THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS

Contracts for Health Care
and Asymmetric Information

Gianni De Fraja

CHE Technical Paper Series 3

Contracts for Health Care
and Asymmetric Information

Gianni De Fraja

December 1996

CENTRE FOR HEALTH ECONOMICS TECHNICAL PAPER SERIES

The Centre for Health Economics has a well established Discussion Paper series which was originally conceived as a means of circulating ideas for discussion and debate amongst a wide readership that included health economists as well as those working within the NHS and pharmaceutical industry.

The introduction of a Technical Paper Series offers a further means by which the Centre's research can be disseminated. The Technical Paper Series publishes papers that are likely to be of specific interest to a relatively specialist audience, for example papers dealing with complex issues that assume high levels of prior knowledge, or those that make extensive use of sophisticated mathematical or statistical techniques.

The series provides a maximum of 20 pages for authors. The content and its format are entirely the responsibility of the author, and papers published in the Technical Paper series are not subject to peer-review or editorial control, unlike those produced in the Discussion Paper series. Offers of further papers, and requests for information should be directed to Frances Sharp in the Publications Office, Centre for Health Economics, University of York.

THE AUTHOR

Gianni De Fraja is Reader in Economics at the University of York. He holds a DPhil from the University of Oxford, and has worked at the Universities of Leicester and Bristol.

Contracts for Health Care and

Asymmetric Information

Gianni De Fraja*

University of York and C.E.P.R.

December 1996

Abstract

The paper presents a general model of contracting for health services. A purchaser (NHS, insurer) offers the optimal contract to providers (hospitals, GP's), under the constraint of its limited information about the provider's cost. A number of features of such contracts are derived, some surprising: for example, under plausible conditions, the price per case increases with the efficiency of the provider.

Keywords: - Contracts, health services, regulation, NHS.

JEL Number: - I18, L51.

Address for correspondence:
 Department of Economics and Related Studies
 University of York
 Heslington
 York
 YO1 5DD

 tel: +(44) 1904 433767
 fax: +(44) 1904 433759
 e-mail gd4@york.ac.uk

1. Introduction.

Wide reaching reforms have swept the UK National Health Service from the late 80's (see the papers in Culyer et al 1990; Propper 1995 gives a preliminary assessment). One of the main building blocks of these reforms is the separation between providers and purchasers; its aim is to break down "the close personal relationship between employees of the purchaser and those of the provider, thus reducing the influence that providers have over the decisions of the purchasers" (Chalkley and Malcomson 1996, p. 6).

From a theoretical point of view this separation creates the need of formal contractual relationship between provider and purchaser[1,2], where previously informal suggestions or hierarchical command would have sufficed. The aim of this paper is the study of a very stylised contractual relationship between a provider and a purchaser of health services. I consider the provision of one specific and unambiguously defined service, the quality of which does not vary; for the sake of definiteness, one can think of a test or routine surgical operation.[3] The benefit and cost of this operation, however, do vary across patients: on the one hand, some patients may benefit less than others, for example because they are older or affected by other pathologies. On the other hand, the cost of treatment can vary because some patients will need a longer recovery period than others, for some patients the surgery will take longer or need a larger team, and so on. This is realistic. Also realistic is my other assumption that providers differ in efficiency, namely, in the cost of carrying out the operation on identical patients.

My theoretical approach is closely inspired by the modern theory of regulation (see, for an exposition Laffont and Tirole 1993, indeed the paper could be seen as an application of this theory; see also Belli 1996). The central tenet of this theory is that asymmetry of information bedevils contractual relationship and has costly effects, which would not be incurred in the presence of perfect knowledge. The analysis is therefore conducted under the assumption that hospitals have private information, which the purchaser is unable to verify.

The paper aims to provide a theoretical background for the empirical analysis of the contracts agreed between District Health Authorities and "their" hospitals across the country, evaluating the first years of the NHS reforms. In practical terms, the most important message of the paper is a cautionary warning against hurried and simplistic conclusion to be drawn from the analysis of these contracts: some of the conclusions of the paper, taken at face value, would appear very counterintuitive. For example, I show that the price charged for the service is higher (at least up to a point) in more efficient hospitals, that is, in hospitals whose cost of providing the service on a given patient is lower. A brief reflection, however, shows that the reason why this happens is

---

[1] It therefore corresponds to the constitution of a subsidiary part of a multi-divisional company into a separate legal entity. The fact that disputes between the parties are not dealt with by courts but by a special NHS arbitrator is of course irrelevant in the present analysis.

[2] Contracts for services are also important in the US, where large organisation (HMO's, PPO's, etc.) can impose terms on the providers, whereas in the past they simply accepted the provider's posted prices (see Dranove and White 1994, pp 174-176).

[3] Most of the approved list of goods and services which GP fundholders can purchase directly from providers, listed in the National Health Service and Community Care Act 1990 would fit the situation described in the paper.

natural: in the presence of a budget constraint, a purchaser will want to restrict the number of cases treated; allocative efficiency, however, dictates that higher cost providers be required by the purchaser to treat fewer patients. Given their ability to select the case load, they will have an easier case mix than more efficient hospitals. In the presence of asymmetric information, the purchaser does not know the hospital's efficiency, and to stop a low cost hospital from choosing a contract tailored for a high cost provider, it must reward and entice the latter with a higher payment per patient.

I also show that more efficient hospitals, in addition to treating more patients, receive a larger overall budget, and that this effect is strengthened by the asymmetry of information. Thus asymmetric information increases the variability of budget allocation across the providers. It is also shown to reduce the total number of cases treated. By imposing a separation between providers and purchasers, the NHS reforms could arguably be seen as having the effect of increasing the asymmetry of information between them. In this interpretation, the effects of these reforms could be made to correspond loosely to the effects of asymmetric information in the model of the paper. The paper ends showing that, in plausible conditions, the two main types of contracts used in the NHS, that is cost per case and cost and volume contracts, are equivalent: the purchaser can offer the provider a menu of fixed payments plus a payment per case treated, thus delegating the decision on the number of cases to be treated to the provider.

The plan of the paper is as follows: the model is presented in Section 2, the benchmark case of symmetric information in Section 3, the main results in Section 4, and concluding remarks in Section 5. The mathematical derivation of the results are assembled in the Appendix.

2. The Model

I consider a health care purchaser (a District Health Authority) who offers a contract to a provider (hospital). The contract specifies the number of cases that the provider is required to treat and the total price (or, equivalently, the price per treatment) paid by the purchaser for this service. The demand for the service is exogenously given and known to both parties; without loss of generality, it can be normalised to 1. I begin by assuming that, if the purchaser buys services for a given treatment from several providers (e.g. hospitals located in different cities within the purchaser's administrative region), each contract is written independently of all others. This reflect current practice, and it entails no loss in generality as long as the various providers are independent from one another: there are no externalities[4] and unobserved technological shocks are independently distributed among them.

Patients differ according to their ability to benefit from the treatment. Specifically, I assume that each patient is characterised by a parameter $\beta \in \Re$, which denotes her ability to benefit (this could be the increase in QALYs which the treatment determines). β is distributed in the population according to a

---

[4] This is a very strong assumption: it implies that it is not possible to transfer patients from one hospital to another. Further extensions of the model will obviously need to relax it. The paper by Gal-Or (1994) proposes an elegant model for the analysis of such situations.

function $H(\beta)$, with $H'(\beta) = h(\beta)$ and $\lim_{\beta \to -\infty} H(\beta) = 0$, $\lim_{\beta \to +\infty} H(\beta) = 1$.

The purchaser's objective is the maximisation of a utility function, given by:

$$V(T, x(\beta)) = -(1+\lambda)T + \int_{-\infty}^{+\infty} x(\beta)v(\beta)h(\beta)d\beta \qquad (1)$$

where:

- $x(\beta) \in [0,1]$ is the proportion of the patients whose benefit is β who receive the treatment.

- $v(\beta)$ is the social benefit of treating a patient with benefit β. v satisfies $v'(\beta) > 0$, but, in general it need not satisfy $v(\beta) = \beta$, which would imply that the social benefit equals the individual benefit.

- T is the total contractual payment made to the provider.

- λ is the shadow cost of public funds. It reflects the fact that raising funds to pay for public services free at the point of consumption involves a distortionary and administrative cost, so that the overall cost of raising £1 is £$(1+\lambda)$. In most of the literature λ is exogenously given. In the present set up it could be determined as a consequence of an overall budget constraint imposed by the ultimate authority (the Department of Health) on the purchaser, and by the consequent need of the latter to equalise marginal costs across activities. In this case, λ is the multiplier of the purchaser's overall budget constraint, which, if the number of different contracts written is sufficiently large (realistically, since contracts are written for numerous services), can be treated as approximately constant for a single contract.

The standard approach in regulation and procurement is to assume that the purchaser of public services maximise the sum of the utilities of the participants in the industry, including the provider (Laffont and Tirole 1993). This would entail adding U, the provider's utility, to the purchaser's maximand. Doing so might appear in contrast to the intended separation of providers and purchasers, which is the focus of the paper. At any rate, adding the provider's utility, weighted by $\alpha \in (0,1]$, to the purchaser's payoff would not alter the qualitative nature of the results.

With regard to the cost of treatment, it is assumed to depend on two parameters, β and γ. $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ denotes the intrinsic efficiency of the provider, and reflects the hypothesis that hospital differ in their efficiency: if two hospitals are characterised by different values for γ, then they would incur a different cost if they treated the same patient[5]. I assume that cost is decreasing in β: patients who benefit most from the treatment are also the cheapest to treat. In the not unrealistic simplification where β measures the life expectancy, this implies that younger patients (who are expected to survive

---

[5] Factors affecting the efficiency of providers could be the availability of qualified nurses in the areas where the hospital is located, the intrinsic ability of the consultants, the organisational skill of the management, and so on.

longer) are cheaper to treat than older ones[6]. Formally, the provider's monetary cost of treating a given patient is $c(\beta,\gamma)$, satisfying: $c_\beta(\beta,\gamma)<0$, $c_\gamma(\beta,\gamma)>0$, $c_{\gamma\beta}(\beta,\gamma)\leq 0$, $c_{\gamma\gamma}(\beta,\gamma)\geq 0$, where subscripts denote partial derivatives (the assumptions made on the second derivatives ensure the existence of interior solutions). With regard to the distribution of γ, the following standard assumption is made.

Assumption 1: $\gamma \in \left[\underline{\gamma},\bar{\gamma}\right]$ is distributed according to a differentiable function $G(\gamma)$, with $G\left(\underline{\gamma}\right)=0$,

$$G(\bar{\gamma})=1,\ \underline{\gamma}<\bar{\gamma},\ G'(\gamma)=g(\gamma),\ \text{and}\ \frac{d}{d\gamma}\left(\frac{G(\gamma)}{g(\gamma)}\right)>0\ \text{(monotonic hazard rate)}.$$

Moreover, β and γ are independently distributed: the distribution of the potential benefit in the population does not depend on the hospital's intrinsic efficiency. This seems natural.

Assumption 2: Before choosing to treat a patient, the hospital can observe her specific value of β; having observed this patient's β, the hospital can decide to turn her away. If a patient is not treated during the contractual period, then she does not return to the same hospital during the same contractual period, or in subsequent contractual periods.

That is, hospitals can dump patients (Ma 1994, pp 103-108). This certainly appears to be the case in the NHS[7]. From a theoretical point of view, refusing treatment to some patients must inevitably follow from the fact that technology in health care is far more advanced than the taxpayer's willingness to pay for it. The assumption that if a patient is refused treatment by a hospital then she is not treated in the NHS, follows from the assumption that the service is provided by one hospital only. In practice, routine services are provided by several hospitals within a region, and purchasers do transfer patients from one provider to another. Further research needs to consider this case: the analysis of the situation where only one provider is available must therefore constitute a first step of the study of a more complete model where the contracts agreed between a purchaser and its providers are not independent of each other.

With regard to the utility function of the provider, I assume hospitals to be profit maximisers. This might seem objectionable. In the UK, a NHS hospital cannot make any financial surplus and, in the US and elsewhere, many private hospitals are non-profit organisations. However, in the simplified framework considered here, any plausible utility function for the hospital reduces ultimately to profit maximisation (see also Dranove and White 1994, pp 198ff). For example: the managers of a hospital

---

[6] Including some measure of "fitness" would strengthen the case: fitter patients are likely to recover more fully (benefiting more) and more quickly (occupying a hospital bed for a shorter period). One can always think of situations where the opposite hold: for example, when the treatment relieves pain or discomfort, the patients who benefit less are those mildly affected, and hence those who can probably be treated more cheaply.

[7] In a case that has attracted considerable press attention, Cambridge and Huntington Health Authority refused to use NHS funds for a second bone marrow transplant to an 11-year old girl. A private donation allowed the girl to survive a further year (see Lenaghan 1996 for a detailed description of the way dumping differ across the British health authorities).

may want to avoid labour confrontation, to the extent that overmanning allows more pleasant conditions, such as more flexible time or longer tea breaks, additional profit may allow overmanning. Alternatively, even in the situation considered here where there is no scope to vary the quality of the health care provided, if hospital managers are altruistic and care about ancillary services provided to patients, any cash obtained from the purchaser over and above the minimum strictly necessary to provide the health care required can be used to provide things like better food, longer visiting hours, more single rooms, satellite television, and so on. Again, managerial theories of the firm stress how running a larger company yields more utility than running a smaller one. If the same holds for managers of hospitals, then they may use any financial surplus to expand, by expanding the range of services provided. Finally, cost padding and bribery can be easily hidden in presence of profits: a manager may have a friend providing the hospital with, say, laundry services. Extra cash may allow him to overpay for laundry services and personally receive some cash back from his friend. An argument which would, on the surface, seem to contradict this discussion is that doctors might want to have "interesting" cases, which give professional satisfaction and prestige, for example in the form of research publications, and "interesting" cases are likely to be more expensive. The contradiction is only superficial, given the assumption that the service considered is a routine one, where managers are well informed about costing it. It does not rule out that doctors can be keener to treat more "interesting" cases, but it implies that they will be willing to accept a lower wage for the privilege of doing so. In calculating the overall cost of their case-mix a manager will consider this effect. While the study of incentive problems *within* providers is of extreme importance, I concentrate here on the effects of asymmetry of information between provider and purchaser, and rule out any internal incentive problem: the managers of the hospital and the consultants have exactly the same information.

The assumptions that the provider's objective is the maximisation of its profit, and that the purchaser wants to maximise the number of patients treated subject to a shadow cost of public funds imply that the two parties have perfectly aligned objectives with regard to *which* patients should be treated, given the total number agreed upon: there is no conflict of interest in this respect. The health authority, therefore, can delegate the choice of the patients to be treated to the hospital, and it needs neither monitor the choice of patients, nor provide the hospital with costly incentives to choose them appropriately, but it can simply choose n, the number of cases to be treated: the hospital will choose to treat all those patients whose ability to benefit is β^* or higher, with $n = 1 - H(\beta^*)$.[8] It follows that in the purchaser's utility function, $x(\beta) = 0$ for $\beta < \beta^*$, and $x(\beta) = 1$ for $\beta \geq \beta^*$.

3. The Optimal Contract with Symmetric Information

I consider first the benchmark case where the hospital cost parameter, γ, is known to the purchaser. In

---

[8] Suppose not, that is suppose that given $\beta_1 < \beta_2$, patients with $\beta \in [\beta_1, \beta_1 + \varepsilon)$ are treated, and patients with $\beta \in [\beta_2, \beta_2 + \varepsilon)$ are not treated. Then by treating the latter and omitting to treat the former the hospital could be increasing its profit by reducing its cost without modifying the total number of patients treated. Both the provider and the purchaser are strictly better off as a consequence. All statements in the paper should be qualified by noting that they hold except at most in a set of measure zero: in what follows I leave implicit the reference to this qualification.

view of the remark that closed the previous section, the health authority chooses the "cut-off" value of β below which patients are not treated, defined β^S (S for symmetric information), the total payment given to the hospital, T^S, and U, the hospital utility level: the purchaser solves the following problem.

$$\max_{\beta^S, T^S, U} \left\{ -(1+\lambda)T^S + \int_{\beta^S}^{\infty} v(\beta)h(\beta)d\beta \right\}$$

$$\text{s.t.:} \quad U = T^S - \int_{\beta^S}^{\infty} c(\beta,\gamma)h(\beta)d\beta \qquad\qquad U \geq 0$$

The constraint $U \geq 0$ is the standard participation constraint, which guarantees non negative profit to the hospital for the particular service considered. If it were violated, then it would not be possible to break even for this service, and a rational hospital would cease to provide it.

Proposition 1: When the purchaser knows the realised value of γ, the optimal choice of β^S, T^S and U satisfies:

$$(1+\lambda)c(\beta^S,\gamma) = v(\beta^S) \qquad\qquad T^S = \int_{\beta^S}^{\infty} c(\beta,\gamma)h(\beta)d\beta \qquad\qquad U = 0$$

The proof is immediate, and it is therefore omitted. Figure 1 below illustrates the derivation of β^S (disregard for the moment the dashed line): the curve labelled $c(\beta,\gamma)$ is the marginal cost of treating an additional patient, and the curve $v(\beta)/(1+\lambda)$ is the marginal benefit; both cost and benefit are measured in "social" monetary units. β^S is set to equate the two. Figure 1 can be used to determine the effects of changes in the exogenous parameters on variables of interest, such as the number of patients treated, the cost of treating the patient with the least ability to benefit among those treated (who is also the most expensive), the total cost of the service, and the average cost per patient treated. The last of these variables, is simply the price a hospital charges for the given service, and, allowing easy to understand comparison between hospitals, is highly likely to attract considerable public interest.

Corollary 1: $\dfrac{d\beta^S}{d\gamma} > 0$, $\dfrac{dc(\beta^S,\gamma)}{d\gamma} > 0$, $\dfrac{dT^S}{d\gamma} \lesseqgtr 0$, $\dfrac{d}{d\gamma}\left(\dfrac{T^S}{n}\right) \lesseqgtr 0$.

Thus an increase in the hospital intrinsic cost reduces the number of patients treated. Figure 1 illustrates the effects of a change in γ on β^S, the cut-off point and $c(\beta^S(\gamma),\gamma)$, the cost of the most expensive patient treated: still disregarding the dashed line, it is easy to see that an increase

in γ shifts the $(1+\lambda)c(\beta,\gamma)$ curve up, implying a decrease in β^S and an increase in $c(\beta^s(\gamma),\gamma)$: although the last patient treated in a high cost hospital is a patient whose condition is less severe than the condition of the last patient treated in a low cost hospital, the cost of such treatment is higher in a cost hospital. The effect of increase in γ on the total budget allocated to the hospital and on the average price paid for a treatment is ambiguous; as shown in Section 4, this is no longer the case with asymmetric information[9].

4. The Optimal Contract with Asymmetric Information

In this section I study the optimal contract[10] that the purchaser offers when it cannot observe the cost structure of the provider, but only knows distribution of γ. The comparison between the two cases gives the effects of the asymmetry of information.

The purchaser offers the provider a schedule $T(n)$, specifying a total payment, T, depending on the number of cases treated, n, and allows the hospital to choose a point on this schedule. In this case, the point chosen by the hospital gives the number of cases treated and the total payment which would be chosen if the purchaser first specified two schedules $T(\gamma)$, $\beta^*(\gamma)$, giving the total payment T and the cut-off point β^* as a function of the *reported* value of γ, and then asked the hospital to report the value of γ, subsequently offering a contract specifying a total payment of $T(\gamma)$ and a number of cases to be treated $n = 1 - H(\beta^*(\gamma))$. The revelation principle says that the purchaser cannot do better (in terms of maximising his own objective function) than restricting its choice to the schedules satisfy the truth-telling constraint: the hospital is better off reporting the truth than lying. The first step is therefore the determination of the hospital's truth-telling constraint (or incentive compatibility). This is done in the following proposition.

Proposition 2: The hospital's truth-telling constraint is given by:

$$\dot{U}(\gamma) = - \int_{\beta^*(\gamma)}^{\infty} c_\gamma(\beta,\gamma)h(\beta)d\beta \qquad \text{and} \qquad \dot{\beta}^*(\gamma) > 0 \tag{2}$$

The proofs of all the results are confined to the appendix. I use the standard notation $\dot{U}(\gamma) \equiv \dfrac{dU(\gamma)}{d\gamma}$. Note that $\dot{U}(\gamma) < 0$, that is, more efficient hospitals are better off. A further constraint which the purchaser must satisfy is the participation constraint, ensuring that under no circumstances the

---

[9] It is easy to derive other comparative statics results, studying the effects of changes in other parameters of interest, such as the social cost of public funding. This is straightforward, and is omitted.

[10] I study the *optimal* regulation, given the (informational) constraints faced by the purchaser: thus the present model differs from Fenn et al's (1994) analysis of contracts for the NHS, in that they consider two possible regimes of price regulation in symmetric information.

hospital refuses to sign the contract: $U(\gamma) \geq 0$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. In view of the fact that $U(\gamma)$ is decreasing and that rent is costly, it simplifies to:

$$U(\overline{\gamma}) = 0 \tag{3}$$

And the purchaser's objective function becomes:

$$\max_{T(\gamma), \beta^*(\gamma)} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[-(1+\lambda)T(\gamma) + \int_{\beta^*(\gamma)}^{\infty} v(\beta)h(\beta)d\beta \right] g(\gamma)d\gamma$$

Eliminating $T(\gamma)$ from the maximand, using the definition of $U(\gamma)$, an optimal control problem is obtained, whose solution can be obtained using standard techniques, and is given in Proposition 3.

$$\max_{U(\gamma), \beta^*(\gamma)} \int_{\underline{\gamma}}^{\overline{\gamma}} \left[-(1+\lambda)U(\gamma) - \int_{\beta^*(\gamma)}^{\infty} [c(\beta, \gamma) - v(\beta)]h(\beta)d\beta \right] g(\gamma)d\gamma \qquad \text{s.t.: (2) and (3).}$$

Proposition 3: The optimal contract offered by the purchaser in conditions of asymmetry of information satisfies:

$$c(\beta^*(\gamma), \gamma) + \frac{G(\gamma)}{g(\gamma)} c_\gamma(\beta^*(\gamma), \gamma) = \frac{v(\beta^*(\gamma))}{(1+\lambda)} \tag{4}$$

$$T(\gamma) = \int_{\beta^*(\overline{\gamma})}^{\infty} c(\beta, \overline{\gamma})h(\beta)d\beta + \int_{\beta^*(\gamma)}^{\beta^*(\overline{\gamma})} c(\beta, \gamma^*(\beta))h(\beta)d\beta \tag{5}$$

where $\gamma^*(\beta)$ is the inverse function of $\beta^*(\gamma)$.

The proof shows that $\beta^*(\gamma)$ is an increasing function. This implies that a relatively inefficient hospital is required to treat fewer patients; in other words, a patient with low ability to benefit may be refused treatment by a high cost hospital, when a low cost hospital would have treated him. This of course was also the case with symmetric information; the next corollary shows that asymmetry of information exacerbates this effect.

Corollary 2: Let $\beta^s(\gamma)$ denote the optimal cut-off point with symmetric information. Then:
$$\beta^*(\underline{\gamma}) = \beta^s(\underline{\gamma}) \text{ and } \beta^*(\gamma) > \beta^s(\gamma) \text{ for all } \gamma \in (\underline{\gamma}, \overline{\gamma}].$$

[Insert Figure 1 about here]

The dashed curve in Figure 1 depicts the LHS of (4) as a function of β; its intersection with the $v(\beta)/(1+\lambda)$ curve determines the value of β^* for the given value of γ. Unless $\gamma = \underline{\gamma}$, in which case $G(\underline{\gamma}) = 0$, the dashed curve lies above the solid curve $c(\beta,\gamma)$, and therefor e the cut-off point is higher than the corresponding value for the case of symmetric information, β^S.

According to (5), the payment to the hospital is made up of two parts. A first part, independent of γ, is a lump sum equal to the total cost of the least efficient hospital. The second part is less straightforward: it can be interpreted as an incentive payment in the following sense: it constitutes a refund of the cost of treating patients in addition to the number treated by the least efficient hospital; however, each of these additional patients is "valued" at the cost which would be incurred if this patient were treated at the hospital where she would be the last (most serious) patient to be treated. Because the cost of treating inframarginal patients is below this "valuation", an efficient hospital is allowed to keep the cost savings that result from its superior efficiency.

The next corollary studies the relationship between the provider's efficiency and its total budget.

Corollary 3: $\dot{T}(\gamma) = -\dot{\beta}^*(\gamma)c(\beta^*(\gamma),\gamma)h(\beta^*(\gamma)) < 0$.

It was shown in Corollary 1 that an increase in γ has an ambiguous effect on the provider's total budget in the case of symmetric information. In the present situation, however, Corollary 3 shows that efficient hospitals always receive _more_ money than higher cost ones. This would seem surprising. It is however due to the fact that they are required to treat more patients, and that they receive more "profit"; these two effects outweigh the fact that hospitals with lower cost can receive less money to treat a given number of patients and still break even. Although T could be a decreasing function even with symmetry of information, the next Corollary shows that in the presence of asymmetric information, T is on average always steeper.

Corollary 4: Let $T^S(\gamma)$ denote the payment schedule in the symmetric information case. Then: $T(\underline{\gamma}) > T^S(\underline{\gamma})$ and $T(\overline{\gamma}) < T^S(\overline{\gamma})$.

That is the budget of the most efficient hospital is always increased by asymmetric information, that of the least efficient hospital always decreased. Thus an effect of asymmetry of information (and, as argued in the introduction of the separation between purchasers and providers) is therefore an increase in the variability of the of the budget allocation across different hospitals with respect to the allocation which would result in conditions of symmetry of information.

Also note that, since the most efficient provider receives more cash in conditions of asymmetric information, but treats the same number of patients, it follows that, in relatively efficient hospitals, asymmetric information (and the separation of purchasers and providers) increases the cash per patient received by the provider.

This section ends by investigating further the relationship between the total payment received and the number of cases treated. Given the importance attributed by politicians and popular opinion to the issue of prices, this analysis can well be viewed as the main contribution of the paper. I focus on two concepts: the "average" price, and the "marginal price". The average price is simply the total budget divided by the number of cases treated. It is an easy to understand number, and its relationship to γ is described in the following Proposition.

Proposition 4: Define the average price as $t(\gamma) = \dfrac{T(\gamma)}{1 - H(\beta^*(\gamma))}$. Then

$$\frac{dt(\gamma)}{d\gamma} = \frac{\dot{\beta}^*(\gamma)h(\beta^*(\gamma))}{[H(\beta^*(\gamma))]^2}\left[\int_{\beta^*(\gamma)}^{\infty} c_\beta(\beta, \gamma)[1 - H(\beta)]d\beta + U(\gamma)\right].$$ Therefore there exists $\gamma_0 \in [\underline{\gamma}, \overline{\gamma})$

such that $t(\gamma)$ is decreasing for $\gamma \geq \gamma_0$.

In words, the Proposition says that, at least for γ high, the budget per patient treated, that is the *price per treatment*, is increased as the hospital becomes more efficient. Surprising as it may appear, the finding of a positive relationship between the price charged and the efficiency of the hospital has however a natural explanation: efficient hospitals treat more patients, and therefore they treat more difficult cases. This is socially efficient, because they are better hospitals, but of course is more costly, and the hospital must therefore be rewarded for doing so.

Before discussing this further, consider the marginal price. This is the change in the monetary reward that would be received by a hospital if it chose to treat a further patient in addition to those it already treats (see Wilson 1993, p 48). Because the price is non-linear, this differs from the average price. The marginal price is given by the slope of the function, say $\tau(n)$, which relates the total payment received by the provider to the number of cases treated: $\tau(n) = t(\gamma^*(B(n)))$, where $\gamma^*(\beta)$ is defined above as the inverse of $\beta^*(\gamma)$, and $B(n)$ is the inverse of $[1 - H(\beta)]$, that is $B(n)$ is the cut-off point which is necessary to have in order to treat n patients. The following proposition illustrates the shape of this relationship.

Proposition 5: $\tau(n)$ is increasing. Moreover, let $\dfrac{v'(\beta)}{1 + \lambda} > \dfrac{d}{d\gamma}\left(\dfrac{\dfrac{G(\gamma)}{g(\gamma)}c_\gamma(\beta, \gamma)}{c_\beta(\beta, \gamma)}\right)\dfrac{[c_\beta(\beta, \gamma)]^2}{c_\gamma(\beta, \gamma)}$, then $\tau(n)$

is concave.

The condition given in Proposition 5 has a very natural interpretation, as it implies that the cost of the most expensive patient increases as the exogenous cost parameter increases. This was also the case in conditions of symmetric information (see Corollary 1), and it seems a natural result: a marginal increase in γ implies that a hospital can avoid treating its most expensive patient. This reduces its total cost, but, if the condition holds, this reduction in cost is not sufficient to offset the exogenous increase

in the cost.

[Insert Figure 2 approximately here]

Propositions 4 and 5 can be used to determine the shape of the relationship between the total budget and the number of patients treated. This is illustrated in Figure 2. Panel A is the case in which $\gamma_0 = \underline{\gamma}$, panel B the case $\gamma_0 > \underline{\gamma}$. The slope of the dashed line is the average price, the slope of the tangent is the marginal price. The marginal price is always decreasing as the efficiency increases. The average price, on the other hand, increases as γ is reduced: all the way in Panel A, up to γ_0 in Panel B.

As a final remark, the analysis sheds light on the relationship between the two main contracts used in NHS for the present situation: cost per case and cost and volume (a third type, block contract, solidifies pre-reform patterns of referral, and should therefore disappear as contracts with better incentive properties become more widespread; Posnett 1993, p.302). It is easy to see that when the $\tau(n)$ schedule is concave, as in Figure 2, these two types are equivalent, in the sense that the purchaser can indifferently require the provider to treat a given number of cases, or simply offer the provider a "lump sum/price per case" schedule, and allow it to choose its preferred point on this schedule. The contract in this second case specifies the payment of the lump sum and a payment per case, and the provider will select the appropriate number of cases to be treated, $\left[1 - H\left(\beta^*(\gamma)\right)\right]$.

5. Concluding remarks

The paper considers a very simple model of the contractual relationship between a purchaser and a provider of a health service.

One possible way of viewing the paper could be the following: there are several hospitals in the country. Each agrees a contract for the provision of the service considered with its local health authority. Data on caseloads, budgets, and "prices" charged will become available to the Department of Health. After normalisation to take into account differences in demand, the data will show differences among hospital (for an early analysis of such data, see Propper 1996): how are these differences to be interpreted? The analysis of the paper suggests that some of the immediate interpretations one would give are not appropriate. In fact, if the health authorities do offer the optimal contracts, then the relationship between exogenous efficiency and observable variables is often surprising. Lower cost hospital have a higher case load, which is intuitive, but are allocated a higher budget, and the implied price per case is higher than in higher cost hospitals (at least up to a point: very efficient hospitals may reduce their price, see Figure 2B). While counterintuitive, this does have a natural interpretation, as discussed in Section 4. The price charged for services has received much attention; for example, the NHS guidelines require the price to be based on cost. It has been argued (eg Dawson 1995), that this requirement is probably so vague as to be meaningless: the paper reflects this view, both because low cost hospitals treat more expensive patients, with higher β, and because cost can be increased in ways that are not easily observed, it should be treated as endogenous.

Since the focus of the paper is on the effect of the provider's informational advantage, many other facets of the provision of health services are ignored. In particular, I do not consider any issue related to the quality of care, and by assuming separate providers, I abstract from the introduction of competition among different providers. I feel justified in doing so for a number of reasons: firstly, asymmetry of information has been identified by the microeconomic literature as a major source of inefficiency in imperfectly competitive environments, and the technological features of the health markets make it very likely that some parties (e.g. consultants, hospital managers) know something which the purchaser (District Health Authority, HMO, insurer, general practitioner, private patient) does not know, and which is important in their relationship. Secondly, I view this model as a building block of more general models where quality issues and competition between providers can be studied. The framework of the paper could be extended to include quality, and asymmetry of information on the provision of quality. With regard to competition, one way competition could be introduced, is via the franchise of contracts to some among a group of bidders (as suggested by Chalkley and Malcomson 1996, pp 15-16), and the analysis of the paper could extend to the case of several bidders, in the way the Laffont-Tirole analysis of procurement for a single firm (1986) is extended to the n-firm case (1987; see also Laffont and Tirole, 1993, chs. 1, 2 and 7). Last but not least, there are lessons to be learnt from even as simple a model as the one considered here on the role of contracts on the provision of health services. As Dawson (1995) points out, the nature of the market considered, with few agents on each side, implies that it will be governed by contractual relationship, specifying complex price schedules and using any price list made public at most as a guideline or a starting point for negotiation.

References

Apostol, Tom, [1974], *Mathematical Analysis. Second Edition*, (Addison Wesley).

Belli, Paolo, [1996], 'Searching for Optimal Contracts? (Purchasers and Providers in the Health Sector)', mimeo, Università Bocconi, Milan.

Chalkley, Martin, and James M. Malcomson, [1995], 'Contracts and Competition in the NHS', University of Southampton, Disc. Paper n. 9513.

Culyer, Anthony J., Alan Maynard, and John Posnett, [1990], *Competition in Health Care: Reforming the NHS*, (MacMillan).

Dawson, Diane, [1994], 'Cost and Prices in the Internal Market: Markets Vs the NHS Management Executive Guidelines', Centre for Health Economics, University of York, Disc. Paper n. 115.

Dranove, David, and William D. White, [1994], "Recent Theory and Evidence on Competition in Hospital Markets", *Journal of Economics and Management Strategy*, **3** pp. 169-209.

Fenn, Paul, Neil Rickman, and Alistair McGuire, [1994], "Contracts and Supply Assurance in the UK Health Care Market", *Journal of Health Economics*, **13** pp. 125-44.

Gal-Or, Esther, [1994], "Excess Investment in Hospital Capacities", *Journal of Economics and Management Strategy*, **3** pp. 53-70.

Lenaghan, Jo, [1996], *Rationing and Rights in Health Care*, (Institute for Public Policy Research, London).

Laffont, Jean-Jacques, and Jean Tirole, [1986], "Using Cost Observation to Regulate Firms", *Journal of Political Economy*, **94** pp. 614-41.

Laffont, Jean-Jacques, and Jean Tirole, [1987], "Repeated Auctions of Incentive Contracts, Investment and Bidding Parity", *Rand Journal of Economics*, **19** pp. 516-37.

Laffont, Jean-Jacques, and Jean Tirole, [1993], *A Theory of Incentives in Procurement and Regulation*, (MIT Press, Cambridge, Massachusetts).

Ma, Ching-to A., [1994], "Health Care Payment Systems: Cost and Quality Incentives", *Journal of Economics and Management Strategy*, **3** pp. 93-112.

Posnett, John, [1993], "The Political Economy of Health Care Reform in the United Kingdom", in: R. J. Arnould, R. F. Rich & William White, (eds), *Competitive Approaches to Health Care Reform*, The Urban Institute Press, pp. 293-312.

Propper, Carol, [1995], "Regulatory Reform of the NHS Internal Market", *Health Economics*, **4** pp. 77-83.

Propper, Carol, [1996], "Market Structure and Prices: The Responses of NHS Hospitals to Competition", *Journal of Public Economics*, forthcoming.

Wilson, Robert B., [1993], *Nonlinear Pricing*, (Oxford University Press, Oxford).

Appendix

Corollary 1: Totally differentiate the first condition in Proposition 1: $c_\beta(.)d\beta + c_\gamma(.)d\gamma = \dfrac{v'(\beta)d\beta}{(1+\lambda)}$; from which: $\dfrac{d\beta^S}{d\gamma} = \dfrac{c_\gamma(.)}{-c_\beta(.)+v'(\beta)/(1+\lambda)} > 0$. This establishes the first part. Consider the second next: $\dfrac{dc(\beta,\gamma)}{d\gamma} = \dfrac{d\beta^S}{d\gamma}c_\beta(.)+c_\gamma(.) = \dfrac{c_\gamma(.)v'(\beta)}{-(1+\lambda)c_\beta(.)+v'(\beta)} > 0$.

The change in the total cost and in the cost per case can now be seen to be undetermined:

$$\frac{dT^S(\gamma)}{d\gamma} = \frac{d}{d\gamma}\left(\int_{\beta^S(\gamma)}^{\infty} c(\beta,\gamma)h(\beta)d\beta\right) = -\frac{d\beta^S}{d\gamma}c(\beta^S,\gamma)h(\beta^S) + \int_{\beta^S(\gamma)}^{\infty} c_\gamma(\beta,\gamma)h(\beta)d\beta$$

$$\frac{d}{d\gamma}\left(\frac{T^S(\gamma)}{1-H(\beta^S(\gamma))}\right) =$$

$$= \frac{1}{[1-H(.)]^2}\left\{\frac{d\beta^S}{d\gamma}h(\beta^S)\left[\int_{\beta^S}^{\infty} c_\beta(\beta,\gamma)H(\beta)d\beta - c(\beta^S,\gamma)\right] + [1-H(\beta^S)]\int_{\beta^S}^{\infty} c_\gamma(\beta,\gamma)h(\beta)d\beta\right\}$$

Both these expressions can, in general, take either sign. ∎

Proof of Proposition 2: Let $U(\gamma)$ devote the hospital's utility if it truthfully reports cost γ:

$$U(\gamma) = T(\gamma) - \int_{\beta^*(\gamma)}^{\infty} c(\beta,\gamma)h(\beta)d\beta \tag{A1}$$

Let $\varphi(\gamma,\hat{\gamma})$ be the hospital's utility when it has cost γ but makes a false report $\hat{\gamma}$. $\varphi(\gamma,\hat{\gamma}) = T(\hat{\gamma}) - \int_{\beta^*(\hat{\gamma})}^{\infty} c(\beta,\gamma)h(\beta)d\beta$. The first order condition for truth-telling requires that $\left.\dfrac{\partial\varphi}{\partial\hat{\gamma}}\right|_{\hat{\gamma}=\gamma} = 0$ for all $\gamma \in [\underline{\gamma},\overline{\gamma}]$, that is:

$$\dot{T}(\gamma) + \dot{\beta}^*(\gamma)c(\beta^*(\gamma),\gamma)h(\beta^*(\gamma)) = 0 \tag{A2}$$

Differentiate (A1) to get

$$\dot{U}(\gamma) = \dot{T}(\gamma) + \dot{\beta}^*(\gamma)c\big(\beta^*(\gamma),\gamma\big)h\big(\beta^*(\gamma)\big) - \int_{\beta^*(\gamma)}^{\infty} c_\gamma(\beta,\gamma)h(\beta)d\beta$$

Using (A2) the first statement is obtained.

Following Laffont and Tirole (1993, p 63), we derive the following condition for truth-telling.

Let $\phi(\beta,\gamma) = \int_{\beta}^{\infty} c(\tilde{\beta},\gamma)h(\tilde{\beta})d\tilde{\beta}$. Incentive compatibility implies that for β any pair γ_1 and γ_2, the following must hold:

$$T(\gamma_1) - \phi\big(\beta^*(\gamma_1),\gamma_1\big) \ge T(\gamma_2) - \phi\big(\beta^*(\gamma_2),\gamma_1\big)$$

$$T(\gamma_2) - \phi\big(\beta^*(\gamma_2),\gamma_2\big) \ge T(\gamma_1) - \phi\big(\beta^*(\gamma_1),\gamma_2\big)$$

Adding up and integrating, we get:

$$\int_{\gamma_1}^{\gamma_2} \int_{\beta^*(\gamma_2)}^{\beta^*(\gamma_1)} \phi_{12}(x,y)dxdy \ge 0$$

Now $\phi_{12}(x,y) = -c_\gamma(x,y)h(x) < 0$, and hence $\beta^*(\gamma)$ must be non-decreasing. It can also be shown (again see Laffont Tirole 1993, p 121) that if $\beta^*(\gamma)$ is non-decreasing, then the first order condition is sufficient for optimality. ∎

Proposition 3: The hamiltonian of the problem is:

$$H = \left\{ -(1+\lambda)U(\gamma) - \int_{\beta^*(\gamma)}^{\infty} \big[(1+\lambda)c(\beta,\gamma) - v(\beta)\big]h(\beta)d\beta \right\} g(\gamma) + \mu(\gamma) \int_{\beta^*(\gamma)}^{\infty} c_\gamma(\beta,\gamma)h(\beta)d\beta$$

From which $\mu(\gamma) = (1+\lambda)G(\gamma)$, and (4) is obtained. Totally differentiate (4):

$$\left[c_\beta(.) - \frac{v'(\beta)}{(1+\lambda)} + \frac{G(\gamma)}{g(\gamma)}c_{\gamma\beta}(.) \right]d\beta + \left[c_\gamma(.) + \frac{G(\gamma)}{g(\gamma)}c_{\gamma\gamma}(.) + \frac{d}{d\gamma}\left(\frac{G(\gamma)}{g(\gamma)} \right)c_\gamma(.) \right]d\gamma = 0$$

$$\frac{d\beta^*}{d\gamma} = \frac{\left[1 + \frac{d}{d\gamma}\left(\frac{G(\gamma)}{g(\gamma)}\right)\right]c_\gamma(.) + \frac{G(\gamma)}{g(\gamma)}c_{\gamma\gamma}(.)}{-c_\beta(.) + \frac{v'(\beta)}{1+\lambda} - \frac{G(\gamma)}{g(\gamma)}c_{\gamma\beta}(.)}$$

Thus if $c_{\gamma\gamma}(.) \geq 0$ and $c_{\beta\gamma}(.) \leq 0$, then $\dfrac{d\beta^*}{d\gamma} > 0$ and the second order condition for truth-telling are satisfied. To derive (5), note that

$$T(\gamma) = U(\gamma) + \int_{\beta^*(\gamma)}^{\infty} c(\beta,\gamma)h(\beta)d\beta \tag{A3}$$

With

$$U(\gamma) = \int_{\gamma}^{\bar{\gamma}} \int_{\beta^*(\tilde{\gamma})}^{\infty} c_\gamma(\beta,\tilde{\gamma})h(\beta)d\beta d\tilde{\gamma} \tag{A4}$$

Developing (A4):

$$U(\gamma) = \int_{\gamma}^{\bar{\gamma}} \int_{\beta^*(\tilde{\gamma})}^{\infty} c_\gamma(\beta,\tilde{\gamma})h(\beta)d\beta d\tilde{\gamma} + \int_{\gamma}^{\bar{\gamma}} \int_{\beta^*(\tilde{\gamma})}^{\beta^*(\tilde{\gamma})} c_\gamma(\beta,\tilde{\gamma})h(\beta)d\beta d\tilde{\gamma}$$

$$U(\gamma) = \int_{\beta^*(\bar{\gamma})}^{\infty} \int_{\gamma}^{\bar{\gamma}} c_\gamma(\beta,\tilde{\gamma})h(\beta)d\tilde{\gamma}d\beta + \int_{\beta^*(\gamma)}^{\beta^*(\bar{\gamma})} \int_{\gamma}^{\gamma^*(\beta)} c_\gamma(\beta,\tilde{\gamma})h(\beta)d\tilde{\gamma}d\beta$$

See Apostol (1974, p 400) for the derivation of the second term.

$$U(\gamma) = \int_{\beta^*(\bar{\gamma})}^{\infty} h(\beta)[c(\beta,\bar{\gamma}) - c(\beta,\gamma)]d\beta + \int_{\beta^*(\gamma)}^{\beta^*(\bar{\gamma})} h(\beta)[c(\beta,\gamma^*(\beta)) - c(\beta,\gamma)]d\beta$$

$$U(\gamma) = \int_{\beta^*(\bar{\gamma})}^{\infty} c(\beta,\bar{\gamma})h(\beta)d\beta + \int_{\beta^*(\gamma)}^{\beta^*(\bar{\gamma})} c(\beta,\gamma^*(\beta))h(\beta)d\beta - \int_{\beta^*(\gamma)}^{\infty} c(\beta,\gamma)h(\beta)d\beta$$

which gives (5), from (A3). ■

Corollary 2: This follows immediately from (4) and Figure 1, noting that the second term on the LHS is positive, except at $\gamma = \underline{\gamma}$, and is zero in symmetric information. ■

Corollary 3: From (A3) above $\dot{T}(\gamma) = \dot{U}(\gamma) - \dot{\beta}^*(\gamma)c(\beta^*(\gamma),\gamma)h(\beta^*(\gamma)) + \int_{\beta^*(\gamma)}^{\infty} c_\gamma(\beta,\gamma)h(\beta)d\beta$,

and the Corollary follows. ■

Corollary 4: It is $\beta^*(\underline{\gamma}) = \beta^S(\underline{\gamma})$ and $U(\underline{\gamma}) > 0$, hence $T(\underline{\gamma}) = U(\underline{\gamma}) + \int_{\beta^*(\underline{\gamma})}^{\infty} c(\beta,\gamma)h(\beta)d\beta >$

$\int_{\beta^*(\underline{\gamma})}^{\infty} c(\beta,\gamma)h(\beta)d\beta = T^S(\underline{\gamma})$.   From   Corollary   2, $\beta^*(\bar{\gamma}) > \beta^S(\bar{\gamma})$,   thus

$$T(\bar{\gamma}) = \int_{\beta^*(\bar{\gamma})}^{\infty} c(\beta,\gamma)h(\beta)d\beta < \int_{\beta^S(\bar{\gamma})}^{\infty} c(\beta,\gamma)h(\beta)d\beta = T^S(\bar{\gamma}).$$ ■

Proposition 4: Expand $\dfrac{dt(\gamma)}{d\gamma}$:

$$\frac{dt(\gamma)}{d\gamma} = \frac{1}{\left[1 - H(\beta^*(\gamma))\right]^2}\left[\dot{T}(\gamma)H(\beta^*(\gamma)) + h(\beta^*(\gamma))\dot{\beta}^*(\gamma)T(\gamma)\right] =$$

$$= \frac{h(\beta^*(\gamma))\dot{\beta}^*(\gamma)}{\left[1 - H(\beta^*(\gamma))\right]^2}\left[\left[1 - H(\beta^*(\gamma))\right]c(\beta^*(\gamma),\gamma) + \int_{\beta^*(\bar{\gamma})}^{\infty} c(\beta,\gamma)h(\beta)d\beta + U(\gamma)\right]$$

Integrate by part the first two terms in the square bracket, write $-c(\beta^*(\gamma),\gamma)$ as $\int_{\beta^*(\gamma)}^{\infty} c_\beta(\beta,\gamma)d\beta$

and the Proposition follows. ■

Proposition 5:

$$\frac{d\tau}{dn} = \dot{T}\left(\gamma^*(B(n))\right)\gamma^{*\prime}(B(n))B'(n) = \dot{T}\left(\gamma^*(B(n))\right)\frac{1}{\dot{\beta}^*\left(\gamma^*(B(n))\right)}\frac{-1}{h\left(\beta^*\left(\gamma^*(B(n))\right)\right)} =$$

$$= c(B(n),\gamma) > 0$$

and

$$\frac{d^2\tau}{dn^2} = c_\beta(.)B'(n) + c_\gamma(.)\gamma^{*\prime}(.)B'(n) = \frac{c_\beta(.)\dot{\beta}^*(.) + c_\gamma(.)}{-h(B(n))\dot{\beta}^*\left(\gamma^*(B(n))\right)}$$

Next expand $\dfrac{dc\left(\beta^*(\gamma),\gamma\right)}{d\gamma}$.

$$\frac{dc\left(\beta^*(\gamma),\gamma\right)}{d\gamma} = \dot{\beta}^*(\gamma)c_\beta(.) + c_\gamma(.) =$$

$$= \frac{1}{A}\left\{ c_\beta(.)c_\gamma(.)\frac{d}{d\gamma}\left(\frac{G(\gamma)}{g(\gamma)}\right) + \frac{G(\gamma)}{g(\gamma)}c_{\gamma\gamma}(.)c_\beta(.) + \frac{v'\left(\beta^*(\gamma)\right)c_\gamma(.)}{(1+\lambda)} - \frac{G(\gamma)}{g(\gamma)}c_{\gamma\beta}(.)c_\gamma(.)\right\}$$

where $A = -c_\beta(.) + \dfrac{v'\left(\beta^*(\gamma)\right)}{1+\lambda} - \dfrac{G(\gamma)}{g(\gamma)}c_{\beta\gamma}(.) > 0$. The above can be written as:

$$\frac{1}{A}\left\{ \left[c_\beta(.)\right]^2\frac{d}{d\gamma}\left(\frac{\frac{G(\gamma)}{g(\gamma)}c_\gamma(\beta,\gamma)}{c_\beta(\beta,\gamma)}\right) + \frac{v'(\beta*(\gamma))c_\gamma(.)}{1+\lambda}\right\}$$

and the Proposition follows. ∎

Figure 1
Optimal cut-off point, with symmetric and asymmetric information
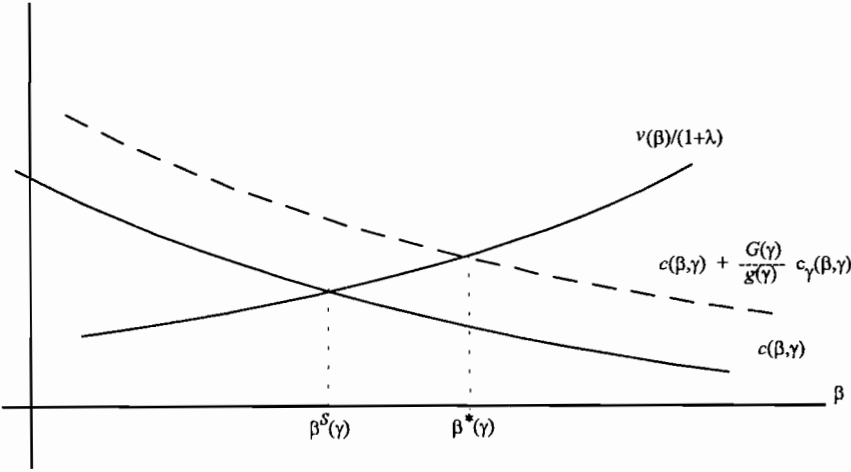
Figure 2
Total Payment as a function of the number of patients treated.
(a): price always increasing as cost decreases; (b) price first increases then decreases as cost decreases.

(a) (b)