

**The payment of hospital services:
a waiting lists model.**

ALESSANDRA FERRARI

Department of Economics, University of Reading, Reading RG6 6AA.

Abstract

This paper analyses the incentive properties of prospective payment systems for hospital contracts, a key feature in many health systems' reforms. Building on current literature, the model explicitly allows for the existence of waiting time, modelled as adversely affecting patients' utility and therefore reducing social welfare. The model shows that rewarding hospitals for their demand leads to the first best solution, identified with respect to the relevant quality and quantity variables. The additional separate payment of a price per case is instead required when the social cost of waiting is introduced alongside the private costs.

JEL classification: I11, L11.

1. Introduction.

The last two decades have witnessed major changes in the organisation of health services. Various factors, among which the aging of the population and technical progress, led to growing demand levels and expectations; then to discontent, cost increases and general financial crisis. In general terms this scenario has characterised

(and still characterises) many European countries: thus the need to reform the system. Unlike the US, that heavily relies upon private insurance, most European countries share strong traditional equity aims that descend from the consideration of health as a social good; as a consequence European reforms typically did not affect the general finance of health services, based on income tax or social insurance, but the mechanisms that regulate their provision (Saltman *et al.* 2002).

As regards the provision of hospital services, almost invariably the main change consisted in the introduction of prospective payment systems (PPS) that pay a fixed price per unit of output (price per case), in place of the traditional cost-reimbursement rules based on historical costing. By making hospitals the residual claimants for their costs PPS give strong incentives to economic efficiency, although this could be at the expenses of quality levels. The US introduced this system in 1983, along with a system of DRGs (diagnostic related groups) to classify cases and their costs and therefore define their prices. The UK was one of the first European countries to reform their health system, with the White Paper *Working for Patients* in 1990 that introduced an internal market for hospital services. Even though the “quasi-market” system was subsequently eliminated, the main changes introduced by the reform remained in place: the clear separation of purchasers (demand) and providers (supply) of hospital services, the independent hospital trusts, the use of contracts, or now “agreements”, for the payment of their services via PPS. Other countries have then followed, and are following, along very similar lines, and the debate about different payment systems for hospital services is then on top of their political agendas.

This paper contributes to the literature on the properties of PPS by explicitly allowing for the existence of waiting lists. Building on a model by Chalkley and Malcomson

(1998a) waiting time is introduced as a measure of hospitals' demand taking into consideration the adverse effects that it has on patients utility, and therefore on social welfare. The structure of the paper is as follows. Section 2 identifies the framework of analysis and reviews the existing literature on the topic. Section 3 analyses more in detail the work of C&M: this is the basis for the waiting lists model developed in Section 3. Section 4 relaxes one of the assumptions of the model and analyses the case in which private and social costs are different. Conclusions are in Section 5. Appendices 1, 2 and 3 discuss different extensions to the main model.

2. The literature on PPS.

There is a very large literature on the incentive properties of different payment systems to hospitals, mainly but not exclusively from the US (good reviews of the main issues of contracting in the NHS are in Barker et al., 1996, and in Chalkley and Malcomson, 1996a,b, 1998c). Its main characteristic is that of dealing with a multitask agency problem (Holmstrom and Milgrom, 1991) that is expressed in at least two potentially conflicting objectives: reducing the costs of the service without reducing the quality level. Cost reimbursement rules, that pay the hospitals on the basis of the costs they actually incurred, lack the incentives for cost reduction. On the other hand prospective payment systems make the hospital the residual claimant for its costs giving powerful incentives to reduce them (Laffont and Tirole, 1993). However, this can lead to too low a quality level, as quality is costly but usually unmeasurable and unenforceable by contract. This is the key problem discussed by the literature. The superiority of PPS is acknowledged in all cases, although more complex forms of payment can be devised depending on the particular assumptions

made, or characteristics considered: whether demand comes from privately insured patients, the degree of asymmetric information between purchasers and providers (about their costs and/or the quality level of the service), the heterogeneity of providers or patients, and so on.

A good general set up of the problem is given by Chalkley and Malcomson (1998a,b). The model describes the contractual relationship between a self-interested provider, that maximises its financial surplus, and a purchaser that maximises a social welfare function that depends on the number of cases treated and on the quality level of the treatment. The original demand for treatment comes from an exogenously given number of patients but the actual, final demand comes from the purchaser who buys hospital services for them. The purchaser is assumed to know the costs and to correctly perceive the quality level of hospital services. The model shows that a PPS that rewards hospitals on a price per case basis is not optimal, unless the system is demand constrained, because it leads to the treatment of too many patients; the first best solution is achieved by rewarding hospitals separately for the people they treat and for those that ask for treatment. The result holds true under different capacity constraints as well as when assuming different quality dimensions. If instead buyers do not infer quality correctly (1998b) only a second best solution can be achieved, by means of a mixed system that combines the fixed price per case to a partial cost reimbursement. Block contracts (simpler agreements that paid the hospitals a predefined amount of money in exchange for a total amount of services, very common in the first years of the UK reform) can be used only when assuming, following Newhouse (1970), that hospitals are not profit-maximising units but have ethical concerns, and only if the system is demand-constrained.

More or less complicated prospective payment systems, that allow for different prices or for a cost reimbursement element, are also the solutions to the models of Ellis and McGuire (1986, 1991), Ma (1994), Rogerson (1994), Ellis (1998) and De Fraja (2000).

An issue that has not been given much consideration in the analysis of contracts is that of waiting time. In the contracting context, the existence of waiting lists has been usually considered as an indirect measure of quality, often to explain the demand for private insurance (Besley *et al.*, 1999; Propper 2000). As mentioned in the introduction, the aim of this paper is to contribute to the existing literature by extending the model of Chalkley and Malcomson (1998a) (C&M from now on) to the case in which an explicit role is given to the existence of waiting time; this will be considered as a measure of hospital demand, taking into consideration the adverse effects that it has on patients utility, and therefore on social welfare.

3. The C&M model.

The situation modelled by C&M is the definition of a contract between a purchaser of hospital services and a self-interested hospital. The purchaser buys the services for the patients, and it maximises a social welfare function. The hospital maximises a utility function that is increasing in its financial surplus. The contracting system and the determination of the purchaser's budget are taken as given, and no ethical or professional considerations enter the hospital's utility function.

The logic and the structure of the model can be summarised as follows. The two objective functions are identified and maximised with respect to the relevant

variables, so that two sets of first order conditions (F.O.C.) are identified. As the hospital's utility function depends on the (kind of) payment it receives, the point is to choose a payment system that equates the two sets of F.O.C., so that the hospital makes a choice that maximises social welfare. This result is defined as a first best solution.

More in detail, define $b(x,q)$ as the benefit perceived by the purchaser of treating x patients with quality level q . This is increasing in both variables and strictly concave in q .

$C(x,q,f)$ is the cost for the hospital and is assumed to be increasing, convex in x and strictly convex in q ; f is the effort in cost reduction. The cost is assumed to be known to the purchaser, whereas q cannot be monitored and so cannot be enforced by contract.

Utility for the hospital is given by

$$U = s - v(x,q,f)$$

where s is the financial surplus and v a disutility component. Reservation utility for the hospital is \bar{v} .

The social welfare function can be written as

$$(1) \quad W = [b(x,q) - s - C(x,q,f)] + [s - v(x,q,f)] - \alpha[C(x,q,f) + s]$$

where $\alpha > 0$ is to allow for distortions from raising revenue from taxation. The purchaser is maximising (1) subject to the constraint

$$(2) \quad s - v(x,q,f) \geq \bar{v}$$

Rearranging (1) and observing that it is a decreasing function of s , so that (2) will always hold as an equality, it can be rewritten as

$$(3) \quad W = b(x, q) - (1 + \alpha)[C(x, q, f) + v(x, q, f)] - \alpha v$$

The hospital objective function is to maximise

$$(4) \quad H = s - v(x, q, f) \quad \text{s.t.} \quad B(x, y, C) - C(x, q, f) - s = 0$$

$B(\cdot)$ is the total payment received by the hospital and it can be made dependent on the number of cases x , on the demand $y(q)$, which is assumed to be increasing in quality, or on costs depending on the contract chosen, but it cannot be directly a function of q . So for example with a simple price per case contract $B = px$; with cost reimbursement $B = C(x, q, f)$; with partial cost sharing $B = px + \phi C$, and so on. Thus hospital revenues can be affected by quality only via the effect that this has on demand.

The model consists of maximising (3) and (4) with respect to x , q and f under the further constraint that

$$(5) \quad x \leq \bar{x}(q)$$

where $\bar{x}(q)$ is the maximum number of patients that can be treated by a hospital and is given by

$$\bar{x}(q) = \min[y(q), k]$$

that is the minimum value of demand $y(q)$ and the capacity (maximum possible supply) of the hospital, k .

If (3) subject to (5) is maximised at x^* , q^* and f^* , the problem is to find the payment system $B(\cdot)$ that maximises (4) subject to (5) at exactly the same values. Mathematically, this is done by substituting the solutions to the purchaser's F.O.C. in the hospitals' F.O.C. A first best solution exists if there is a payment system that

guarantees that the equations hold. This first best solution means that hospitals will choose to treat the optimal number of patients, at the optimal quality level.

The following results are obtained by C&M.

1) For the case of a demand constrained system, that is when

$$x^* = y(q^*) < k$$

the payment of a lump-sum transfer T (required to ensure that (2) always holds as an equality) and the use of a price per case contract of the form $B = px$ leads to the desired result if the price $p = \partial B / \partial x = B_x^1$ is set as

$$(6) \quad B_x = b_x - \alpha(C_x + v_x) + \frac{b_q - \alpha(C_q + v_q)}{y_q}$$

evaluated at the socially optimal values. That is, (6) reflects the marginal benefit of treating an additional patient and the marginal benefit of quality, calculated from the marginal increase in demand, all net of the cost of tax distortion.

2) If the system is unconstrained or capacity constrained, that is respectively

$$x^* \leq \bar{x}(q^*)$$

and

$$x^* = k < y(q^*)$$

then the optimal pricing rule will be given by

$$(7) \quad \begin{aligned} B_x &= b_x - \alpha(C_x + v_x) \\ B_y &= \frac{b_q - \alpha(C_q + v_q)}{y_q} \end{aligned}$$

¹ All variables with a subscript represent a derivative

Two different prices are in (7), a price per case B_x and a price per patient demanding treatment B_y . The optimal price per case equals the marginal benefit at the number of cases treated; the optimal price per patient demanding treatment reflects the marginal benefit of quality for the marginal patient; both prices are net of the marginal cost of tax distortion.

The rationale of the results is the following. A fixed price per case set as in (6) leads to the efficient outcome. Hospitals increase the quality level up to q^* in order to attract patients, and treat all the patients demanding treatment at that quality level. This works only if the system is demand constrained. In the case of an unconstrained system, the optimal number of cases to treat at the optimal quality level is lower than the number of people demanding treatment. If hospitals are rewarded for the number of people they treat, they will have an incentive to treat too many. Rewarding them separately for the number of treatments and for the number of referrals for future treatment (i.e. for their demand) solves the problem. A similar argument, but for opposite reasons, applies in the case of a capacity constraint. If the hospital cannot treat more than k patients it doesn't need to increase its level of quality up to q^* in order to attract them, and will therefore choose too low a quality level. Finally, a payment system like (7) works efficiently in all cases, i.e. whether the system is constrained or unconstrained.

In all cases the optimal setting of the price makes hospitals choose f^* . This comes from the fact that being the residual claimants of their costs gives them the incentives to cost minimisation.

4. The waiting lists model.

For a competitive market and with convex costs, C&M show that a linear pricing system can lead to the first best solution: hospitals have an incentive not to reduce quality levels in order to attract patients, which are their source of income. Quality does not need to be monitored or specified in the contract; all that is required is observability of demand and that this is responsive to quality. A reputation effect reinforces the argument in the case of short term contracts.

The assumption of demand observability is strictly linked with the discussion of the role of waiting lists, which are considered below. The assumption of demand's responsiveness to quality is less critical when demand does not come directly from patients but is filtered by medical professionals, who are in a better position to collect and process information about the quality of treatment provided by different hospitals. This is equivalent to considering health care as a search good and not an experience good (as it is for the single patient), thus avoiding much of the informational problems that are typical of the sector.

The main change brought to the C&M model is the introduction of waiting lists. These are considered as a measure of patients' demand, and thus an indirect measure of quality, taking into consideration that they negatively affect patients' utility. The model is constructed as follows.

If patients get treatment of quality q , their utility from this treatment can be defined as

$$(8) \quad U = q$$

If treatment is postponed by w periods, where w is the time length of the waiting list, then (8) must allow for a discounting factor reflecting time preference, and for a factor reflecting the risk of a worsening of patients conditions while waiting.

Define r the time preference factor and h the conditional probability of “dying” at time t , having survived until then², the expected utility of a patient is

$$(9) \quad U = qe^{-w(r+h)}$$

Now define g as the number of people in front of a patient in the queue; its change over time will be given by

$$(10) \quad \frac{dg}{dt} = -x - hg$$

where x is the number of cases treated. Considering a steady state situation in which x (and later y , the demand) is constant, then (10) can be solved to obtain

$$(11) \quad g(t) = \left(G + \frac{x}{h} \right) e^{-ht} - \frac{x}{h}$$

Equation (11) means that the number of people in front of a person in the queue is a function of G (the number of people at the beginning, when $t = 0$) and it decreases with time, partly because they are treated partly because they die. When $t = w$ then $g(t) = 0$ and (11) can be solved with respect to w , i.e.

$$(12) \quad w = \left(\ln \frac{Gh + x}{x} \right) \frac{1}{h}$$

Moreover, the number of people in a queue changes over time as

² In particular, $F(t)$ is the probability distribution function of dying at any point in time; $e^{-ht} = 1 - F(t)$ is the probability of surviving t periods, and so $h = \frac{dF(t)/dt}{1 - F(t)}$ is the conditional probability.

$$(13) \quad \frac{dG}{dt} = y - x - hG$$

where y is the number of people joining a queue, or each hospital's demand. In steady state, when G is constant then

$$(14) \quad G = \frac{y - x}{h}$$

Substituting (14) into (12) and then into (9) finally gives

$$(15) \quad U_i = q_i \left(\frac{x_i}{y_i} \right)^{\frac{r+h}{h}}$$

Equation (15) is an expression of the expected utility of joining a queue at hospital i , and in equilibrium this value will be the same for every hospital.

In the social welfare function the benefit is now the expected utility of a cohort of patients asking for treatment at all hospitals at time 0; hospitals' costs are discounted by a factor $e^{-rw} = (x/y)^{r/h}$ as they will be incurred only after w periods. The same discount rate r is assumed for both patients and hospitals, but these assumptions will be relaxed in the next section.

The social welfare function can be now written as

$$(16) \quad W = \sum_i \left\{ y_i q_i \left(\frac{x_i}{y_i} \right)^{\frac{r+h}{h}} - (1 + \alpha) [C_i(x, q, f) + v_i(x, q, f)] \left(\frac{x_i}{y_i} \right)^{\frac{r}{h}} - \alpha \bar{v} \right\}$$

In this case \bar{v} is the present value of the hospitals' reservation utility.

On the hospitals' side, the objective function is like in C&M but with the cost-discounting factor. The demand function $y(x,q)$ for each hospital is derived from (15) as

$$(17) \quad y_i = \left(\frac{q_i}{U} \right)^{\frac{h}{r+h}} x_i$$

where U is the level of utility available at other hospitals.

Substituting (17) into the discount factor formula finally gives

$$(18) \quad H = B(x, y(x, q)) - [C(x, q, f) + v(x, q, f)] \left(\frac{U}{q} \right)^{\frac{r}{r+h}}$$

which is the hospitals' objective function. Following on C&M results, in (18) the payment system $B(\cdot)$ has been directly specified as a function of the number of cases treated and the volume of demand (the proof of the non-optimality of other payment schemes is redundant and therefore not included).

The simultaneous maximisation of (16) and (18) with respect to x , q and f leads to the following system of F.O.C.

$$(19) \quad W_x = \left[\frac{r+h}{h} q - (1+\alpha)(C_x + v_x) - \frac{r}{hx} (1+\alpha)(C + v) \right] \left(\frac{x}{y} \right)^{\frac{r}{h}} = 0$$

$$(20) \quad W_q = \left[x - (1+\alpha)(C_q + v_q) \right] \left(\frac{x}{y} \right)^{\frac{r}{h}} = 0$$

$$(21) \quad W_f = -(1+\alpha)(C_f + v_f) \left(\frac{x}{y} \right)^{\frac{r}{h}} = 0$$

$$(22) \quad H_x = B_x + \left(\frac{U}{q} \right)^{\frac{r}{r+h}} \left(\frac{q}{U} B_y - C_x - v_x \right) = 0$$

$$(23) \quad H_q = \left[B_y \frac{hx}{(r+h)U} - (C_q + v_q) + \frac{r(C+v)}{(r+h)q} \right] \left(\frac{U}{q} \right)^{\frac{r}{r+h}} = 0$$

$$(24) \quad H_f = -(C_f + v_f) \left(\frac{U}{q} \right)^{\frac{r}{r+h}} = 0$$

To proceed (19) is equated with (22), (20) with (23) and (21) with (24).

First of all, (21) and (24) will always be the same, that is hospitals have an incentive to maximise the effort in cost reduction, i.e. with a prospective payment system they will always choose f^* . As regards the choice of x and q , solving with respect to the price variables leads to

$$(25) \quad \begin{aligned} B_x &= 0 \\ B_y &= \frac{U}{1+\alpha} \frac{r+h}{h} - \frac{r(C+v)U}{hqx} \end{aligned}$$

Rearranging the equations, another way to express the result for B_y is

$$(26) \quad B_y = \frac{U_q y}{y_q} - \frac{e^{-rv} \left[\alpha(C_q + v_q) + \frac{C+v}{q} \frac{r}{r+h} \right]}{y_q}$$

Equation (26) is the same price per patient demanding treatment as in C&M (see equation (7)) except for the presence of the discount factor and its sensitivity to quality changes. As could be expected, the optimal price, and hence the marginal benefit for the hospital, equals the marginal social benefit, net of a portion α of the marginal cost. This result indicates that it is optimal to pay the hospital a fixed sum T and then a fixed price for every person asking for treatment by entering its waiting list, and a zero price for the number of people actually treated.

This apparently counterintuitive result stems from the way demand was modelled. On the one hand, hospitals have an incentive to increase the quality level of their service: they are paid on the basis of their demand and, given the assumptions, a higher quality attracts more people. On the other hand, however, more people demanding treatment translate into a longer waiting time, and this in turn acts negatively on demand. In order to counteract this effect hospitals will decide the number of people they want to treat, as the more patients are treated the shorter is the waiting time. In other words the adjustment towards equilibrium can be described as an adjustment towards an optimal waiting time; rewarding hospitals on the basis of their demand gives them immediate and direct incentives over q and also indirect incentives over x . Also, in the presence of waiting time targets it could counterbalance the potential incentive to find ways to exclude (new or current) patients from the queue, giving them the funding in advance to treat them. As shown more in detail in Appendix 1, if only a price per case were paid only a second best solution could be achieved, in which a lower quality level is traded off with a higher number of treatments. For example hospitals could reduce the length of stay and move towards a “quick” treatment of patients on a day basis, not a totally unfamiliar scenario in reformed systems.

The first best solution can be achieved also when allowing for hospitals and patients to have different discount factors for time preference, and when modelling the market as an oligopoly; in both cases the only difference is the price level, which is higher the more patients dislike being in a queue, or the fewer providers are on the market. These two cases are discussed respectively in Appendices 2 and 3.

A different result is instead obtained when relaxing the implicit assumption that social and private costs are the same and the “cost of the loss” of patients getting worse or

dying while waiting is introduced in the social costs. This is explored in the next section.

4. The social “cost of the loss”.

So far, the only costs that have been considered are the “production” costs of the hospital and its disutility. It is realistic however to consider that, from a social welfare point of view, the fact that people might get worse while they wait and eventually die has a cost, other than their own cost from waiting. In other words, not only people, but also society, are worse off the longer they have to wait; the events of getting worse and eventually dying carry a cost, that includes rather unmeasurable as well as more measurable elements (for example the payment of benefits to people when they cannot work). Call this “cost of the loss” γ . This has to be introduced in the social welfare function that the purchaser is assumed to be maximising. This means that social and private costs are now different.

The cumulative distribution of the probability of dying at time t is

$$F(t) = 1 - e^{-ht}$$

If γ is the social cost of the loss, the expected social cost today of having y persons getting worse between now and w is given by

$$(27) \quad yh\gamma \int_0^w e^{-(r+h)t} dt$$

From (27) the value of the social cost of the loss is

$$(28) \quad y\gamma \frac{h}{r+h} \left[1 - \left(\frac{x}{y} \right)^{\frac{r+h}{h}} \right]$$

A new expression for the welfare function can now be rewritten as

$$(29) \quad W = \sum_i \left\{ y_i q_i \left(\frac{x_i}{y_i} \right)^{\frac{r+h}{h}} - (1+\alpha) \left[(C_i + v_i) \left(\frac{x_i}{y_i} \right)^{\frac{r}{h}} - y_i \gamma \frac{h}{r+h} \left(1 - \left(\frac{x}{y} \right)^{\frac{r+h}{h}} \right) \right] \right\}$$

Applying the same procedure as in all previous cases, the result is now different, i.e.

$$(30) \quad \begin{aligned} B_x &= \gamma \\ B_y &= \frac{U}{1+\alpha} \frac{r+h}{h} - \frac{r(C+v)U}{hqx} \end{aligned}$$

The value of the price per person asking for treatment (B_y) is the same as in (25), but in this case a positive price per case (B_x) is required, which is equal to the social cost of the loss itself.

To interpret this result it is necessary to analyse what effect the introduction of γ has had on the equilibrium. To do so, let's consider the following equations from the F.O.C. for welfare maximisation:

$$(31) \quad W_x = \frac{r+h}{h} q - (1+\alpha)(C_x + v_x) - (1+\alpha) \left(\frac{r}{h} \frac{C+v}{x} - \gamma \right) = 0$$

$$(32) \quad W_q = x - (1+\alpha)(C_q + v_q) = 0$$

The effect that the introduction of γ has had on the optimal values of x and q can be calculated by totally differentiating (31) and (32), which results in

$$(33) \quad \begin{aligned} & \left[-(1+\alpha)(C_{xx} + v_{xx}) - (1+\alpha) \frac{r}{h} \left(\frac{(C_x + v_x)x - (C+v)}{x^2} \right) \right] dx + \\ & + \left[\frac{r+h}{h} - (1+\alpha)(C_{xq} + v_{xq}) - (1+\alpha) \frac{r}{h} \frac{(C_q + v_q)}{x} \right] dq = -(1+\alpha)d\gamma \end{aligned}$$

$$(34) \quad [1 - (1+\alpha)(C_{xq} + v_{xq})] dx + [-(1+\alpha)(C_{qq} + v_{qq})] dq = 0$$

In matrix notation this can be rewritten as

$$(35) \quad H \begin{pmatrix} dx \\ dq \end{pmatrix} = \begin{bmatrix} -(1+\alpha)d\gamma \\ 0 \end{bmatrix}$$

where H is the Hessian matrix of the second order derivatives of the welfare function.

From (35) finally get

$$(36) \quad \frac{dx}{d\gamma} = \frac{(1+\alpha)^2 (C_{qq} + v_{qq})}{|H|}$$

$$(37) \quad \frac{dq}{d\gamma} = \frac{(1+\alpha)[1-(1+\alpha)(C_{xq} + v_{xq})]}{|H|}$$

From the second order conditions the determinant of the Hessian matrix is known to be positive, so that

$$\frac{dx}{d\gamma} > 0 \quad \text{always}$$

$$\frac{dq}{d\gamma} > 0 \quad \text{if} \quad (C_{xq} + v_{xq}) < \frac{1}{1+\alpha}$$

$$\frac{dq}{d\gamma} < 0 \quad \text{if} \quad (C_{xq} + v_{xq}) > \frac{1}{1+\alpha}$$

This means that the introduction of γ always produces an increase in the number of patients to treat, whereas the effect on the quality level depends on the size of the cross derivative of the hospital cost and disutility function. To shed more light on the above, consider the cross derivative W_{qx} as calculated from (32); it can be seen that when

$$(38) \quad (C_{xq} + v_{xq}) < \frac{1}{1+\alpha}$$

then $W_{xq} > 0$, that is the marginal benefit of quality increases with the number of cases and the sign of $\frac{dq}{dx}$ (obtained via the total differentiation of (32)) is positive. The opposite holds in the case

$$(39) \quad (C_{xq} + v_{xq}) > \frac{1}{1 + \alpha}$$

This means that if (38) is true then an increase in the number of cases makes an improvement in quality more valuable to society. The increase in the optimal number of cases brought about by γ will therefore lead to a higher level of quality. The opposite holds in the case when (39) is true. Another way of looking at this is via the behaviour of the average cost function. From the convexity assumptions of the total cost function made at the beginning, the average cost per case $(C+v)/x$ is an increasing function of both x and q , that is the average cost of treating a patient gets higher the more patients are treated and/or the higher is the quality level.

The implication of (38) is that the average cost will tend to increase with quality less when the hospital treats more people. If instead (39) is true then the increase in the average cost per case will be higher the higher is the number of patients treated. In the first case, a social welfare function that includes the cost of the loss will be maximised by increasing both quantity and quality; in the second case, the increase in the optimal number of cases will lower the socially optimal quality level.

Even if this last hypothesis might seem more reasonable to expect, the opposite cannot be ruled out a priori. This could be the case for example if in order to increase the number of patients to treat the hospital has to buy and use some particular

equipment, or hire some specialised staff, and this has beneficial effects on the quality level of treatment that can be increased at a lower extra cost.

In conclusion, the social cost of the loss introduces a difference between the social and the private cost functions, and the optimal number of cases is increased. As a consequence, a positive price per case is now required to give hospitals correct incentives. The optimal price per entry in the waiting list is determined in the same way as before, and its value will depend on whether the optimal quality level has increased or decreased.

5. Conclusions.

This paper has analysed the structure of a prospective payment system to hospitals taking explicitly into account the existence of waiting lists. Building on a paper by C&M, waiting time has been modelled as adversely affecting patients' (and thus the social welfare maximising purchaser's) utility. The model shows that rewarding hospitals for their demand, as opposed to their output, leads to a first best solution. The additional payment of a price per case is instead required when the social cost of waiting is introduced alongside the private costs. In both cases the payment of a positive price for every person entering the queue is essential: the equilibrium is reached via the direct incentive that this gives towards quality, in order to attract demand. The incentive to treat the socially optimal number of patients comes from the need to reduce the average waiting time that arises from a higher demand but at the same time negatively affects it.

As will be detailed in Appendix 1, the payment of a price per case leads instead to a second best solution, in which a lower quality could be traded off with a higher quantity. Hospitals could for example reduce the average length of stay (lower quality

level) and/or move towards the treatment of patients on a day-basis, which could in turn increase the probability of being readmitted, and therefore possibly increase the average waiting time. This scenario is actually not totally unfamiliar to the reformed English NHS with the additional risk of hospitals hiding true information about the real length of their waiting lists when specific targets about these are imposed. The payment of a price per demand might reduce the extent of this last problem but its proper discussion goes beyond the aim of this paper, as it involves other issues, such as the determination of the purchaser's budget, or the quality of their information. The paper is subject to other limitations, among which the assumption of perfect quality observability by purchasers, the modelling of quality as a one-dimensional variable, or the exclusion of the probability of readmission; all these are acknowledged and the extensions left to future research.

Acknowledgments

I would like to thank my PHD supervisor Prof. Norman Ireland, whose guidance and advice have assisted considerably in the direction and completion of this work, during and after my PhD. Special thanks also go to my examiners Prof. John Cubbin and Prof. Tom Weyman-Jones, and to Prof. Pedro Pita-Barros for his useful comments as discussant of an earlier version of this work.

Appendix 1

Payment of a positive price per case and a zero price per demand.

This appendix analyses the case in which the hospital is paid a positive price per case, B_x , and no price per person entering their queue, i.e. no price per demand B_y .

The F.O.C from the maximisation of (16) and (18) become

$$(A.1) \quad W_x = \left[\frac{r+h}{h}q - (1+\alpha)(C_x + v_x) - \frac{r}{hx}(1+\alpha)(C+v) \right] \left(\frac{x}{y} \right)^{\frac{r}{h}} = 0$$

$$(A.2) \quad W_q = \left[x - (1+\alpha)(C_q + v_q) \right] \left(\frac{x}{y} \right)^{\frac{r}{h}} = 0$$

$$(A.3) \quad W_f = -(1+\alpha)(C_f + v_f) \left(\frac{x}{y} \right)^{\frac{r}{h}} = 0$$

$$(A.4) \quad H_x = B_x - \left(\frac{U}{q} \right)^{\frac{r}{r+h}} (C_x + v_x) = 0$$

$$(A.5) \quad H_q = - \left(\frac{U}{q} \right)^{\frac{r}{r+h}} \left[(C_q + v_q) - \frac{r(C+v)}{(r+h)q} \right] = 0$$

$$(A.6) \quad H_f = -(C_f + v_f) \left(\frac{U}{q} \right)^{\frac{r}{r+h}} = 0$$

Equations (A.3) and (A.6) are always the same, so the hospital is putting the required level of effort in cost reduction.

Equation (A.1) implies that

$$(C_x + v_x) = \frac{r+h}{h} \frac{q}{1+\alpha} - \frac{r(C+v)}{hx}$$

and equation (A.4) implies that

$$B_x = \left(\frac{U}{q}\right)^{\frac{r}{r+h}} (C_x + v_x)$$

One can therefore set

$$B_x = \left(\frac{U}{q}\right)^{\frac{r}{r+h}} \left(\frac{r+h}{h} \frac{q}{1+\alpha} - \frac{r(C+v)}{hx}\right)$$

to ensure that hospitals choose the optimal value of $(C_x + v_x)$. This does not guarantee that they will also choose $x=x^*$, or that indeed it is optimal that they do so. In fact, now no instrument exists to give hospitals direct incentives to the right level of quality, as the comparison of (A.2) and (A.5) shows. As the cost function is strictly convex in q , hospitals will therefore choose $q^H < q^*$. Having no direct control over it, the purchaser will take this lower level of quality as given. The effect that this sub-optimal quality level has on the welfare function and on the final equilibrium can be seen by totally differentiating (A.1) that gives

$$dW_x = W_{xq}dq + W_{xx}dx$$

$$\text{so that } dx/dq = -W_{xq}/W_{xx}$$

Concavity of the welfare function means that $W_{xx} < 0$. For symmetry in cross derivatives from (A.2) one gets

$$W_{xq} = 1 - (1+\alpha) C_{xq}$$

which will be positive or negative depending on the sign and size of C_{xq} . If $C_{xq} > 0$ (as one might expect) and $C_{xq} > 1/(1+\alpha)$ then $W_{xq} < 0$, which implies $dx/dq < 0$. In other words, the reduction in the quality level can lead to an increase in the optimal quantity level, i.e. the one that re-maximises the social welfare function.

This second best solution can now be achieved by setting the (new) optimal price per case as usual.

Appendix 2

Different discount factor.

If people's discount factor for time preference is different from that of hospitals, call the former δ and the latter r , and the objective functions can be re-written as

$$(A.7) \quad W = \sum_i \left[y_i q_i \left(\frac{x_i}{y_i} \right)^{\frac{\delta+h}{h}} - (C+v) \left(\frac{x_i}{y_i} \right)^{\frac{r}{h}} (1+\alpha) \right]$$

$$(A.8) \quad H = B(x,y) - (C+v) \left(\frac{U}{q} \right)^{\frac{r}{\delta+h}}$$

Proceeding as in the rest of the paper, the following results are obtained

$$(A.9) \quad B_x = 0$$

$$(A.10) \quad B_y = \frac{U}{1+\alpha} \frac{\delta+h}{h} - \frac{r(C+v)}{hx} \left(\frac{U}{q} \right)^{\frac{r+h}{\delta+h}}$$

which is the same as

$$(A.11) \quad B_y = \frac{U_q y}{y_q} - \frac{e^{-rw} [\alpha(C_q + v_q) + \frac{r}{\delta+h} \frac{C+v}{q}]}{y_q}$$

The result and its rationale are therefore unchanged. It is still optimal to reward hospitals on the basis of their demand, paying them a fixed price per person entering

the queue, calculated as in section 1. In this case, as can be seen from (A.10) or (A.11), an increase in δ will have a positive effect on the price B_y , meaning that a stronger dislike for waiting leads to a higher payment per patient to the hospital. Unfortunately, the calculation of the effect of δ on the equilibrium is inconclusive. Intuitively one might expect the higher marginal revenue to translate into an incentive to increase x and q , in order to increase y , possibly leading to higher equilibrium values of both quality and quantity. However, the outcome where either only x or q is increased cannot be ruled out.

Appendix 3

Oligopoly of supply.

This appendix considers the case of an oligopolistic market with a finite number m of homogeneous hospitals. Whereas the homogeneity assumption might still be simplifying, the hypothesis of a finite number of hospitals is quite realistic, especially for local market conditions. The case is modelled as a Cournot-Nash game, in which hospitals decide their quality-quantity mix at the same time, taking the others' decisions as given.

This interaction between hospitals can be introduced in the model via a respecification of their demand function. In a symmetric case with m hospitals the demand of each hospital in equation (17) becomes

$$(A.12) \quad y_i = \left(\frac{q_i}{U} \right)^{\frac{h}{r+h}} x_i \quad \forall i$$
$$i = 1, \dots, m$$

and the total market demand is

$$Y = \sum_i y_i$$

As a symmetric equilibrium is assumed, (A.12) can be equivalently rewritten as

$$(A.13) \quad y_i = \frac{\left(\frac{q_i}{U} \right)^{\frac{h}{r+h}} x_i}{\sum_j \left(\frac{q_j}{U} \right)^{\frac{h}{r+h}} x_j} Y \quad \forall i$$

Equation (A.13) shows that every decision made by hospital i about its quality or quantity level will affect it both directly, through the effect on its demand, and

indirectly, through the effect on the others' market shares. Differentiating (A.13) with respect to x and q gives

$$y_x = \frac{y}{x} \left(\frac{m-1}{m} \right)$$

$$y_q = \left(\frac{h}{r+h} \right) \left(\frac{y}{q} \right) \left(\frac{m-1}{m} \right)$$

The same procedure as in section 1 is then used (maximise the social welfare function and the hospitals' utility function and equate their F.O.C.) but using the above expressions for the marginal variation of demand with respect to quantity and quality, because this specification explicitly takes into consideration the number of firms and their market shares. This gives the following results:

$$B_x = 0$$

$$(A.14) \quad B_y = \frac{m}{m-1} \left(\frac{U}{q} \right)^{\frac{r}{r+h}} \left[\frac{x}{(1+\alpha)} \frac{(r+h)}{xy} - \frac{r(C+v)}{hy} \right]$$

This is exactly the same result of (25) multiplied by $m/(m-1)$.

Thus again, in the case of an oligopoly it is also optimal to pay hospitals a zero price for the number of cases treated and to reward them instead on the basis of their demand. It is therefore only the price per entrant to be influenced by the number of hospitals on the market. As can be seen by differentiating (A.14) with respect to m , the higher is this number the lower is the price and viceversa, provided that $m \geq 2$. That is, the result does not hold in the case of a monopolistic provider. This is quite intuitive, as the whole model works on the adjustments of demand. In the case of a monopoly demand is fixed, as patients do not have any choice as of where to go to receive treatment. Formally this means that both y_x and y_q are zero. From the objective

function of the hospital it can be easily seen that it would provide too low a quality level, even though a positive price per case could induce it to treat the right number of patients.

Bibliography

ALLEN, R. and GERTLER, P. (1991). Regulation and the provision of quality to heterogeneous consumers: the case of prospective prices of medical services. *Journal of Regulatory Economics*, 3, 361-375.

APPLEBY, J. (1994). *Developing contracting: a national survey of DHAs, boards and NHS trusts*. NAHAT, Birmingham.

BAGGOTT, R. (1994) *Health and Health Care in Britain*, St.Martin's Press.

BARKER, K., CHALKLEY, M., MALCOMSON, M.J. and MONTGOMERY, J. (1996). Contracting in the NHS; legal and economic issues. *Discussion Papers in Economics and Econometrics*, no. 9622, University of Southampton.

BARTLETT, W. and LE GRAND, J. (1992). The impact of NHS reforms on hospital costs. In *Studies in decentralisation and quasi-markets*, n.8, SAUS, University of Bristol.

BARTLETT, W. and LE GRAND (1994a). *Quasi-markets and Social Policy*, MacMillan.

BARTLETT, W. and LE GRAND, J. (1994b), The performance of trusts. In ROBINSON, R. and LE GRAND, J. (1994). *Evaluating the NHS reforms*. The King's Fund Institute.

BESLEY, T., HALL, J. and PRESTON, I. (1999). The demand for private health insurance: do waiting lists matter? *Journal of Public Economics*, 72, 155-181.

CHALKLEY, M. and MALCOMSON, M.J. (1996a). Competition in the NHS quasi-markets. *Oxford Review of Economics Policy*, 12 (4), 89-99.

CHALKLEY, M. and MALCOMSON, M.J. (1996b). Contracts for the National Health Service. *Economic Journal*, 106 (439), 1691-1701.

CHALKLEY, M. and MALCOMSON, M.J. (1998a). Contracting for health services with unmonitored quality. *Economic Journal*, 108 (449), 1093-1110.

CHALKLEY, M. and MALCOMSON, M.J. (1998b). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17 (1), 1-19.

CHALKLEY, M. and MALCOMSON, M.J. (1998c). Government purchasing of health services. *Discussion Papers in Economics and Econometrics*, 9821, University of Southampton.

CULYER, A.J. (1971). The nature of the commodity health care and its efficient allocation. *Oxford Economic Papers*, 23 (2) 189-211.

CULYER, A.J. (1991). *The Economics of Health*. E.Elgar Publishers.

DE FRAYA, G. (2000). Contracts for health care and asymmetric information. *Journal of Health Economics*, 19, 663-677.

DoH, (1989). *Working for Patients*, HMSO, London (Cm 555).

ELLIS, R.P. and MC GUIRE, T. (1986). Provider behaviour under prospective reimbursement. *Journal of Health Economics*, 5, 129-151.

ELLIS, R.P. and MC GUIRE, T. (1991). Optimal Payment systems for health services. *Journal of Health Economics*, 9, 375-396.

ELLIS, R.P. (1998). Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics*, 17, 537-555.

ELWOOD, S. (1996). Full-cost pricing rules within the NHS. *Management Accounting Research*, 7, 25-51.

FATTORE, G. (1999), Cost containment and health care reforms in the British NHS. In MOSSIALOS, E. and LE GRAND, J. (1999). *Health Care and Cost Containment in the European Union*, Ashgate, USA.

HOFFMEYER, U.K. and MC CARTHY, T.R. (1994). *Financing Health Care*. Vol.1, Dordrecht, London, Kluwer.

HOLMSTROM, B. and MILGROM, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organisation*, 7, 24-52.

LAFFONT, J.J. and TIROLE, J. (1993). *A Theory of Incentives in Procurement and Regulation* MIT Press, Cambridge, MA.

LE GRAND, J. (1994). Evaluating the NHS Reforms. In ROBINSON, R. and LE GRAND, J. (1994). *Evaluating the NHS reforms*. The King's Fund Institute.

MA, C.A. (1994). Health care payment systems: cost and quality incentives. *Journal of Economics and Management Strategy*, 3, 93-112.

MOSSIALOS, E. and LE GRAND, J. (1999). *Health Care and Cost Containment in the European Union*, Ashgate, USA.

NEWHOUSE, J.P.(1970). Towards a theory of non-profit institutions: an economic model of a hospital. *American Economic Review*, 60, 64-74.

PAULY, M. and REDISCH, M. (1973). The not-for-profit hospital as a physicians' cooperative. *American Economic Review*, 63, 87-99.

PROPPER, C. (1994). Quasi-markets, contracts and quality in health and social care: the US experience. In BARTLETT, W. and LE GRAND (1994a). *Quasi-markets and Social Policy*, MacMillan.

PROPPER, C. (1996). Market structure and prices: the responses of hospitals in the UK NHS to competition. *Journal of Public Economics*, 61, 307-335.

PROPPER, C. (2000). The demand for private health care in the UK. *Journal of Health Economics*, 19, 855-876.

ROGERSON, W. (1994). Choice of treatment intensities by a non profit hospital under prospective pricing. *Journal of Economics and Management Strategy*, 3, 7-51.

SALTMAN, R. (2002). The Western European experience e with health care reform. *European Observatory on Heath Care Systems*.