XIII CONFERENZA
STATO O MERCATO?
Intervento pubblico e architettura dei mercati
Pavia, Università, 5 - 6 ottobre 2001

ROSELLA LEVAGGI

ARCHITECTURE OF MARKETS FOR HEALTH CARE WITH QUALITY CONTROL AND COST CONTAINMENT

pubblicazione internet realizzata con contributo della COMPAGNIA di San Paolo

Società italiana di economia pubblica

Dipartimento di economia pubblica e territoriale – Università di Pavia

Architecture of markets for health care with quality control and cost containment

Rosella Levaggi
Dipartimento di Scienze Economiche, Università di Brescia
Via S. Faustino 74/b, 25122 BRESCIA
Tel. 030 2988825 - Fax. 030 2988837
e-mail: levaggi@eco.unibs.it

Abstract

La riforma del SSN attuata in Italia a partire dai primi anni '90, così come le analoghe riforme che hanno interessato i principali paesi europei, ha determinato la separazione delle funzioni di acquisto di prestazioni sanitarie dalla produzione delle stesse.

Tale processo ha avuto delle importanti implicazioni per l'organizzazione del sistema sanitario a più livelli.

In questo lavoro ci si occupa della scelta delle forme contrattuali ottime per gestire il rapporto fra fornitori ed acquirenti di prestazioni sanitarie. All'interno di tale relazione, le caratteristiche che si vogliono esaminare sono il problema della qualità del servizio offerto ed il costo delle prestazioni. Il problema viene esaminato al punto di vista dell'acquirente il quale deve determinare un contratto ottimo in termini di qualità e costo della prestazione in un ambiente in cui entrambe le variabili non sono osservabili. La letteratura ha preso in esame diverse figure contrattuali per risolvere il problema.

In questo articolo si propone una soluzione diversa. Per quanto riguarda la qualità, si propone di far utilizzare all'acquirente i segnali che derivano dalla scelta dei pazienti circa la struttura in cui farsi ricoverare. La quantità di prestazioni effettuate e la mobilità possono essere quindi degli indici da cui si può ricavare la qualità.

Per quanto riguarda il problema del rimborso delle prestazioni, l'articolo esamina il problema in un contesto di asimmetria informativa in cui vengono confrontati contratti incentivanti, yardstick competition e forme di oligopolio misto.

L'analisi così svolta dimostra che certe forme di yardstick competition, utilizzate per le public utilities non si adattano molto ai contratti sanitari.

Per quanto riguarda gli altri contratti, la scelta dipende dalla struttura dell'informazione a disposizione dell'acquirente.

Introduction

In western countries a substantial proportion of expenditure for health care is financed by the public sector. Since the first oil crisis in 1975,[1] the objective to rationalise and control expenditure has become a priority for any effective government policy.

Health care systems have been widely reformed and in most countries a separation between purchasing and delivering the service has been operated in order to mimic the structure of a competitive market. However, health care presents peculiarities that prevent reaching a first-best solution.

Purchasing at government level presents problems from the point of view of the incentives for both parties to act in a competitive way. The structure proposed for the internal market is only virtually competitive because the separation of functions does not correspond to a separation of interests as in the competitive market. Laffont and Tirole (1995) show that procurement contracts might not be incentive compatible because of collusion among the parties; efficiency is in fact used to evaluate the performance of the agency that does not receive the surplus of its cost minimisation actions. Levaggi (1999) shows that technical problems prevent the implementation of an optimal incentive structure in procurement contracts where a binding budget constraint exists. Even if we do not take into account these problems, achieving an optimal allocation is particularly difficult in this market context.

The cost of the health care services is determined by the quality of care, by the ability of the patient to take advantage of health care and by the effort of the medical staff. The relationship between quality and health gains is unobservable since it depends on personal characteristics of the patients which are often unpredictable.

The literature has recently proposed contractual forms that might solve the problem. As concerns quality, the literature studies the problem with models where hospitals are sensitive to their own reputation, i.e. reputation enters in their utility function. In this case, the quality of the care to be supplied, although not observable, becomes a relevant variable for the hospital which will then try to keep it at a minimum standard[2].

Another branch of the literature assumes that the suppliers of health care are competing for patients in a space (either physical or related to the characteristics of their products). This literature uses the basic assumptions of the Hotelling competition and is related to the new theory of patient's choice and consumer's search. In its more naive formulation, it is assumed that the patient can observe the quality of the health care he receives and chooses his supplier accordingly[3].

---

[1] and more recently from the '90's

[2] See Malcomson and Chalckley (2000) for a review.

[3] See Gravelle (1999)

The other important area of uncertainty in health care is represented by the cost of the service. The problems related to cost containment have been differently approached and several solutions have been proposed. Yardstick competition and benchmarking seem to be the most popular instruments in this context, but they might not be so effective as it was first thought. The problem is related to the relative incentive that can be given to firms that reveal their true cost and also to the problem of bankruptcy.

In the context of public utilities, the policy of pricing at the average observed cost for other firms in the same industry is optimal for two main reasons: the product they offer is homogeneous and these utilities are usually public companies quoted on the stock exchange where the market can give effective sanctions to the management that is inefficient.

For hospitals, a straight application of yardstick competition is not possible. In public health care systems they are non profit institutions and the product that is supplied is not homogeneous: the case mix can vary between hospitals and often the same case mix can be treated using techniques that have different costs. The use of alternative procedures is determined by the human capital of the hospital and the cost is not always an indicator of efficiency[4]. Finally, it must be considered that if the hospital runs into debt, it cannot be left gone bankrupt because the community cannot be left without medical care[5]; in this case in fact the consequences of a bad management would be borne by the citizens instead of the management.

The use of yardstick competition in health care should then follow the softer rules proposed by the public choice literature for voting local administrators as suggested by Shleifer (1985), Salmon (1987), Besley and Case (1995). However, in this context the mechanism of yardstick competition might not be optimal as shown by Rocaboy et al. (2000)

In this article the cost minimising properties of alternative systems to reimburse hospital treatments will be discussed. The design of the scheme is made from the standpoint of a benevolent Health Authority (HA) that wants to provide a service to his population at the least possible cost. Both quality and the cost cannot be observed, but quality can be inferred through the choice of patients. Hospitals are in fact assumed to compete for patients according to the rules of the Hotelling competition. Consumers receive health care free of charge, but they have to pay for their mobility costs to get to the hospital. The Health Authority reimburses the hospitals for the service they provide. This contract, which is made in a context of asymmetry of information, can be formulated using different schemes that this paper wants to examine as far as their incentive

[4] As an example, we can think to the U.K. system where the training of doctors is organised using the clinical model. In this environment the hospitals that can attract young doctors to be trained can get labour at a cost lower than other hospitals. The same argument can be used for attracting in small local hospitals doctors that can do the newest techniques

[5] On this point see Dawson and Howart (1997)

properties are concerned. In particular, we will consider three different schemes, namely an incentive compatible scheme, yardstick competition and direct management. The first type of contracts derives from the standard agency literature and is the classical solution offered by the literature to the problem of asymmetry of information. The idea of yardstick competition derives form the literature on regulation of public utilities which shows that it is possible to use information on costs deriving from different industries that produces the same basic good to reduce the information rent of the agents. Finally direct management allows to get the relevant information on the costs, i.e. the principal becomes informed on the uncertain parameter.

Given this solution in terms of quality and location, HA can monitor the quality of care offered indirectly through the observation ex-post of mobility patterns of patients. If hospitals are optimally located, mobility should not be observed. In actual fact some mobility might exist because hospitals might not be optimally located, but consistent inflows/outflows of patients from the catchment area of an hospital need to be studied and interpreted in this light.

In this theoretical model HA is able to predict the number of patients to be admitted to hospital; a clause on the number of cases to be reimbursed is the best instrument to control the hospital. While a constraint on the minimum number of cases is a reasonable instrument to secure the quality standard, the use of an instrument like a quantity cap might not be optimal as it will be shown in the section 3.

2. The model: quality control

The quality aspect of the model are regulated by a competition à la Hotelling. The patient that has to be admitted to hospital will have to face two different types of costs:

- the cost of treatment that is financed through an income tax, i.e. there is no charge for the consumer when he is receiving the treatment;
- a cost related to mobility, i.e. with the distance that the consumer has to travel to go to hospital and get the treatment.

Consumer can observe the quality of the care they receive, hence they will choose to go to the hospital that maximises the difference between quality and travelling cost.

The model assumes the most simple case of competition in which the patients are uniformly distributed along a line with a length of 1. Each patient is indexed by $x \in [0,1]$, so that x represents patient located at point x from the origin. Patients have the same valuation of quality characteristics and incur the same marginal distance cost T. The N identical consumers are distributed uniformly around the line of length 1 with density m The line represents the geographic extension of an

Health Authority that has to provide hospital services for the population and that does so using two hospitals A and B that are optimally located in order to make all the patients needing a treatment go to an hospital with the lowest possible cost.

Quality is a parameter that the regulator cannot observe, but in this case it can use the behaviour of consumers as an indicator of the quality of the care produced by the hospital. It will be in fact assumed that consumers can observe, directly or through the choice of their GPs, the quality of the care they receive and that they go to the hospital that allows them to maximise their surplus, defined as the difference between the increment in utility they experience from receiving health care and the cost they have to incur to get access to the service.

Hospital A is located a units of distance from point 0 and hospital B is located b units of distance from 1. To go to an hospital, a patient has to pay a transportation cost T per unit of distance d. The utility function of a patient located at point x may be defined as:

$$U_x = \begin{cases} \varphi(q_A) - p_A - T|x-a| & \text{if patient is admitted to hospital A} \\ \varphi(q_B) - p_B - T|x-(1-b)| & \text{if patient is admitted to hospital B} \end{cases} \qquad (1)$$

where $\varphi(q_A)$ is the monetary equivalent gain from consuming hospital services of quality q in region A; p_A and p_B are respectively prices of hospital A and hospital B; and $T|x-a|$ and $T|x-(1-b)|$ are travel costs.

Let us consider the location of the hospitals as proposed in figure 1 and an environment in which, as it is usually the case in the presence of public health care system, the service if free at the point of use, ie. $p_A=p_B=0$

Several methods can be proposed to find the optimal location of hospital and the quality of the service provided. The traditional approach would be to define the market share of each hospital[6]. In our case, a sequential approach is more suitable to get the solution.

As a first step, the location of hospital is found in order to minimise the travelling cost of the marginal consumer, i.e. of the patient who is located at the furthest possible distance from an hospital. Given that travelling costs are symmetric, from figure 1 we can observe that the optimal location will be ¼ for hospital A and ¾ for hospital B.

The second step consists of finding the minimum quality requirement q that makes the marginal patient go to the hospital.

[6] See Gravelle (1998).

Since the maximum distance to travel is ¼ and the cost are symmetric this means that $q_a=q_b$ and that the quality that makes the marginal patient going to hospital is given by the solution of the following equation:

$$\varphi(q) - \frac{1}{4}T = 0 \qquad\qquad q^* = \varphi^{-1}(\frac{1}{4}T)$$

2. The model: the reimbursement schemes

Every hospital is paid by a Government agency that is allocated from Central Government a specific budget for health care. The Agency wants to minimise cost for supplying the treatment to all the patients, but cannot observe quality of the service or the cost. Once the optimal quality q* has been defined, HA has to set up the optimal reimbursement policy. The cost incurred by the hospital to produce health care is assumed to be a linear function of quality, patients' characteristics and the effort of the medical staff. The unit cost function can be written as:

$$C_i = \beta_i + q^* - e_i$$

where β_i is related to the ability of the patient to take advantage of health care, q^* is the quality and e_i is the effort of the medical staff. β_i is a random variable; to simplify the analysis it will be assumed that β can take only two values, β_l for a patient with low morbidity and β_h for a patient with high morbidity[7]. Both events have a known probability equal to p and (1-p) respectively.

The environment in which the contract has to be made is one of asymmetry of information since the hospital has better information on the recovery speed of the patient hence on the cost of health care. The hospital pays $\beta_i + q_i$ for taking care of a patient of recovery speed i, but this cost can be lowered through the effort e_i of the management of the hospital. The effort produces a disutility additive in patients but increasing in the effort , i.e.

$$f(e) > 0; \qquad f'(e) > 0; \qquad f''(e) > 0$$

The hospital management participates to the production process only if the reward they receive, net of the cost of production produces a positive utility:

$$t_i - C_i - f(e_i) \geq 0$$

2.1 *Cost reimbursement schemes when information is symmetric*

To start with a benchmark model without asymmetry of information will be presented. Given the assumption on the cost and utility function, it is possible to define the optimal contract for just one patient and this solution can then be replicated for all the cases treated.

[7] It is assumed that lmorbidity is correlated with the recovery speed of the patient and hence with cost. If morbidity is low, the rocovery rate is high and hence cost is low.

In an environment characterised by uncertainty but symmetric information, HA has to define an optimal state contingent payment scheme so as to minimise the resources that are necessary to make the hospital supply care of a given quality to the patient. The problem can be written as:

$$\begin{aligned}
&Min \quad pt_l + (1-p)t_h \\
&s.t. \\
&C_i = q^* + \beta_i - e_i \qquad\qquad i = l, h \\
&t_i - C_i - f(e_i) \geq 0
\end{aligned}$$

The solution of this problem is presented in appendix two and can be written as:

$$\begin{aligned}
&f'(e_i) = 1 \\
&t_i^* = C_i^* + f(e_i^*)
\end{aligned}$$

which can be interpreted as follows: in both states of the world, the management is asked to make an efficient, cost-minimising effort in exchange for his reservation utility. This result is the typical solution of a Stackelberg problem. In the presence of symmetric information, in fact, the HA can act as a Stackelber leader and the hospital is a follower.

The payment scheme just presented can be easily translated into a linear payment scheme of the form: $\alpha + \delta_1 C_l . + \delta_2 (C_d - C_l)$[8]. In this case, we will have $\alpha = f(e^*)$; $\delta_1 = \delta_2 = 1$

The following scheme corresponds to cost reimbursement, the typical way in which public-run hospitals have been financed in the recent past; the scheme was also used at first in the U.S. to finance health care supplied under the Medicare programme. Ellis and McGuire (1986) and other authors have widely criticised this method of payment for its cost inflation properties and for not being able to make hospitals produce and efficient level of care for their patients; for this reason in America the scheme has been substituted by a straight prospective payment scheme.

2.2 *The Incentive Compatible schemes*

In this section we consider the effect of the introduction of asymmetry of information on the game described so far. The cost of health care provision, as much as the outcome in terms of improved health, depends on the specific characteristics of each patient that can be observed only by the hospital and that can be used by this latter agent to pursue the maximisation of his objective function [9]. In this article, it will be assumed that, although at the moment in which the contract is stipulated both parties do not know the realisation of β, at some stage before the hospital makes his effort this parameter can be observed by the management and becomes private information to this part. The cost to provide health care is observed by both parties ex post and the effort can be

[8] C_d is the cost declared by the hospital, i.e. either C_l or C_h

inferred for a given declared β. In this environment an asymmetry of information arises and the hospital commands a rent on the private information (the exact realisation of β) he can withhold. HA has to change its strategy in the optimal reimbursement scheme if he wants to avoid the hospital to cheat by declaring the patient is always high morbidity (β_h) in order to reduce his effort and deliver care at cost C_h also when the patient is low morbidity (β_l). In this article we will consider three different alternatives, namely an incentive compatible scheme, yardstick competition and direct management. These solutions have different costs in terms of incentives and also in terms of organisation and control of the system. In what follows, we will concentrate on the costs in terms of rent that the hospital can command while the discussion of the cost of control and design will be left to the following section.

2.2.1 The traditional incentive compatible solution

The first scheme to be considered is an incentive compatible reimbursement to the hospital. The private information it commands means that the optimal effort and reimbursement schemes have to be altered to take account of the asymmetry of information.

If Central government was offering the scheme presented in the previous section in the presence of asymmetry of information, the hospital would always declare that β_h has occurred and it would get an extra utility equal to $f(q^* + \beta_h - C_h^*) - f(q^* + \beta_l - C_h^*)$ if the patient is of type *l*. The expected payoff for each patient for the hospital will then be equal to:

$$pay^{ch} = p[f(q^* + \beta_h - C_h^*) - f(q^* + \beta_l - C_h^*)]$$

When asymmetry of information is explicitly considered, the environment in which the provider operates must be modified. The technology of production and recovery can be observed by both players; when the contract is signed both parties share the same beliefs about the realisation of β , but the provider will be able to observe its value before making his effort. The principal's objective in this new environment can be written as:

$$\begin{aligned}
&Min \quad pt_l + (1-p)t_h \\
&s.t. \\
&C_i = q^* + \beta_i - e_i \qquad\qquad i = l,h \\
&t_i - C_i - f(e_i) \geq 0 \\
&t_i - C_i - f(e_i) \geq t_j - C_j - f(e_{ij})
\end{aligned}$$

where $f(e_{ij})$ represents the effort compatible with declaring C_j when the true state of the world is i. Two constraints characterise the problem: the first means, as before, that the hospital receives a

[9] See, for example, Levaggi (1999)

reward producing at least his reservation utility; the second constraint, which is also called the incentive compatibility constraint, means that the hospital has the incentive to reveal truthfully the state of the world that has occurred and to do his effort accordingly. The solution is characterised by the following conditions that are derived in appendix 2:

$$f'(q^* + \beta_l - C_l) = 1$$
$$f'(q^* + \beta_h - C_h) = 1 - p[1 - f'(q^* + \beta_l - C_h)] < 1$$
$$t_l^{**} = C_l^* + f(q^* + \beta_l - C_l^*) + [f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**})]$$
$$t_h^{**} = C_h^{**} + f(q^* + \beta_h - C_h^{**})$$
$$U_l = f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**})$$
$$U_h = 0$$

The efficient level of effort is now required to the management only in the best state of the world. If the patient is low recovery speed, the hospital will instead deliver an effort short than the optimal one. This scheme allows to satisfy the incentive compatibility constraint with the management receiving his reservation utility only the worst state of the world. If the better state occurs, the hospital receives an extra payment that in economic terms represents the rent for the information it commands. Also in this case, it is possible to use a linear contract to reimburse the hospital:

$$\alpha = [f(q^* + \beta_h - C_h^{**})] + f(q^* + \beta_l - C_l^*) - f(q^* + \beta_l - C_h^{**})]$$
$$\delta_1 = 1$$
$$\delta_2 = 1 - \frac{f(q^* + \beta_l - C_l^*) - f(q^* + \beta_l - C_h^{**})}{C_l^* - C_h^*}$$

In this case, the information rent that the hospital posses makes it receive the following expected payoff:

$$pay^{ICC} = p[f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**})]$$

which correspond to the utility above the reservation level that is received for each patient that is high recovery.

This formula is quite similar to the system adopted in Italy to reimburse hospital care. After the reform in 1995, hospitals are reimbursed through a DRG-related system which should be a prospective payment scheme. In Italy however we have introduced a correction through a supplementary payment to the hospital if the cost of the case is higher than a specific threshold.

The traditional literature that studies health contracts [10] suggests to solve the asymmetry of information suing a prospective payment system based on the expected cost of each treatment. In this case the hospital gets all the surplus and is incentivated in making all the necessary cost minimising procedures. This system can be effectively used for large contracts or for homogeneous patients: in both cases in fact the average actual cost will be close to its expected value. When these conditions are not met, the risk of using a prospective payment system is that the hospital might go bankrupt even if it is efficient[11]. From an economic point of view another important difference exists: in the prospective payment the ability of the patient to recover is not observed even ex-post. This information is extremely valuable to the HA for designing incentive schemes in the future and also for making its health policies more effective. Higher than expected recovery costs, when truly observed, might be the signal of morbidity or environment related problems on which the HA might want to investigate.

2.2.2 Yardstick competition

Yardstick competition is an instrument that the literature studying pricing policies for public utilities widely uses to reduce asymmetry of information about the firms' cost. The idea is to compare the cost of each outlet with those of others facing a similar technology. The environment in which these models are set is one where the regulator has to define a pricing policy for several public utilities that do not compete in the product market because each of them has a spatial monopoly. The technology of production is common to the industry, but the cost of production depends on parameters that can be observed only by each firm. The ability of the regulator to extract this private information depends on how private is the information on this parameter, i.e. it depends on the correlation between the realisation of a predetermined value for one firm and for the other firms. If the shock is perfectly correlated, for example, it is possible to show that the regulator can extract from the agent all the rent while if these shocks are not correlated, the best solution for the principal is to apply a straight incentive compatible scheme. Yardstick competition contracts are applied in the U.S. to Medicare contracts, whereby the reimbursement received by an hospital for a given treatment is based on the average cost of the treatment in similar hospitals. This system can however be successfully applied only if we are prepared to accept that firms can go bankrupt or that they can have a deficit in the short run. Both conditions are not easy to be applied to the internal market for health care where yardstick competition has to be modified in order to take account of these limits.

For health care markets, the structure of yardstick competition that is more suitable is that of yardstick competition in public decisions as studied by the public choice literature. In our case the mechanism could be described in a two-period framework.

---

[10] For a review see Nordhaus (1995) and Malcomson and Chalckley (2000).

Period 1: The two hospitals A and B set the cost for the treatment they offer on the basis of a declared β and the set level of quality q^*. For ease of exposition, it will be assumed that it represents local characteristics of the population to be treated and that individuals are homogeneous with respect to this parameter whose realisation has then to be announced only once.

Period 2: After considering the cost and the quality of the treatment offered by both hospitals, the Health Authority decides whether to leave the management run the hospital or if a change is necessary. The choice among the two alternatives is based on a relative performance evaluation of the two hospitals.

In this environment each hospital has the alternative to cheat on the true realisation of his β and that of developing a cooperation among hospitals in order not to reveal its private information. The choice of cheating on q^* is in fact not viable. If the hospital lowers its level of quality, some people will decide not to be cured and this is an element that we have assumed HA can observe. In this model we will consider the simplest case in which there is no coordination among the hospitals; the decision of each of them on whether to cheat is the solution to a straightforward Nash game.

To develop our model we will use the same basic assumptions presented in section 2 as concerns the cost function, the utility of the hospital and the ability of the patients to recover quickly. For the specific application to yardstick competition it will be also assumed that the parameter shock is not correlated among periods and that all the other costs and functions do not change from the two periods.

The parameters on the ability of the patient to recover quickly are specific to each hospital, but there is a degree of correlation r among them, so that the observation of a parameter in hospital a makes the realisation of the same level more probable in the other hospital. In particular, we can define the joint probabilities of the event β_l β_h as:

$$\pi(\beta_h^i, \beta_h^j) = (1-p)(1-p(1-r))$$
$$\pi(\beta_l^i, \beta_l^j) = p[(1-(1-p)(1-r)]$$
$$\pi(\beta_l^i, \beta_h^j) = \pi(\beta_h^i, \beta_l^j) = p(1-p)(1-r)$$

An opportunistic behaviour for hospitals can be defined as follows: in both periods they declare a value for β equal to β_h and they get a payment and extra utility correlated with this behaviour. However, the possibility of cheating, and hence the payoff in the second period, is subordinated to being confirmed. The possibility of being reconfirmed depends on the value of the shocks in A and B and on the behaviour of both hospitals.

[11] Or it might have a profit and being inefficeint.

First of all, it is necessary to define the payoff the hospital receives in each period by his cheating. From section 2.1 we can recall that the expected payoff for cheating is equal to $pay^{ch} = p[f(q^* + \beta_h - C_h^*) - f(q^* + \beta_l - C_h^*)]$

The outcome of the game and the payoff of each hospital depends on the realisation of β and on their behaviour. Since each hospital has just two strategies (cheat or not cheat), the are four possible outcomes for the game that can be summarised as follows:

<u>a) Both hospitals do not cheat in period one</u>

In this case, the hospital reveals the true recovery parameter in the first period while in the second period he can cheat. However, due to the rules of the game, the hospital management might not be re-elected even if he tells the truth. The probability of being re confirmed depends on the cost revealed by the other hospital. If both hospitals tells the truth, the probability for the hospital management of not being re-confirmed depends on the probability of the event: one hospital has a low recovery rate while the other has a high recovery rate which can be written as:

$$\phi_i = 1 - \pi(\beta_h^i, \beta_l^j) = 1 - p(1-p)(1-r) \leq 1$$

As we can note, being honest does not always assure to the hospital management its being reconfirmed and this is a serious disincentive to reveal the true state of the world that has occurred. The payoff for hospital A can be written as:

$$E(pay) = 0 + \sigma[1 - p(1-p)(1-r)]pay^{ch}\frac{N}{2}$$

and similarly for hospital B.

<u>b) Both hospitals cheat in period one</u>

The policy of cheating in period one corresponds to announcing that the patient is a low recovery also when this is not the case. Since both hospitals announce the same β, both management will be confirmed, so that the expected payoff for both hospitals can be written as:

$$E(pay) = \frac{N}{2}(pay^{ch} + \sigma pay^{ch}) = \frac{N}{2}pay^{ch}(1+\sigma)$$

<u>c) Just one of the two hospitals cheat</u>

Let us assume that hospital A cheats and always declares that the patients are low recovery. The hospital receives a higher utility in the first period with probability *p*, but the management might not be re-confirmed if the other hospital declares that his patients were high recovery. The probability of being reconfirmed is then equal to:

$$q_i = \pi(\beta_h^i, \beta_h^j) + \pi(\beta_l^i, \beta_h^j) = 1 - p$$

and the expected payoff is equal to:

$$E(pay) = \frac{N}{2} pay^{ch} + \sigma(1-p)\frac{N}{2} pay^{ch} = \frac{N}{2} pay^{ch}(1+\sigma(1-p))$$

Hospital B in this case decides to reveal the true recovery rate of his patients in the first period and, since hospital A always declares that his patients are low recovery, his management will always be re confirmed. The payoff for hospital B can then be written as:

$$E(pay) = 0 + \sigma \frac{N}{2} pay^{ch}$$

The interactions between the two hospitals can now be described in a matrix that presents the possible outcomes of the game.

Table 1: The possible outcomes of the game

		B	
		Not cheat	Cheat
A	Not cheat	A: $\sigma[1-p(1-p)(1-r)]\frac{N}{2} pay^{ch}$ B: $\sigma[1-p(1-p)(1-r)]\frac{N}{2} pay^{ch}$	A: $\sigma\frac{N}{2} pay^{ch}$ B: $\frac{N}{2} pay^{ch}(1+\sigma(1-p))$
	Cheat	A: $\frac{N}{2} pay^{ch}(1+\sigma(1-p))$ B: $\sigma\frac{N}{2} pay^{ch}$	A: $\frac{N}{2} pay^{ch}(1+\sigma)$ B: $\frac{N}{2} pay^{ch}(1+\sigma)$

From table 1, it is easy to show that the opportunistic outcome, with both hospital declaring that the patients are high cost, is always a Nash equilibrium unless p=0, in which case playing an opportunistic game would have the same effect as not cheating; if the probability of having high recovery patient is equal to zero, the model in fact collapses to a game with certainty and symmetric information.

The negative outcome of the solution of this game is determined by the fact that revealing the truth does not get to the hospital the possibility of extracting his rent in the second period while at the same time cheating, even when this is an unilateral decision, cannot be always punished. In order to make the hospital play in a non opportunistic way, it is necessary to give more incentive to the management to reveal the truth either by using fines if the cheating is discovered or by rewards if the truth is told. In the first case we can observe that if a fine equal to $[1-\sigma p(1-p(1-p)(1-r)\frac{N}{2}]$ was imposed, not cheating might become a weak Nash equilibrium.

From the point of view of our analysis, the reward schemes that can be used are more interesting. One possible way to incentivate the management to tell the truth is to merge the hospitals allowing the management of the more productive one to extract in the second period the rent of both hospitals. In this game, if the hospitals merge together, HA knows that in the second period there will not be any possibility to observe, even indirectly β, hence the scheme on which to reimburse the hospital will have to be the incentive compatible one. The expected payoff for the hospital can then be written as:

$$E(pay) = pay_1 + \sigma\, q(z\frac{N}{2}pay^{ch} + (1-z)Npay^{ICC})$$

where z represents the probability that the other management is reconfirmed and $pay^{ICC} < pay^{ch}$. The possible outcomes of the new game are presented in table 2

Table 2: The possible outcomes of the game with takeover

		B	
		Not cheat	Cheat
A	Not cheat	A: $\sigma[1-p(1-p)(1-r)]$ $\{\frac{N}{2}pay^{ch}(1-p(1-p)(1-r)+Npay^{ICC}p(1-p)(1-r)\}$ B: $\sigma[1-p(1-p)(1-r)]$ $\{\frac{N}{2}pay^{ch}(1-p(1-p)(1-r)+Npay^{ICC}p(1-p)(1-r)\}$	A: $\sigma[\frac{N}{2}pay^{ch}(1-p)+pNpay^{ICC}]$ B: $\frac{N}{2}pay^{ch}(1+\sigma(1-p))$
	Cheat	A: $\frac{N}{2}pay^{ch}(1+\sigma(1-p))$ B: $\sigma[\frac{N}{2}pay^{ch}(1-p)+pNpay^{ICC}]$	A: $\frac{N}{2}pay^{ch}(1+\sigma)$ B: $\frac{N}{2}pay^{ch}(1+\sigma)$

As in the previous case, the incentive alone is not sufficient to make the non opportunistic behaviour be a Nash equilibrium. A combined system using a penalty if the management is suspected of cheating has to be used. This policy would however have another important extra cost: the merger of the two hospital would have another important cost for HA. The marginal patients going to the hospital that is going to be closed will no longer accept health care since the gain in terms of utility becomes negative. To assure that all the patients accept to be admitted to hospital, the quality has to be raised to $\varphi^{-1}(\frac{3}{4}T)$ or both hospitals have to be closed and one located at ½ has to be opened supplying health care with quality $\varphi^{-1}(\frac{1}{2}T)$. It is clear that both options have a great cost in terms of additional resources; in both cases, if the differential between the cost for high and low recovery patients is not very high, the best policy for HA might be that of leaving hospital free to cheat.

2.2.3 Direct management.

By using this policy, HA decides to manage directly one of the two hospitals in order to get better information on the realisation of β. The policy of running the hospital directly might have consequences on the cost of care that we will not explicitly consider. Our analysis will be focused on the gain in the incentive mechanisms that can achieve through an imperfect observation of the recovery rate. In the yardstick competition model, the observation of the parameter β makes the hospital that is not under the direct management of the hospital reveal the truth if the correlation between β_l and β_h is equal to one and there is no discount rate, i,e. σ=0. In this environment, in fact, the behaviour of the hospital that cheats is always found out and the gain in the first period has to be offset by the loss in gain in the second period. It is interesting to note that if there is a discount rate, even being found out cheating and not being reconfirmed is a sufficient incentive to reveal the truth. In this case, in fact, the payoff for cheating in the first and the second period are the same and by not cheating the hospital has to give up the payoff in the first period. If we consider the model with a fine for the management that is found cheating, if the fine is at least equal to $[1-\sigma p(1-p(1-p)(1-r)\frac{N}{2}]$, the hospital will not have any incentive to cheat and it will always reveal its private information.

If we consider the classical model with asymmetry of information, the new incentive scheme that has to be given to the hospital to reveal the true β depends on the conditional probability of β once this parameter has been observed for the other hospital. The incentive compatible schemes will then be two depending on the observed β for A. The problem for HA can then be written as:

$$
\begin{aligned}
&Min \quad z_{A=i}t_l + (1-z_{A=i})t_h \qquad\qquad i = l,h \\
&s.t. \\
&C_i = q^* + \beta_i - e_i \qquad\qquad i = l,h \\
&t_i - C_i - f(e_i) \geq 0 \\
&t_i - C_i - f(e_i) \geq t_j - C_j - f(e_{ij})
\end{aligned}
$$

where $z_{A=i}$ is the conditional probability of a high/low recovery parameter for hospital B given that in A the observed parameter was high/low. The solution for the game is similar to the one presented in appendix 2 for the classical incentive compatible problem.

$$
\begin{aligned}
&f'(q^* + \beta_l - C_l) = 1 \\
&f'(q^* + \beta_h - C_h) = 1 - z_{A=i}[1 - f'(q^* + \beta_l - C_h)] < 1 \\
&t_l^{**} = C_l^* + f(q^* + \beta_l - C_l^*) + [f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**})] \\
&t_h^{**} = C_h^{**} + f(q^* + \beta_h - C_h^{**}) \\
&U_l = f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**}) \\
&U_h = 0
\end{aligned}
$$

In this case, it is interesting to note the role played by z in determining the two incentive schemes. z in fact assumes the following two values:

$$
\begin{aligned}
&z_{A=l} = p + (1-p)r \\
&z_{A=h} = p(1-r)
\end{aligned}
$$

In the first case, having observed β_l for A, this event is more probable than β_h and for this reason the effort in this occurrence is reduced, C_h^{**} increases, but the incentive given to B for his information rent is reduced.

In the second occurrence, β_h increases its probability of occurring and for this reason, the effort in this state is increased, hence increasing the efficiency of the game. Of course, the cost is reduced and the incentive for the first occurrence is increased, but this event should occur with a lower probability. The use of information on one hospital has then the advantage to tailor the incentive to the other hospital, but do not avoid the problem of cheating.

4. Discussion

The model presented in the previous section shows that even if we assume that quality can be indirectly controlled through the behaviour of patients, designing a contractual form that allows to reach an efficient allocation of resources is rather difficult. The problem arises from the presence of asymmetry of information as regards the cost of health care. Health care is in fact an input itself in the process of patient's recovery and the relationship between health care and health gains is not observable. The reforms that have affected health care in the recent past seem to be oriented towards the introduction of mechanisms to mimic private competitive markets. The first step in this direction was the separation of purchasing and providing activities with the creation of specific agency devoted to these functions. The first implementations of the internal market have not been as successful as expected for a number of reasons and as a result, Central Government has tried to introduce more competition in the purchasing and providing side of the market[12]. However, the problem seems to be more complicated than these naive formulations might let us think. The comparison of costs in different areas, an instrument that has been widely used for benchmarking and yardstick competition, does not seem to be an effective instrument for giving incentives and sanctions to hospitals. Cost might vary even with the same case mix simply because the procedures used to treat the patients are different[13]. Even if it was possible to compare costs, it would be impossible to make hospitals go bankrupt and close because the negative consequences of this policy would fall on the patients rather than on the inefficient management. Yardstick competition in this context does not seem to be a feasible alternative: its use in fact has to be limited to the process of re confirming the management of the hospital and the results, in line with the analogous literature on local governments, show that designing an effective incentive scheme in this context is rather difficult and very costly The feasible alternatives are the use of a classical incentive compatible scheme whose cost, in terms of welfare is lower than yardstick competition. The expected payoff for the hospital, which correspond to the extra cost the Health Authority has to incur is equal to $\frac{N}{2}pay^{ICC} = \frac{N}{2}p[f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**})]$ in each period. To compare the payoff with the one that can be obtained using yardstick competition we can use the same discount factor introduced for that model so that the expected payoff over two periods can be written as $\frac{N}{2}pay^{ICC}(1+\sigma)$. For yardstick competition, with or without the incentive of closing the other hospital[14], the payoff of each hospital is equal to $\frac{N}{2}pay^{ch}(1+\sigma)$ which is greater than the incentive compatible payoff because $pay^{ch} > pay^{ICC}$. The difference $\frac{N}{2}(1+\sigma)[pay^{ch} - pay^{ICC}]$ represents a straight gain in terms of welfare because both models assumes the same level of quality of health care.

[12] One might think to the U.K. system were the function of purchasing health care was left to health agencies (DHA) in competition with GP's practice (GP fundholders). In Italy a degree of competition is introduced between the public and the private sector, both being able to supply public health care.

[13] There are a number of treatments that can be supplied in a day hospital form or as normal admission. The use of these alternative techniques is often determined by the experience and competence of the medical staff.

[14] The solution is the same since we have noted in the previous section that the incentive is not sufficient to make hospital not cheat. If both hospitals cheat, the management will be confirmed and they will share the market.

The direct management of one hospital requires in general less resources to be spent in payoffs for the hospital that is not run by the Health Authority. With an incentive compatible scheme, direct management of one hospital allows to reduce the payoff for the hospital; in this case in fact the incentive scheme is based on the conditional probability of the event after having observed the recovery parameter for the other hospital. Direct management of one hospital and yardstick competition is not a welfare improving strategy since this scheme is not effective for avoiding the cheating; if we use the model of yardstick competition with a fine on the hospital that cheats, the feasible solution of the Nash game becomes for the hospital that is not directly controlled to tell the truth[15].

As we noted in the introduction, quality and cost issues can be separated using patients migration patterns which allows to define the contract of each hospital in terms of number of cases admitted. As we noted at the end of section 2, it was argued that quantity caps to hospital reimbursement schemes should be introduced to control indirectly for quality. However, the use of a cap might prove not to be optimal in an asymmetric context. The use of the instrument is justified to avoid a two tier system with hospital competing for low cost patients coming from the neighbour region, but might create dumping in the catchment area. If the information structure is such that the hospital can observe the cost before offering the treatment, a strict cap on quantity might make the hospital dump high cost patients in his area in order to attract low cost from the neighbourhood.

4. Conclusions

The article presented here sets up a model combining the rules of the spatial competition à la Hotelling with several incentive compatible contracts in order to set a scheme allowing to find an efficient (cost minimisation) mechanism to define the cost for treatment in hospital.

Consumer can observe the quality of the care they receive, hence they will choose to go to the hospital that maximises the difference between quality and travelling cost. Hospitals behaviour is determined by the maximisation of the utility of their management which depends on the revenue they receive minus costs. Costs depends on the quality of the services they produce and on the ability of the patient to recover quickly. This information is private to the hospital and determines asymmetry of information.

Every hospital is paid by a Government agency that is allocated from Central Government a specific budget for health care. The Agency wants to minimise cost for supplying the treatment to all the patients, but cannot observe quality of the service or the cost. In this environment the

[15] In this case in fact the hospital knows that the Heath Authority directly observes the relevant parameter for the hospital that it run which then cannot cheat. As it was shown in the text, if one hospital does not cheat and a fine is

solutions proposed by the literature to set up incentive compatible contracts for the hospitals are examined. Quality is kept under control using information on patients choices; HA chooses the location of the hospitals and the quality of care in order to make the market just covered. This allows to make hospitals choose this minimum level because if quality falls short, some patients will choose not to be treated or they will go to the hospital they are not expected to and both behaviours can be observed by HA.

From the model it emerges that yardstick competition, at least in the way it can be thought for health care, has a poor performance. In the presence of an asymmetric shock, the hospital has an incentive to cheat and this result is in line with the most recent public choice literature on this subject. The only way to make not cheating a weak outcome of the game is to impose fines on top of firing the management and this solution might be quite difficult to be implemented. A straight incentive compatible contract seems to be preferred in this case. Direct management allows to reduce the cost to extract private information from the hospitals, but in practice this saving might be offset by the costs the Health Authority has to incur to manage one hospital directly.

imposed, the best policy is not cheating.

Appendix 1

The problem faced by Central Government can be written as:

$$Min \quad pt_l + (1-p)t_h$$

$$s.t.$$

$$C_i = q^* + \beta_i - e_i \qquad\qquad i = l, h$$

$$t_i - C_i - f(e_i) \geq 0$$

Central Government can act as a Stackelberg leader, the inequality for the reservation utility of the hospital can be taken as an equality. The two constraints can then be substituted in the objective function giving:

$$Min \quad Tr = p(q^* + \beta_l - e_l + f(e_l)) + (1-p)(q^* + \beta_h - e_h + f(e_h))$$

The F.O.C. for the problem can be written as:

$$\frac{\partial TR}{\partial e_l} \qquad p[f'(e_l) - 1] = 0$$

$$\frac{\partial TR}{\partial e_h} \qquad (1-p)[f'(e_h) - 1] = 0$$

giving:

$$d'(e_l) = d'(e_h) = 1$$

$$t_l^* = f(e_l^*) + C_l^*$$

$$t_h^* = f(e_h^*) + C_h^*$$

Appendix 2

$$
\begin{aligned}
&Min \quad pt_l + (1-p)t_h \\
&s.t. \\
&C_i = q^* + \beta_i - e_i \qquad\qquad i = l,h \\
&t_i - C_i - f(e_i) \geq 0 \\
&t_i - C_i - f(e_i) \geq t_j - C_j - f(e_{ij})
\end{aligned}
$$

In this case, the problem has to be solved in terms of observable variables such as the cost and not in terms of effort *e* that only the hospital can observe. From the first constraint we can derive that $e_i = q^* + \beta_i - C_i$. The third constraint in the problem, the so called Incentive Compatible Constraint can be written as:

$$
\begin{aligned}
&t_l - C_l - f(q^* + \beta_l - C_l) \geq t_h - C_h - f(q^* + \beta_l - C_h) \\
&t_h - C_k - f(q^* + \beta_h - C_h) \geq t_l - C_l - f(q^* + \beta_h - C_l)
\end{aligned}
$$

The second order conditions on the disutility of the effort allows us to conclude that the second inequality is always satisfied. Let us now observe the first inequality. It states that the net payment to the hospital in the best states of the world has to be at least equal to the payment received in the worst state of the world plus a compensation for the disutility of the effort. Let us now observe the participation constraint. In the worst state of the world, the hospital receives a compensation which is equal to $t_h = C_h + f(q^* + \beta_h - C_h)$. Let us now observe the l.h.s. of equation (). We can observe that $f(q^* + \beta_l - C_h) < f(q^* + \beta_h - C_h)$, hence the utility received in the best state of the world is greater than zero, which in turns means that the first participation constraint is always satisfied. With all this in mind we can write the minimisation problem as:

$$
\begin{aligned}
Min \quad TR = \; & p\{C_l + f(q^* + \beta_l - C_l) + [f(q^* + \beta_h - C_h) - f(q^* + \beta_l - C_h)]\} \\
& + (1-p)[C_h + f(q^* + \beta_h - C_h)]
\end{aligned}
$$

The F.O.C. for the problem can be written as:

$$
\frac{\partial TR}{\partial C_l} = p[1 - f'(q^* + \beta_l - C_l)]
$$

$$
\frac{\partial TR}{\partial C_h} = p[f'(q^* + \beta_l - C_h) - f'(q^* + \beta_h - C_h)] + (1-p)[1 - f'(q^* + \beta_h - C_h)]
$$

giving as solution:

$$
\begin{aligned}
&f'(q^* + \beta_l - C_l) = 1 \\
&f'(q^* + \beta_h - C_h) = 1 - p[1 - f'(q^* + \beta_l - C_h)] < 1 \\
&t_l^{**} = C_l^* + f(q^* + \beta_l - C_l^*) + [f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**})] \\
&t_h^{**} = C_h^{**} + f(q^* + \beta_h - C_h^{**}) \\
&U_l = f(q^* + \beta_h - C_h^{**}) - f(q^* + \beta_l - C_h^{**}) \\
&U_h = 0
\end{aligned}
$$

The two reimbursement schemes can be written as:

$$t_l^{**} = C_l^* + [f(q^* + \beta_h - C_h^{**})] + f(q^* + \beta_l - C_l^*) - f(q^* + \beta_l - C_h^{**})]$$
$$t_h^{**} = C_h^* + f(q^* + \beta_h - C_h^{**})$$

which allows to transform the payment in a optimal linear scheme of the type:

$t = \alpha + \delta_1 C_l + \delta_2 (C_d - C_l)$ by considering that:

$$\alpha = [f(q^* + \beta_h - C_h^{**})] + f(q^* + \beta_l - C_l^*) - f(q^* + \beta_l - C_h^{**})]$$
$$\delta_1 = 1$$
$$\delta_2 = 1 - \frac{f(q^* + \beta_l - C_l^*) - f(q^* + \beta_l - C_h^{**})}{C_l^* - C_h^*}$$

Notation

A = Hospital A
B = Hospital B
a = location on the line of length 1 of hospital A
b = location on the line of length 1 of hospital A
T = transport cost
D= distance
C_i = cost to provide health care to patient of type i
β_i = factor affecting the ability of patients to take advantage of health care
e_i = state contingent effort of the management
q = quality of health care offered
U = utility of the management of the hospital
t_i= state contingent reimbursement scheme
p = probability that the patient is a high recovery (hence low cost)
r = correlation coefficient between β_i^A and β_i^b
π= probability of events correlated with shocks
ϕ= probability of the management of being reconfirmed

References

Besley, T. and A. C. Case (1995) Incumbent behaviour: Vote-seeking, tax setting and yardstick competition, American Economic Review, 85, 25-45

Chalkley, M. and Malcomson, J.H. (2000), Government Purchasing of Health Services, in Culyer, A.J. and Newhouse, J.P. (eds.), *Handbook of Health Economics*, North-Holland, 461-536.

Dawson, D. (1994) Costs and prices in the internal market: markets vs the Management Executive guidelines, *Working Paper n. 115*, Centre for Health Economics, York

Dawson, D. e D. Howart (1994) Competition and the law in the internal market, *mimeo*

Ellis, R. and T. Mc Guire (1986) Provider behaviour under prospective reimbursement, *Journal of Health Economics*, 5(2), 129-51

Gravelle, H. (1999) Capitation contracts: access and quality, *Journal of Health Economics*, 18, 315-340

Laffont, J.J. and J.Tirole (1993) *A Theory of Incentives in Procurement and Regulation, The MIT Press*, Cambridge, MA

Levaggi, R. (1998) Resources allocation in the internal market: whose preferences count?, *Studi Economici*, 65, 105-122

Levaggi, R. Optimal procurement contracts under a binding budget constraint, *Public Choice* , 101, n.1-2, 23-37

Levaggi, R. (1996) NHS contracts: an agency approach, *Health Economics*, 5, 341-352

Newhouse, J. (1996) Reimbursing health plans and health providers: efficiency in production vs. selection, *Journal of Economic Literature*, 34, 1236-63

Feld, L.P., Josselin, J.M., Rocaboy, Y (2000) Yardstick competition: A theoretical model and an empirical analysis for French Regions, Paper presented at Journées AFSE 2000, Marseille, May

Salop, S.C. (1979) Monopolistic competition with outside goods, *Bell Journal of Economics*, 10, 141-156

Salmon, P (1987) Decentralisation as an incentive scheme, Oxford Review of Economic Policy, 2, 24-43

Shleifer, A. (1985) A theory of yardstick competition, Rand Journal of Economics, 16, 319-327

Culyer, A.J., Maynard A.K., Posnett J.W. (eds) (1990) *Competing in health care*, Mc Millan Press, Basingstoke

Maynard, A.K. (1987), “Markets and health care” in Williams, A. (ed) *Health and Economics*, Basingstoke: McMillan Press