists a socially optimal waiting time, w^s, implicitly defined by (31), which is strictly positive and involves a partially covered L-segment, if*

$$C'\left(\frac{\lambda}{n}\right) < \frac{t\lambda}{2n(1-\lambda)(1+\gamma)}, \quad \textit{and}$$

$$C'\left(2(1-\lambda)\frac{v}{t} + \frac{\lambda}{n}\right) > \frac{v}{1+\gamma} + \frac{t\lambda}{2n(1-\lambda)(1+\gamma)}.$$

Proof. First, $x^L = 0$ implies $X(w^s) = \frac{\lambda}{n}$. It follows from (31) that $C'\left(\frac{\lambda}{n}\right) < \frac{t\lambda}{2n(1-\lambda)(1+\gamma)}$ for $x^L > 0$. Second, $x^L = \frac{1}{2n}$ implies $X(w^s) = \frac{1}{n}$. We see from (31) that $C'\left(\frac{1}{n}\right) > \frac{t}{2n(1-\lambda)(1+\gamma)}$ for $x^L < \frac{1}{2n}$. Third, $w^s = 0$ implies $X(0) = 2(1-\lambda)\frac{v}{t} + \frac{\lambda}{n}$. From (31) it is evident that $w^s > 0$ requires $C'\left(2(1-\lambda)\frac{v}{t} + \frac{\lambda}{n}\right) > \frac{v}{1+\gamma} + \frac{t\lambda}{2n(1-\lambda)(1+\gamma)}$. Finally, observe that since, by definition, $2(1-\lambda)\frac{v}{t} + \frac{\lambda}{n} \leq \frac{1}{n}$, it follows that $w^s > 0$ implies $x^L < \frac{1}{2n}$, making the condition for $x^L < \frac{1}{2n}$ redundant. ■

We see that a positive socially optimal waiting time with a partially covered L-segment requires that the cost function C is sufficiently convex. Note also, by comparing Propositions 1 and 4, that the socially optimal waiting time can always be implemented by an appropriate choice of p. This price, denoted p^*, is implicitly given by $w^*(p^*) = w^s$. Noting that $X(w^*) = X(w^s)$, this is given by

$$p^* = \frac{\frac{\partial B(w^s)}{\partial w}}{(1+\gamma)\frac{\partial X(w^s)}{\partial w}} - \alpha\frac{\frac{\partial B(w^*)}{\partial w}}{\frac{\partial X(w^*)}{\partial w}}, \tag{34}$$

or, more extensively,

$$p^* = \frac{tX(w^s)}{(1+\gamma)\,2\,(1-\lambda)} - \alpha t \frac{X\,(w^*) + \frac{\lambda}{t}\left(V - w^* - \frac{t}{2n}\right)}{(2-\lambda)}. \tag{35}$$

Intuitively, higher altruism implies a lower-powered incentive scheme.

5.2 Does competition improve social welfare?

Consider the policy choice of regulated monopolies versus competition in the hospital market. Since, for a given waiting time, the patient surplus $B\,(w)$ is unaffected by this choice of market regime, it is straightforward to see that competition is welfare neutral if the treatment price is set at the level which maximises social welfare, i.e., $p = p^*$. In this case, the effect of competition on equilibrium waiting times will be neutralised by an appropriate adjustment of p, keeping $w^* = w^s$. However, in the general case, where p is not necessarily set at the optimal level, the welfare effect of competition is characterised as follows:

Proposition 5 *Suppose that the competitive patient segment is sufficiently large: $\lambda > 1 - \frac{t}{2n(V-v)}$. Compared with regulated monopolies, hospital competition is then welfare superior if p is sufficiently high, and welfare inferior if p is sufficiently low. This result is qualitatively reversed if $\lambda < 1 - \frac{t}{2n(V-v)}$.*

Proof. Let p^* and p^m be the prices that yield, respectively, $w^* = w^s$ and $w^m = w^s$. We know (Proposition 1) that $\partial w^*/\partial p < 0$, and it is straightforward to show that this also holds under regulated monopolies, i.e., $\partial w^m/\partial p < 0$. From Proposition 2 we know that, if $\lambda > 1 - \frac{t}{2n(V-v)}$, $w^* > w^m$ for all p, implying that $p^m < p^*$. This means that, from a social welfare perspective, waiting time is too long in both regimes if $p < p^m$ and too short in both regimes if $p > p^*$. Since $w^* > w^m$ for all p, it follows that competition is welfare superior if $p > p^*$, while a market regime with regulated monopolies is welfare superior if $p < p^m$. The inverse logic applies for $\lambda < 1 - \frac{t}{2n(V-v)}$. ■

Whether or not hospital competition improves social welfare depends here on the characteristics of the reimbursement system (more specifically, the level of p) and the relative size of the competitive demand segment (λ). An increase in the treatment price always provides the hospitals with stronger incentives for competition, resulting in shorter waiting times. If, in addition, λ is sufficiently large so that $w^* > w^m$, this implies that hospital competition is welfare improving only if the hospitals are given sufficiently strong incentives to compete for patients. However, the somewhat paradoxical reason for this is that competition, in this particular case, leads to *longer* waiting times, not shorter. When the price is sufficiently high, $p > p^*$, waiting times in the competitive equilibrium are shorter than the socially optimal level. However, in a market regime with regulated monopolies, and given $\lambda > 1 - \frac{t}{2n(V-v)}$, waiting times would be even shorter, with a corresponding reduction of social welfare. Obviously, the inverse reasoning applies when $w^* < w^m$.

5.3 Hospital density and welfare

Patient surplus per hospital is given by (27). For a given waiting time, the effect of increased hospital density on patient surplus at the hospital level is given by

$$\frac{\partial B(w,n)}{\partial n} = -\frac{\lambda}{n^2}\left(V - w - \frac{t}{2n}\right) < 0. \tag{36}$$

In the H-segment, an additional hospital implies a lower demand for each hospital and therefore a lower surplus. The corresponding effect of increased hospital density on *aggregate patient surplus* is given by

$$\frac{\partial (nB(w,n))}{\partial n} = n\frac{\partial B(w,n)}{\partial n} + B(w,n) = \frac{\lambda t}{4n^2} + \frac{1-\lambda}{t}(v-w)^2 > 0. \tag{37}$$

For given waiting times, an additional hospital increases total surplus for two reasons. For the H-type, an additional hospital reduces transportation costs (although the demand of H-type patients stays fixed). In addition, there is increased demand from L-type patients.

Aggregating (29), total welfare is

$$\Psi(w,n) = nB(w,n) - (1+\gamma)n[C(X(w,n)) + F]. \tag{38}$$

Assuming that waiting is set at the optimal level, given by (31), the effect on welfare from an additional hospital is

$$\begin{aligned} \frac{\partial \Psi(w^s,n)}{\partial n} &= B + n\frac{\partial B}{\partial n} - (1+\gamma)[C(X)+F] - (1+\gamma)nC'(\cdot)\frac{\partial X}{\partial n} \\ &= W + n\left[\frac{\partial B}{\partial n} - (1+\gamma)C'(\cdot)\frac{\partial X}{\partial n}\right]. \end{aligned} \tag{39}$$

In the last equation, the first term is the additional welfare from an additional hospital. The second term is negative and shows the reduction of patient's surplus in each hospital (since the number of high-type patients is fixed, an extra hospital implies less demand for each hospital). The third term is positive and gives the reduction in the marginal cost from having a lower demand of patients. Notice that, since $w^*(p^*) = w^s$, the first-best solution with respect to optimal hospital density can be implemented under both competition and regulated monopolies, by setting prices at the optimal level.

Suppose that p is exogenous and generally different from p^*. What is the effect of an increase in n on welfare? Inserting the equilibrium level of waiting, $w^*(n)$, into (38), the effect of an additional hospital is given by

$$\frac{\partial \Psi(w^*(n),n)}{\partial n} = W + n\left[\frac{\partial B}{\partial n} - (1+\gamma)C'(\cdot)\frac{\partial X}{\partial n}\right] + n\left[\frac{\partial B}{\partial w} - (1+\gamma)C'(\cdot)\frac{\partial X}{\partial w}\right]\frac{\partial w^*}{\partial n}. \tag{40}$$

Compared with first-best solution, the last term takes into account the effect of n on welfare arising from the equilibrium response in waiting time (this effect is non-zero for $p \neq p^*$). Since $\frac{\partial w^*}{\partial n} < 0$, the last term in (40) is positive if $\frac{\partial B}{\partial w} < (1+\gamma)C'(\cdot)\frac{\partial X}{\partial w}$, and vice versa. Intuitively, if waiting time is too high, then an increase in n will bring the waiting time down towards the optimal level.

6 Conclusions

This study has analysed the impact of hospital competition on waiting times, using a Salop-type model. Our main result is that, compared with a benchmark case of regulated monopolies, hospital competition reduces waiting times only if the competitive patient segment is sufficiently small. Otherwise, if free choice is relevant for a sufficiently large share of the total patient mass (i.e., if the competitive segment is sufficiently high), then competition *increases* waiting times. Therefore we suggest that policies that introduce choice and competition in health care markets may not be as successful as policymakers might expect.

We also find that policies aimed at reducing travelling costs (like reimbursing travel expenses for patients choosing to receive treatment in hospitals outside their catchment area) may surprisingly increase waiting times and reduce overall activity.

According to our analysis, policies aimed at increasing the number of hospital will have the expected effect of reducing waiting times and increasing activity.

Finally, policies that remunerate hospitals according to activity-based funding rules (like Payment by Results in England or DRG pricing in other European countries) combined with hospital competition are socially preferrable to regulated monopolies if the competitive demand segment and the (regulated) treatment price are both either sufficiently large/high or sufficiently small/low.

References

[1] Barros, P., Olivella, P., 2005. "Waiting lists and patient selection", *Journal of Economics and Management Strategy,* 14 (3), 623–646.

[2] Cullis, P., J.G. Jones and C. Propper, 2000, "Waiting and Medical Treatment: Analyses and Policies", Chapter 28 in A. J. Culyer and J. P. Newhouse (eds), *Handbook on Health Economics*, Amsterdam: Elsevier.

[3] Chalkley M., Malcomson J.M.,1998b, "Contracting for Health Services when patient demand does not reflect quality", *Journal of Health Economics*, 17, 1-19.

[4] Dawson, D., Gravelle, H., Jacobs, R., Martin, S., Smith, P.C., 2006, "The effects of expanding patient choice of provider on waiting times: evidence from a policy experiment", *Health Economics*, forthcoming.

[5] Farnworth, M.G., 2003, "A game theoretic model of the relationship between prices and waiting times", *Journal of Health Economics*, 22(1), 47-60.

[6] Ferguson, B., Sheldon, T., Posnett, J. (eds.), 1999. "Concentration and choice in health care", London: Royal Society of Medicine.

[7] Folland, S., Goodman, A.C., Stano, M., 2004. "The economics of health and health care", Upper Saddle River, NJ: Prentice Hall.

[8] Gravelle, H., P.C. Smith and A. Xavier, 2003, "Performance signals in the public sector: the case of health care", *Oxford Economic Papers*, 55, 81-103.

[9] Hammond, P., 1987. Altruism. In: Eatwell, J., Milgate, M., Newman, P. (Eds.), The New Palgrave: A Dictionary of Economics. Macmillan, London, 85-87.

[10] Hoel, M., Sæther, E.M., 2003, "Public health care with waiting time: the role of supplementary private health care", *Journal of Health Economics,* 22, 599–616.

[11] Iversen, T., 1993, "A theory of hospital waiting lists", *Journal of Health Economics,* 12, 55-71.

[12] Iversen, T., 1997, "The effect of private sector on the waiting time in a National Health Service", *Journal of Health Economics,* 16, 381-396.

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

[14] Olivella, P., 2002, "Shifting public-health-sector waiting lists to the private sector", *European Journal of Political Economy*, 19 (1), 103–132.

[15] Martin, S., and P.C. Smith, 1999, "Rationing by waiting lists: an empirical investigation", *Journal of Public Economics*, 71, 141-64.

[16] Siciliani, L., 2005, "Does more choice reduce waiting times?", *Health Economics*, 14(1), 17-23.

[17] Siciliani, L. and J. Hurst, 2004, "Explaining waiting times variations for elective surgery across OECD countries", *OECD Economic Studies*, 38(1), 1-23.

[18] Siciliani, L. and J. Hurst, 2005, "Tackling excessive waiting times for elective surgery: a comparison of policies in twelve OECD countries", *Health Policy,* 72(2), 201-215 .

[19] Siciliani, L. and S. Martin, 2007, "An empirical analysis of the impact of choice on waiting times", 2006, *Health Economics*, forthcoming.

[20] 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(1), 25–51.