Società italiana di economia pubblica
XV CONFERENZA

DIRITTI, REGOLE, MERCATO
Economia pubblica ed analisi economica del diritto

Pavia, Università, 3 - 4 ottobre 2003

THE ROLE OF REPUTATION ON GP'S DECISIONS UNDER DIFFERENT CONTRACTUAL ARRANGEMENTS

MARINA CAVALIERI

Dipartimento di Economia e Metodi Quantitativi - Università di Catania

DIRITTI, REGOLE, MERCATO
Economia pubblica ed analisi economica del diritto

XV Conferenza SIEP - Pavia, Università, 3 - 4 ottobre 2003

pubblicazione internet realizzata con contributo della

COMPAGNIA di San Paolo

società italiana di economia pubblica

dipartimento di economia pubblica e territoriale – università di Pavia

THE ROLE OF REPUTATION ON GP'S DECISIONS UNDER DIFFERENT CONTRACTUAL ARRANGEMENTS

Marina Cavalieri
Dipartimento di Economia e Metodi Quantitativi
Università di Catania
E-mail: mcavali@unict.it

Very preliminary draft. Please, do not quote.

1. Introduction

We develop a theoretical model to analyse how the GP's behaviour is affected by reputational concerns under different payment contracts. There is one important issue that has been largely neglected by the literature until now. The relationship between GP and patient is often an ongoing one, with the latter expecting to search medical services from the same GP not just once but many times. The long-term nature of this relationship provides an important mechanism for maintaining a correct professional behaviour, namely the GP's concern for his future revenue (utility) through the effect on his reputation. The basic idea is that the GP may be deterred from exploiting his short-term advantage by the "punishment" in terms of reputation that reduces his long-term payoff. In short, a GP can face a situation in which there is a trade-off between either choosing to behave in such a way to obtain a certain payoff for a given period of time or taking advantage of the situation and obtaining a higher payoff during the first period but a lower one afterwards. The GP's intertemporal decision depends on the value he attaches to the present compared to the future, that is from the discount factor. The more the GP is patient, the more he will prefer that strategy that do not relentlessly exploit short-term gains.

In the literature, a distinction has been made between two different forms of reputation in health care (Chalkley and Malcomson, 2000): *reputation for characteristics* and *reputation for past behaviour*. The former refers to some characteristics that are innate to the provider or that, once acquired (in the form of investments), have a long-lasting positive impact on the quality of services. We believe that this typology of reputation has more relevance in the provision of secondary care than in the primary care context. The latter refers to the possibility of renewing or not a given relationship on the basis of the behaviour undertaken by the provider in the past. In this study, we apply this concept to the relationships between patients and GPs and not to the contracts between public purchasers of health services and providers.

The role of reputation as an instrument that induces providers to offer a quality above the minimum under the threat of a contract disruption, has been analyzed by Montefiori (2002). The study uses an infinitely repeated game where purchasers try to infer from the provider's past choices his future behaviour. In a previous work, Montefiori (2001) investigates the possibility that patients' demand is influenced by the provider's reputation. The framework is however different from ours. Here, the provider is a hospital whose reputation at a given time depends on its previous reputation, the

real quality provided, the number of patients treated and the expenditure on advertising. The author uses a dynamic equation to describe how reputation builds up over time.

This work is also related to the many analyses on the physician's response to different forms of contracts. Ma (1994), Ma and McGuire (1997) and Chalkley and Malcomson (1998) have all considered the issue within an agency framework. In the first two models, patients' demand reflect the present quality of care but none of them consider explicitly the existence of a repeated relationship between provider and patient. Moreover, our model departs from these since we have neglected the issue of how a payment system may affect the provider's cost reducing efforts. The emphasis here is not upon cost-quality trade-offs. In this respect, our approach is closer to the recent contribution of Jelovac (2001), where she designs an incentive scheme for an opportunistic physician who makes diagnoses and provides treatments. Like our setting, she examines a context of double moral hazard: with hidden action (hidden diagnosis effort) and with hidden information (hidden treatment type). However, at least one major aspect distinguishes her analysis from ours. Jelovac does not allow the patient to choose another physician than the one he is assigned to (no demand reaction exists).

The paper proceeds as follows. The basic analytical framework is set out in section 2. Section 3 analyzes the impact of reputation on the GP's behaviour under a capitation system. In section 4, we consider the same issue but for a GP who is paid according to a FFS scheme. The different results are then compared in section 5. The final section concludes. Proofs are in the Appendices.

2. The framework

We consider an utility maximizing GP who practices for two periods of time. In the first period, an exogenous number of patients $(n_1 = \hat{n})$ is assigned to the GP. Patients registered with the GP are considered homogeneous with respect to all their characteristics but one: they suffer from an illness that can be either serious, requiring a medical treatment, or minor, requiring no treatment by the GP but only the adoption of some precautions by the patient himself.

Nature assigns patients to the seriously ill type with a given probability (p) which, for the sake of simplicity, is supposed to be equal to $1/2$. It is assumed that the GP is aware of this probability in the form of a common knowledge (e.g. prevalence of the illness in the population) but he does not know exactly which the serious cases are. Patients observe only the symptom but are not able to understand the severity of their illness. Therefore, they seek health care from the GP.

In accordance with the previous literature (Jelovac, 2001; Garcia Mariñoso and Jelovac, 2003; González, 2002), the GP's medical activity is supposed to consist of two different phases: that of diagnosis (which includes also an analysis of patient's case history and the provision of medical advice) and that of treatment. While performing the diagnosis, the GP exerts a certain level of effort, $e \in [0,1]$, which yields him a signal about the severity of the patient's condition. We should interpret this variable not only as the time but also as the diligence, care and attentiveness spent by the GP in conducting the diagnosis. Thus, the higher the effort, the greater the number of serious cases identified by the GP among all his patients. This proportion, $\mu(e)$, may be defined by the following explicit function:

$$\mu(e) = \frac{1+e}{2}(p) = \frac{1+e}{2}(1/2) = \frac{1+e}{4}$$

which is increasing in e, with $\mu(0) = 1/4,\ \ \mu(1) = 1/2,\ \ \mu'(\mathrm{e}) > 0$ and $\mu''(\mathrm{e}) = 0$. The assumption $\mu(0) = 1/4$ implies that, when exerting the lowest effort $(e = 0)$, the GP discovers by chance only one out of two individuals who actually suffer from the serious illness, or, equivalently, a quarter of all the patients registered with him. On the contrary, when exerting the maximum effort $(e = 1)$, the GP makes accurate diagnoses and is able to detect all the existing serious cases.

In performing the diagnosis, the GP bears some disutility of effort, denoted by $g(e)$, where $g(0) = g'(0) = 0,\ \ g'(e) \geq 0$ and $g''(e) > 0$. Because the effort is not contractible, the GP will not be directly reimbursed for it. Therefore, a problem of moral hazard with hidden action arises since the GP may have an incentive to undersupply effort in diagnosis.

After he has gathered information about the seriously ill patients, the GP provides them with a medical treatment (T)[1]. Treatments differ from each other with respect to their length. We indicate with $\underline{T} > 0,$ the minimum treatment (in terms of length) required to fully recover a seriously ill patient. Ethical concerns prevent the GP from supplying a level of treatment less than the minimum one and greater than an upper bound $\overline{T}$[2]. It is supposed that $\underline{T} \neq \overline{T}$ and that the provision of a level of treatment higher than $\underline{T}$ does not add extra health benefits to patients but, depending on the existing remuneration system, only brings monetary advantages to the GP. Thus, a second problem of moral hazard occurs since the GP may be encouraged to exploit the patient information asymmetry for intensifying the treatment beyond the minimum level ("Supply Induced Demand" problem).

Nevertheless, treating patients is not without cost for the supplier. The monetary cost of treatment is assumed to be $c(T)$, which is increasing, differentiable and convex. It includes also the remuneration of the GP's labour (valued in terms of cost-opportunity). Fixed costs are not considered given that their presence is marginal in primary care.

Reputation enters in this model since patients' demand is supposed to respond to the quality of services provided by the GP. It is well known that the particular nature of the health service (that of an "experience good") does not allow patients to observe all the aspects of quality while receiving it. However, the existence of repeated physician-patient relationships as in the primary care context may help to partially overcome this problem. Once the treatment has been provided, it seems reasonable to assume that, although some dimensions of quality continue to be largely ignored, patients could judge the GP's work at least by whether or not they have recovered from the illness and by the length of the treatment received. Indeed, it is in the patients' interest not only to

---

[1] Once the seriously ill patients have been identified, we assume that GP's ethics preclude him from hide this private information for opportunistic reasons. Moreover, we exclude the possibility for the GP to refer patients to a specialist (Brasseur, 2000; Garcia Mariñoso and Jelovac, 2003). Thus, all the detected serious cases are treated by him.

[2] Alternatively, we could consider $\overline{T}$ as an external constraint imposed by the regulator either to control the unnecessary treatment costs due to the GP's opportunistic behaviour (especially under a FFS system), or to protect patients' health. Indeed, a too prolonged treatment could be harmful for the patient.

recover but also to do it as soon as possible, in order to minimize the opportunity costs of being ill. In reality, patients' recovery does not depend solely on the GP's decisions and actions but also on factors external to his control. Hence, patients not recovering from a certain illness do not know whether to blame their problems on physician's malfeasance or simply on bad luck. However, no role exists in this model for this kind of uncertainty and patients still ill are supposed to automatically infer that the GP has "cheated".

During period 2, patients suffer again from the same illness and the GP has to perform new diagnoses and to provide the seriously ill ones with new treatments. The demand reaction in the second period may be formalized as $n_2(\mu(e_1), T_1)$, where n_2 is differentiable, increasing in the proportion of serious cases detected by the GP during the first period and decreasing in the level of the treatment previously provided. Thus, whenever a poor service is supplied during the first period (low $\mu(e_1)$ and/or high T_1), patients may decide to move to another GP, and may discourage other potential patients from choosing that provider. In both cases, the GP's opportunistic behaviour is punished by a decrease in his future income. On the opposite, an high value of $\mu(e_1)$ together with a low value of T_1 will allow the GP to retain his patients and even to attract new ones ("word of mouth" effect).

Concerning the preferences, we impose that the GP is risk neutral with respect to money, has a utility function separable in money and effort, and a reservation utility level normalized to zero. Contrary to other models (Dionne and Contandropoulos, 1985, among others), we do not include patients' welfare in the GP's utility[3]. However, altruism in medical activity may certainly help to alleviate the moral hazard problems.

The aim of the following sections is that of describing the GP's behaviour under different remuneration systems both in presence or not of reputational effects. Therefore, the analysis of the mentioned reputational model is presented alongside with the analysis of a more simple scheme, where patient's demand do not react to the GP's previous choices and reputation does not matter. Two different "pure" types of payment contracts are considered: a capitation system and a FFS system. A salary-based mechanism is ignored since, under it, no role exists for reputation. Whatever decision the GP takes during the first period, a rise in patients' demand does not increase his payment.

---

[3] There is no consensus in the literature about how to take account of the role of physician's altruism or ethics. Ma and McGuire (1997) assume that the physician must provide health benefits above a certain threshold (given a health shock). This is the approach that we have used in this model, where the GP provides as minimum level of treatment, the one required to fully recover a seriously ill patient. An alternative way (Chalkley and Malcomson, 1998) is adding to the physician's utility a benevolent component $z(q)$, assumed to be non-decreasing in quality and concave. For a fully self-interested physician $z(q)$ is equal to zero. For a completely benevolent physician $z(q) = b(q)$, that is equal to the patient's benefit. This formulation allows for intermediate levels of benevolence, where $z(q) = \alpha b(q)$, with $0 < \alpha < 1$.

3. The GP's behaviour under a capitation contract

In this section, we assume that the health authority pays the GP according to a capitation system. Therefore, the GP annually receives a fixed amount of money (R) for each patient registered with his practice. The GP must meet the monetary cost of treatment out of this payment and can retain any residual.

Let us first consider the GP's effort and treatment decisions when reputation does not play any role, that is when the number of patients is supposed to be constant in each period and equal to $\hat{n}$. Since the GP's utility in the second period does not depend on the choices made in the first period (i.e. there is no demand reaction), maximizing the overall utility over the two periods is equivalent to maximizing each period's utility. Thus, the GP's problem may be simply reduced to:

$$\underset{e_t, T_t}{Max}\ U_t(e_t, T_t) = R\hat{n} - g(e_t)\hat{n} - c(T_t)T_t\mu(e_t)\hat{n} \qquad (t = 1, 2)$$

$$\text{s.t.} \quad \begin{aligned} -T_t &\leq -\underline{T} \\ T_t &\leq \overline{T} \\ e_t &\leq 1 \end{aligned}$$

$$\text{and} \quad e_t \geq 0$$

where the subscript t refers to each of the two periods of time, $R\hat{n}$ is the total revenue the GP receives by the health authority for providing care to all the patients in his list, $g(e_t)\hat{n}$ denotes the total disutility of effort when performing $\hat{n}$ diagnoses and $c(T_t)T_t\mu(e_t)\hat{n}$ is the total cost of providing the treatment only to those seriously ill patients identified by the GP through the diagnosis process.

The GP's decisions are provided by the following proposition:

Proposition 1. *Under a capitation contract and in absence of reputational effects, the GP's optimal effort and treatment decisions for each period imply:*

- *to exert the minimum level of effort* $(e_t^* = 0)$;
- *to provide seriously ill patients with the minimum treatment* $(T_t^* = \underline{T})$.

Proof. See Appendix 1.

These results may be intuitively explained. Indeed, any level of effort higher than the minimum, which is here normalized at zero, increases both the GP's non-monetary and monetary costs; the latter increase because of the rise in the proportion of serious cases that the GP is able to discover and, consequently, has to treat. With reference to the treatment decisions, once the seriously ill patients have been identified, treating them longer than the necessary gives only rise to extra costs. In return for these higher

costs, the GP is not entitled to any additional remuneration, which remains fixed for that given number of patients $(\hat{n})$.

By choosing $e_t^* = 0$, the GP does not bear any disutility of effort and discovers and treats half of the patients who actually suffer from the serious illness (or, equivalently, a quarter of all the patients in his list). Obviously, we are assuming here that the patients who are not cured by the GP do not react to this fact by choosing another doctor in the second period. We could either think that these patients are completely passive or imagine a situation where there is no real competition among GPs (e.g. the costs of shifting to another provider are very high). Consequently, the above optimal solutions are valid for each period and, more generally, are long-term ones. The GP's utility in each period is equal to:

$$U_t(0,\underline{T}) = R\hat{n} - \frac{1}{4}c(\underline{T})\underline{T}\hat{n} \qquad (\mathrm{t} = 1,2) \qquad (1)$$

which depends positively on the amount of the capitation fee and negatively on the average variable cost of providing the minimum treatment.

The above results are, however, somewhat largely explored by the existing literature (Ellis and McGuire, 1990 and 1993). Capitation contracts are supply-side cost sharing arrangements that give the GP incentives to reduce monetary costs in order to obtain a higher net payment. Therefore, the capitation system does not lead GP to misrepresent the patient's problem by providing unnecessary treatments. On the contrary, in absence of reputational effects, the problem becomes that of an underprovision of effort during the diagnosis[4], since gathering information about the patient's condition causes disutility to the GP but does not confer any present or future benefit on him.

Suppose now that patients' demand responds with a lag of one period to the GP's medical decisions. Then, the GP maximizes his utility over the two periods and his intertemporal optimization problem can be formulated as:

$$\underset{e_e, e_2, T_1, T_2}{Max}\ U(e_1, e_2, T_1, T_2) = U_1(e_1, T_1) + \beta U_2(e_1, e_2, T_1, T_2) = R\hat{n} - g(e_1)\hat{n} - c(T_1)T_1\mu(e_1)\hat{n} +$$
$$+ \beta(Rn_2(\mu(e_1), T_1) - \mathrm{g}(\mathrm{e}_2)n_2(\mu(e_1), T_1) - c(T_2)T_2\mu(e_2)n_2(\mu(e_1), T_1))$$

$$\text{s.t} \quad \begin{aligned} -T_1 &\le -\underline{T} \\ T_1 &\le \overline{T} \\ e_1 &\le 1 \\ -T_2 &\le -\underline{T} \\ T_2 &\le \overline{T} \\ e_2 &\le 1 \end{aligned}$$

[4] As mentioned, the risk of an underprovision of treatment, which is big under a capitation system, is avoided in this model by fixing a minimum level of treatment.

$$\text{and} \qquad e_1 \geq 0$$
$$e_2 \geq 0$$

where the parameter $0 < \beta \leq 1$ represents the GP's subjective discount factor and measures the extent to which the GP values future utility relative to current utility (i.e. how patient he is). This parameter may be also written as $\beta = 1/(1+\rho)$, where ρ indicates the GP's rate of time preference and varies from individual to individual. For most individuals, ρ is positive, which means that $(1+\rho)$ is greater than 1 and that greater weight is attached to the utility obtained in period 1 rather than in period 2. More rarely, it may be the case that the individual regards the current and future utility equally, in which case ρ takes the value of zero.

Before solving this problem of nonlinear programming, it is, however, possible to predict some of the expected results. Since the medical activity is supposed to end up with the second period, the GP will not have incentives to establish a reputation in the last time. Thus, the GP's optimal effort decision in period 2 will be exactly the same as the one found in case of a capitation contract without reputation ($e_2^* = 0$). Any other level of effort apart from zero, will imply for the GP additional costs (both monetary and in terms of disutility) that will not be compensated by future benefits in terms of additional demand. Furthermore, during the second period, seriously ill patients will receive the minimum treatment since only under this condition the GP will be able to save on costs.

More difficult is to foresee what the level of effort and treatment will be during the first period. Regarding the effort, two opposite effects must be taken into consideration by the GP. On one hand, the less the effort, the lower the costs (monetary and non) that the GP will have to face and the higher his utility in period 1; on the other, any decrease in the level of effort will result in some seriously ill patients not being correctly diagnosed and, thus, treated. These patients will be dissatisfied of the care received by their doctor and will presumably decide to deregister, giving rise to a reduction of the GP's revenue during the second period. Moreover, once the GP's reputation has worsened, other potential patients will be prevented from choosing him as their provider. However, the less the patients in period 2, the less the costs of providing the minimum treatment for those of them that are found seriously ill. Therefore, under the hypothesis that, during the second period, the loss of revenue deriving from a bad reputation is higher than the saving in treatment costs, the GP's effort decision becomes basically the choice between either obtaining a higher utility during the first period and a lower one afterwards, or the opposite. The final decision depends not only on the dimension of the two effects stated above but also on the GP's rate of time preference.

A similar argument applies to the choice of the treatment in the first period. In this case, a lower treatment will allow the GP both to reduce the current cost of providing care to the serious cases and to enhance his reputation. By satisfying his patients' expectations of a fast recovery, the GP will avoid the risk of patients moving to other providers and will even attract new patients, with a concomitant gain of revenue. Nevertheless, as the number of patients during the second period increases, the number of serious cases detected will raise too. Hence, the costs of providing them with the minimum treatment will be higher.

The next proposition summarizes the solutions to this maximization problem.

Proposition 2. *Under a capitation contract and in presence of reputational effects, the GP's optimal effort and treatment decisions over the two periods are as follows:*

- *to provide severe cases with the minimum treatment during the first period* $(T_1^* = \underline{T})$;

- *to exert the minimum level of effort during the first period* $(e_1^* = 0)$ *if:*

$$\beta \frac{\partial n_2}{\partial e_1}(R - \frac{1}{4}c(\underline{T})\underline{T}) \leq c(\underline{T})\underline{T}\mu'(0)\hat{n}; \tag{2}$$

- *to exert such a positive level of effort during the first period* $(0 < e_1^* \leq 1)$ *that:*

$$\beta \frac{\partial n_2}{\partial e_1}(R - \frac{1}{4}c(\underline{T})\underline{T}) = \hat{n}(g'(e_1^*) + c(\underline{T})\underline{T}\mu'(e_1^*)); \tag{3}$$

- *to exert the maximum level of effort during the first period* $(e_t^* = 1)$ *if:*

$$\beta \frac{\partial n_2}{\partial e_1}(R - \frac{1}{4}c(\underline{T})\underline{T}) > \hat{n}(g'(1) + c(\underline{T})\underline{T}\mu'(1)); \tag{4}$$

- *to exert the minimum effort during the second period* $(e_2^* = 0)$;

- *to provide severe cases with the minimum treatment during the second period* $(T_2^* = \underline{T})$.

Proof. See Appendix 2.

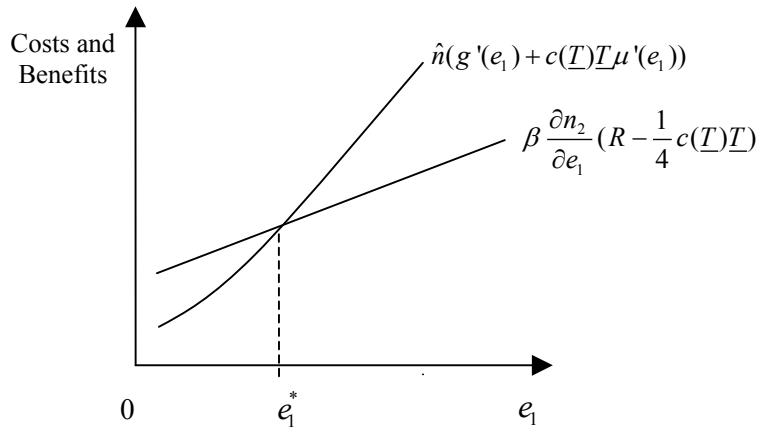
The most interesting results that can be extracted from Proposition 2 concerns the choice of the levels of treatment and effort during the first period. The former depends on both the tomorrow discounted net costs of increasing the treatment (i.e. the discounted difference between the loss of capitation fee due to the reduced number of patients and the savings in the costs of providing a quarter of them with the minimum treatment) and the today additional costs of supplying that treatment to the serious cases. Since the tomorrow discounted net costs must be always positive (i.e. the loss of revenue is greater than the savings) in order to satisfy the GP's participation constraint (see Appendix 2), it follows a strong incentive for the GP to keep down the overall costs by reducing the treatment to the minimum.

The GP's effort in period 1 is equal to zero only if the tomorrow discounted net benefit deriving from increasing the effort (that is the discounted difference between the additional capitation fees and the cost of providing a quarter of the additional patients with the minimum treatment) is not higher than the today monetary cost of exerting an effort (i.e. the cost of providing the minimum treatment for the additional serious cases).

Therefore, the GP finds convenient to minimize the effort to zero so as not to bear at least any disutility of it.

Viceversa, whenever equation (3) is fulfilled, the GP chooses to exert a positive level of effort, which can also be equal to the maximum. He increases the effort up to the level where the tomorrow discounted net benefit of exerting that effort is exactly equal to the today disutility and monetary costs (see figure 1). Going beyond this point, the GP is not able any more to maximize his overall utility which starts from there onward to decrease. A level of effort equal to 1 is also chosen by the GP whenever condition (4) is satisfied. In this case, at $e_1^* = 1$, the second period discounted net benefit of exerting the maximum effort still exceeds the first period non-monetary and monetary costs implied by that effort.

Figure 1 – Capitation contract with reputation: effort decision in period 1



The optimal level of effort depends strongly on the capitation fee. The lower the capitation fee, the lower the level of effort. Moreover, the case of a zero effort is more plausible when the subjective discount factor (β) is low (i.e. the GP is impatient) and the demand does not react heavily to an increase in effort (i.e. one or both the derivatives $\partial n_2 / \partial \mu$ and $\partial \mu / \partial e_1$ are low). All these changes determine a downward shifting of the curve of the tomorrow discounted net benefit in figure 1. On the contrary, a lower disutility of effort increases the GP's exertion while the effects of a decrease in the cost of providing the minimum treatment cannot be easily predicted.

Leaving undefined the optimal level of effort in period 1, the GP utility over the two periods of time assumes the following general form:

$$U(e_1, e_2, T_1, T_2) = U_1(e_1^*, \underline{T}) + \beta U_2(e_1^*, 0, \underline{T}, \underline{T}) = R\hat{n} - g(e_1^*)\hat{n} - c(\underline{T})\underline{T}\mu(e_1^*)\hat{n} + \\ + \beta(Rn_2(\mu(e_1^*), \underline{T}) - \frac{1}{4}c(\underline{T})\underline{T}n_2(\mu(e_1^*), \underline{T})) \quad (5)$$

where the term $g(e_1^*)\hat{n}$ becomes zero when the GP does not exert any effort during the first period.

By comparing the optimal solutions obtained under the two capitation models, we can conclude that either the absence or the presence of reputational effects does not influence the level of treatment which is always the minimum. The risk of an excessive treatment length (moral hazard on the supply side) is thus avoided simply because of the contract design, which makes the GP bear the full cost of his decision making. The presence of reputational effects can only restrain, under specific circumstances, the GP's opportunistic attitude towards an underprovision of effort in diagnosis.

4. The GP's behaviour under a fee-for-service contract

Under a FFS contract, the GP receives a reimbursement from the health authority in the form of a fee (F) for each unit of treatment (in terms of length) provided to his seriously ill patients. No remuneration is given for those patients that are not treated.

As in the case of a capitation contract, when there are no reputational effects and patients' demand does not respond to the GP's past choices, the optimization problem for each of the two periods of time becomes:

$$\underset{e_t, T_t}{Max}\ U(e_t, T_t) = FT_t\mu(e_t)\hat{n} - g(e_t)\hat{n} - c(T_t)T_t\mu(e_t)\hat{n} \qquad (t = 1,2)$$

$$\text{s.t.} \quad \begin{aligned} -T_t &\le -\underline{T} \\ T_t &\le \overline{T} \\ e_t &\le 1 \end{aligned}$$

$$\text{and} \quad e_t \ge 0$$

where the subscript t refers either to the first or to the second period of time and $FT_t\mu(e_t)\hat{n}$ is the total reimbursement the GP obtains from treating those serious cases (namely $\mu(e_t)$) identified among his patients.

From the above maximization problem, it can be gathered that exerting a positive level of effort implies for the GP to bear some disutility and, at the same time, increases the proportion of seriously ill patients detected during the diagnosis. Treating these additional cases gives rise to extra monetary costs but also to extra compensation. Therefore, in choosing the optimal level of e, the GP should weigh the monetary and non-monetary costs and the revenue deriving from exerting the effort. With regards to the treatment decision, once a given proportion of serious cases has been identified, providing them with a treatment higher than the minimum is costly. However, in return for these costs, the GP is entitled to a greater remuneration. Again, the GP has to balance these two conflicting effects.

The mathematical analysis helps to shed light on the possible results of this problem, that are summarized by the following proposition:

Proposition 3. *Under a FFS contract and in absence of reputational effects, the GP's optimal effort and treatment decisions for each period require:*

A. whenever $F = c(\underline{T})$,

- *to exert the minimum level of effort* $(e_t^* = 0)$ *and to provide seriously ill patients with the minimum treatment* $(T_t^* = \underline{T})$;

B. whenever $F > c(\underline{T})$,

- *to exert such a positive level of effort* $(0 < e_t^* \leq 1)$ *that:*

$$T_t^* \mu'(e_t^*)\hat{n}(F - c(T_t^*)) = g'(e_t^*)\hat{n}; \tag{6}$$

- *to exert the maximum level of effort* $(e_t^* = 1)$ *if:*

$$T_t^* \mu'(1)\hat{n}(F - c(T_t^*)) > g'(1)\hat{n}; \tag{7}$$

- *to provide severe cases with such a level of treatment* $(\underline{T} \leq T_t^* \leq \overline{T})$ *that:*

$$F = c(T_t^*) + c'(T_t^*)T_t^*; \tag{8}$$

- *to provide severe cases with the minimum treatment* $(T_t^* = \underline{T})$ *if:*

$$F < c(\underline{T}) + c'(\underline{T})\underline{T}\ ; \tag{9}$$

- *to provide severe cases with the maximum treatment* $(T_t^* = \overline{T})$ *if:*

$$F > c(\overline{T}) + c'(\overline{T})\overline{T}. \tag{10}$$

Proof. See Appendix 3.

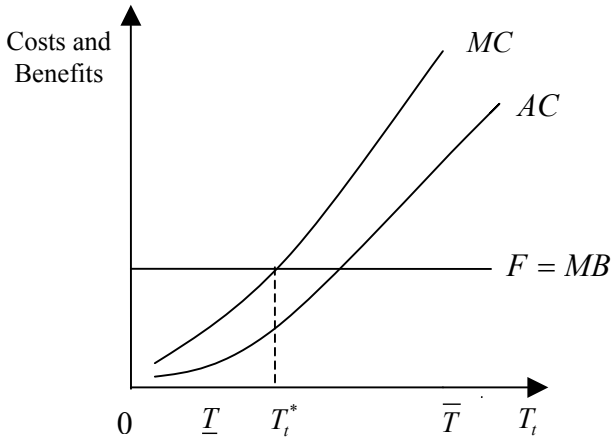
The interpretation of point A is quite clear. Whenever the fee is equal to the average cost of providing the minimum treatment $(F = c(\underline{T}))$[5], the GP does not find profitable to treat patients more intensively (i.e. the marginal benefit does not cover the marginal cost of treatment). Choosing the minimum treatment allows the GP to avoid at least the risk of a negative net revenue. Furthermore, since providing a positive level of effort is costly in terms of utility and does not increase the net income (the fee covers just the cost of the treatment provided to each of the additional serious cases), no effort is exerted in diagnosis. Therefore, the optimal effort and treatment decisions are exactly the same as the ones found under a capitation contract without reputation, except for the fact that the level of the GP's utility is now zero in both periods.

The second part of Proposition 3 states that, whenever $F > c(\underline{T})$, the GP chooses to increase the level of treatment up to that point $(\underline{T} \leq T_t^* \leq \overline{T})$ where the marginal benefit of providing it equals the marginal cost (see figure 2). Indeed, any other level of

[5] The case of $F < c(\underline{T})$ has been ruled out since it represents an unacceptable contract for the GP.

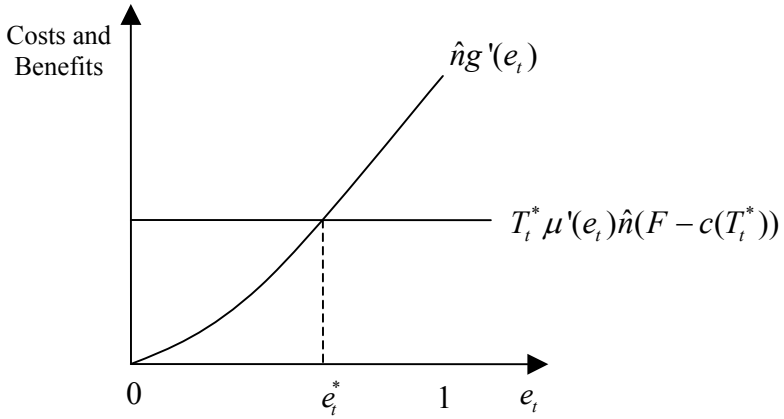
treatment implies for the GP a loss of revenue and, therefore, of utility. Optimal solutions are also to provide the minimum treatment if, at $\underline{T}$, the marginal benefit is less than the marginal cost and the maximum treatment if, at $\overline{T}$, the opposite situation arises.

Figure 2 – FFS contract without reputation: treatment decision



Concerning the effort, the GP opts for that positive level where the marginal cost of exerting it for all his patients equates the marginal net benefit (the difference between the fee and the average cost of treatment) of providing the selected treatment to the additional serious cases. Figure 3 shows a possible equilibrium. The maximum level of effort $(e_t^* = 1)$ is also chosen whenever equation (7) is satisfied, that is when providing the highest level of effort is still convenient for the GP in terms of utility.

Figure 3 – FFS contract without reputation: effort decision



All other things being equal, the combination of effort and treatment adopted by the GP depends on the amount of the fee-for-service (F), which thus affects the final

level of his utility. The higher the FFS, the higher, in general, the levels of both treatment and effort. Therefore, any attempt by the health authority to increase the fee in order to induce a higher effort, will also encourage the GP to intensify the treatment. Nevertheless, under specific circumstances, it is possible to overcome this trade-off and make the GP both exert the maximum level of effort and provide the minimum treatment. This happens when the fee is lower or equal to the marginal cost of providing the minimum treatment and such that the condition $\underline{T}\mu'(1)\hat{n}(F - c(\underline{T})) \geq g'(1)\hat{n}$ is satisfied. The problem is here that the marginal disutility of effort differs from individual to individual. Thus, any fee that guarantees the provision of the minimum level of treatment, will encourage some GPs to choose the maximum effort and others to opt for a lower level of effort. It has to be noticed that the level of effort also increases with the marginal productivity of it $(\mu'(e_t))$.

Whenever the number of patients during the second period depends on the professional reputation established by the GP during the first period, the optimization program may be transformed into:

$$\underset{e_1,e_2,T_1,T_2}{Max}\ U(e_1,e_2,T_1,T_2) = U_1(e_1,T_1) + \beta U_2(e_1,e_2,T_1,T_2) = FT_1\mu(e_1)\hat{n} - g(e_1)\hat{n} - c(T_1)T_1\mu(e_1)\hat{n} +$$
$$+ \beta(FT_2\mu(e_2)n_2(\mu(e_1),T_1) - g(e_2)n_2(\mu(e_1),T_1) - c(T_2)T_2\mu(e_2)n_2(\mu(e_1),T_1))$$

$$\begin{aligned} & -T_1 \leq -\underline{T} \\ & T_1 \leq \overline{T} \\ \text{s.t} \quad & e_1 \leq 1 \\ & -T_2 \leq -\underline{T} \\ & T_2 \leq \overline{T} \\ & e_2 \leq 1 \\ \\ \text{and} \quad & e_1 \geq 0 \\ & e_2 \geq 0 \end{aligned}$$

where the parameter $\beta = 1/(1+\rho)$ is the GP's discount factor $(0 < \beta \leq 1)$ and ρ denotes the subjective rate of time preference.

It is easy to predict that the GP's effort and treatment decisions during the second period will be similar to the ones obtained in absence of reputational effects (see Proposition 3). Since the GP is not expected to practice for a further period, his medical choices will not have any consequence on future demand and revenue. The GP's concern in period 2 will thus be only that of maximizing his current utility.

A different argument applies to the GP's choices during the first period. Generally speaking, exerting a positive level of effort requires the GP to bear some disutility during the first period but also allows him to discover a higher number of serious cases who have to be treated but for whom he receives an additional remuneration. Moreover, a positive level of effort contributes to strengthen the GP's reputation and increases patients' demand. A higher number of patients in the second period implies additional diagnoses and, thus, non-monetary costs. However, it also produces some monetary

benefits to the GP, insofar as the fee is higher than the cost of providing the serious cases with treatment.

Concerning the treatment, a level higher than the minimum increases the costs during the first period and, because of the worsening in the GP's reputation, gives rise to a loss of patients, and thus revenue, during the second period. On the opposite, it also entitles the GP to a higher compensation in the first period and allows him to save on the total disutility costs of the second period (less patients means less diagnoses). Clearly, the GP has to consider all the consequences of his actions before taking a decision.

The following proposition characterizes all the optimal solutions to this case:

Proposition 4. *Under a FFS contract and in presence of reputational effects, the GP's optimal effort and treatment decisions over the two periods are as follows:*

A. whenever $F = c(\underline{T})$,

- *to minimize in both periods the level of effort* $(e_1^* = e_2^* = 0)$ *and treatment* $(T_1^* = T_2^* = \underline{T})$;

B. whenever $F > c(\underline{T})$,

- *to exert the minimum level of effort during the first period* $(e_1^* = 0)$ *if:*

$$\mu'(0)\hat{n}T_1^*(F - c(T_1^*)) + \beta(T_2^*\mu(e_2^*)\frac{\partial n_2}{\partial e_1}(F - c(T_2^*)) \leq \beta g(e_2^*)\frac{\partial n_2}{\partial e_1} \qquad (11)$$

where, depending on the levels of T_1^* *and* T_2^**, the term* $\mu'(0)\hat{n}T_1^*(F - c(T_1^*))$ *may be negative;*

- *to exert such a positive level of effort during the first period* $(0 < e_1^* \leq 1)$ *that:*

$$\mu'(e_1^*)\hat{n}T_1^*(F - c(T_1^*)) + \beta(T_2^*\mu(e_2^*)\frac{\partial n_2}{\partial e_1}(F - c(T_2^*)) = g'(e_1^*)\hat{n} + \beta g(e_2^*)\frac{\partial n_2}{\partial e_1} \qquad (12)$$

where, depending on the levels of T_1^* *and* T_2^**, the term* $\mu'(e_1^*)\hat{n}T_1^*(F - c(T_1^*))$ *may be negative;*

- *to exert the maximum level of effort during the first period* $(e_1^* = 1)$ *if:*

$$\mu'(1)\hat{n}T_1^*(F - c(T_1^*)) + \beta(T_2^*\mu(e_2^*)\frac{\partial n_2}{\partial e_1}(F - c(T_2^*)) > g'(1)\hat{n} + \beta g(e_2^*)\frac{\partial n_2}{\partial e_1} \qquad (13)$$

where, depending on the levels of T_1^* *and* T_2^**, the term* $\mu'(1)\hat{n}T_1^*(F - c(T_1^*))$ *may be negative;*

- *to provide severe cases, during the first period, with such a level of treatment* $(\underline{T} \leq T_1^* \leq \overline{T})$ *that:*

$$\mu(e_1^*)\hat{n}(F - c(T_1^*) - c(T_1^*)T_1^*) - \beta g(e_2^*)\frac{\partial n_2}{\partial T_1} = -\beta(\mu(e_2^*)\frac{\partial n_2}{\partial T_1}T_2^*(F - c(T_2^*)) \quad (14)$$

where, depending on the level of T_1^* *and* T_2^**, the term* $\mu(e_1^*)\hat{n}(F - c(T_1^*) - c(T_1^*)T_1^*)$ *may be positive, null or even negative;*

- *to provide severe cases with the minimum treatment during the first period* $(T_1^* = \underline{T})$ *if:*

$$\mu(e_1^*)\hat{n}(F - c(\underline{T}) - c(\underline{T})\underline{T}) - \beta g(e_2^*)\frac{\partial n_2}{\partial T_1} < -\beta(\mu(e_2^*)\frac{\partial n_2}{\partial T_1}T_2^*(F - c(T_2^*)) \quad (15)$$

where, depending on the level of T_2^**, the term* $\mu(e_1^*)\hat{n}(F - c(\underline{T}) - c(\underline{T})\underline{T})$ *may be positive, null or even negative;*

- *to provide severe cases with the maximum treatment during the first period* $(T_1 = \overline{T})$ *if:*

$$\mu(e_1^*)\hat{n}(F - c(\overline{T}) - c(\overline{T})\overline{T}) - \beta g(e_2^*)\frac{\partial n_2}{\partial T_1} > -\beta(\mu(e_2^*)\frac{\partial n_2}{\partial T_1}T_2^*(F - c(T_2^*)) \quad (16)$$

where, depending on the level of T_2^**, the term* $\mu(e_1^*)\hat{n}(F - c(\overline{T}) - c(\overline{T})\overline{T})$ *may be positive, null or even negative;*

- *to exert such a positive level of effort during the second period* $(0 < e_2^* \leq 1)$ *that:*

$$T_2^*\mu'(e_2^*)n_2(F - c(T_2^*)) = g'(e_2^*)n_2; \quad (17)$$

- *to exert the maximum level of effort during the second period* $(e_2^* = 1)$ *if:*

$$T_2^*\mu'(1)n_2(F - c(T_2^*)) > g'(1)n_2; \quad (18)$$

- *to provide severe cases with such a level of treatment during the second period* $(\underline{T} \leq T_2^* \leq \overline{T})$ *that:*

$$F = c(T_2^*) + c'(T_2^*)T_2^*; \quad (19)$$

- *to provide severe cases with the minimum treatment during the second period* $(T_2^* = \underline{T})$ *if:*

$$F < c(\underline{T}) + c'(\underline{T})\underline{T}; \quad (20)$$

- *to provide severe cases with the maximum treatment during the second period* $(T_2^* = \overline{T})$ *if:*

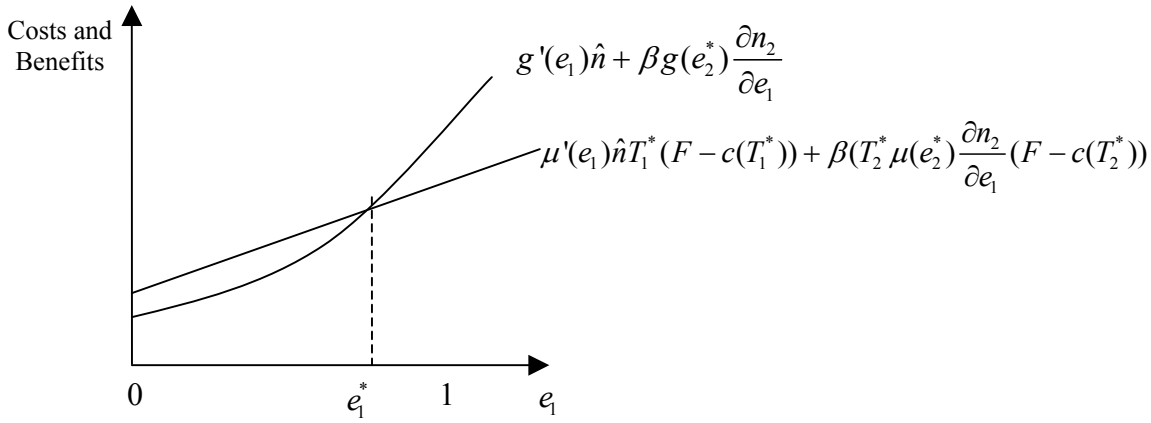
$$F > c(\overline{T}) + c'(\overline{T})\overline{T}. \tag{21}$$

Proof. See Appendix 4.

There are several insights in Proposition 4 that are worth mentioning. Firstly, when the FFS is equal to the average cost of providing the minimum treatment, we know from Proposition 3 that the level of effort and treatment are minimized during the second period. Since the fee is not enough high to cover even the marginal cost of providing the minimum treatment, the GP does not choose to increase the level of treatment beyond $\underline{T}$ during the first period. Moreover, exerting a positive level of effort entails greater disutility costs but does not give advantages in terms of more net revenue during neither the first period or the second one. The positive effect of reputation on the level of effort is neutralized by the absence of an income reward for the treatment of the seriously ill patients. Hence, the overall GP's utility is zero.

Secondly, whenever the fee is higher than the average cost of providing the minimum treatment, Proposition 3 secures that, during the second period, the level of effort is always positive and the treatment is such that, in general, the marginal benefit (i.e. the fee) is equal to the marginal cost of providing it. More difficult is to examine the GP's decisions during the first period. Regarding the effort, the GP chooses to exert such a positive level that the sum of the discounted marginal net benefits (i.e. the difference between the fee and the average cost of treatment) deriving from treating the additional serious cases both in the first and in the second period equates the sum of the total marginal disutility of exerting that level of effort in the first period and of the additional discounted disutility borne in the second one. Figure 4 shows that an increase in the FFS shifts upward the line of the discounted net benefits and determines a higher level of effort. A lower disutility of effort can produce the same result by shifting downward the discounted cost curve.

Figure 4 – FFS contract with reputation: effort decision in period 1



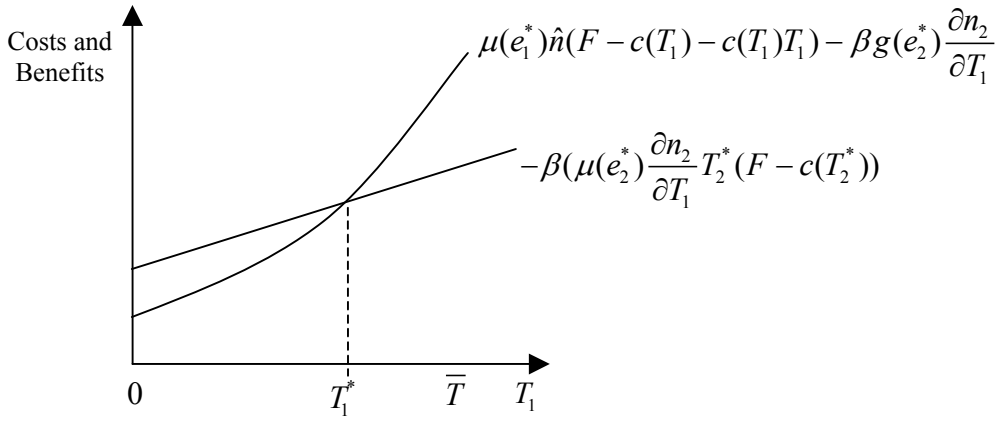
Occasionally, the level of treatment selected by the GP during the first period may be such that the fee is not high enough to cover the average cost of providing it (at that level of treatment, the fee is lower than both the marginal and the average cost). This

means that exerting a positive level of effort does not produce a monetary net benefit during the first period. However, the discounted marginal net benefit deriving from treating the additional seriously ill patients during the second period may be still sufficient to cover all the costs (including the monetary losses of period 1)[6].

A zero level of effort is chosen if disequation (11) is fulfilled, that is when the total discounted monetary net benefits of exerting an effort are lower or equal to the total discounted non-monetary costs. On the opposite, whenever the total discounted monetary net benefits deriving from exerting the maximum effort are still higher than the total discounted disutility costs, the GP decides to maximize his exertion.

The optimal treatment choice during the first period requires the GP to compare the benefits deriving from the provision of a higher level of treatment with the related costs. The former comprise both the marginal net benefit (i.e. the difference between the fee and the marginal cost of treatment) of providing a certain level of treatment to the seriously ill patients during the first period and the discounted savings in non-monetary costs due to the reduction of patients' demand during the second period. The latter includes the discounted loss of revenue resulting from the worsening in the GP's reputation (see figure 5). The optimal level of treatment in period 1 decreases with an increase in both the level of effort exerted in period 1 and the disutility of effort, and with a reduction in the level of effort exerted in period 2.

Figure 5 – FFS contract with reputation: treatment decision in period 1



It may occur that, given the optimal level of treatment chosen in the second period, the marginal net benefit of providing a certain level of treatment to the seriously ill patients during the first period is negative (i.e. the fee is lower than the marginal cost of treatment). Therefore, providing that level of treatment during the first period implies for the GP to incur in losses of revenue in both periods which, however, may be counterbalanced by the reduction in disutility costs caused by a decrease in the number of diagnoses performed in the second period[7]. Lastly, whenever disequation (15) or (16)

[6] This situation is more likely to happen when the disutility cost of exerting the effort is low.

[7] Again, this situation is more likely to arise when the disutility cost of exerting the effort is very high and the fee is low (though $F > c(\underline{T})$).

are fulfilled, then the GP decides to provide respectively the minimum or the maximum treatment.

5. Comparative statics

The analysis focuses here on the comparison of the solutions obtained under the different schemes. The purpose is that of defining the effect of reputation on the GP's effort and treatment decisions and understanding under which circumstances reputation is more likely to restrain the GP's double moral hazard.

Under a capitation system, the presence of reputation does not induce a change in the level of treatment, which is always minimum. However, it may represent a stimulus for the GP to choose a level of effort different from zero. This happens if the tomorrow discounted net benefit deriving from exerting a positive level of effort is higher than the today cost of providing the additional serious cases with the minimum treatment. A positive level of effort is more likely to occur when the capitation fee and the subjective discount factor are high, and patients are very sensitive to the care spent in conducting the diagnosis. The same results may also be obtained through a decrease in both the marginal productivity of the GP's effort and the initial number of patients in his list. Table 1 summarizes these results.

Table 1 – Comparison of the GP's effort and treatment decisions under a capitation system

	Without reputation	**With reputation**
Effort	0	≥ 0 (>0 if $\beta \frac{\partial n_2}{\partial e_1}(R - \frac{1}{4}c(\underline{T})\underline{T}) > c(\underline{T})\underline{T}\mu'(0)\hat{n}$)
Treatment	minimum	minimum

The case of a FFS contract is more controversial (see table 2). As mentioned, when the fee is equal to the average cost of providing the minimum treatment, the presence or absence of reputation does not make any difference. In both cases, effort and treatment are minimized. When the fee is higher than the average cost of providing the minimum treatment, the GP always exerts a positive level of effort, whether he has reputational concerns or not. However, the existence of a demand reaction may induce him to intensify the effort, provided that, during the second period, the discounted net benefit deriving from treating the additional serious cases is higher than the discounted disutility due to the additional diagnoses. The choice of a higher level of effort depends positively on the fee and the productivity of effort, and negatively on the disutility of effort. Both the discount factor and the patients' sensitivity to a change in effort do not play any role in taking this decision.

Concerning the treatment, the GP can provide whatever level he desires. Nevertheless, the risk of loosing patients in the future may prevent him from oversupplying treatment, as long as, during the second period, the discounted loss of revenue resulting from the worsening in reputation is greater than the discounted

savings in non-monetary costs. A lower treatment is more likely to be provided if both the fee and the productivity of effort are high, and the disutility of effort is low. Again the discount factor and patients' reaction to a change in the level of treatment do not influence the intensity of care.

Table 2 – Comparison of the GP's effort and treatment decisions under a FFS system

	Without reputation	**With reputation**
$F = c(\underline{T})$		
Effort	0	0
Treatment	minimum	minimum
$F > c(\underline{T})$		
Effort	>0	>0 (higher than the case without reputation if $\beta(T_2^* \mu(e_2^*)\frac{\partial n_2}{\partial e_1}(F - c(T_2^*)) > \beta g(e_2^*)\frac{\partial n_2}{\partial e_1})$
Treatment	$\underline{T} \le T \le \overline{T}$	$\underline{T} \le T \le \overline{T}$ (lower than the case without reputation if $-\beta(\mu(e_2^*)\frac{\partial n_2}{\partial T_1}T_2^*(F - c(T_2^*)) > -\beta g(e_2^*)\frac{\partial n_2}{\partial T_1})$

Lastly, we should analyse the GP's decisions in presence of reputation under the two contractual arrangements. With a fee-for-service component in the payment system, the GP's choice involves, in general, a more treatment-intensive practice style. Indeed, a FFS contract is not able by itself to restrain the provision of unnecessary treatment as it is a capitation contract. The minimum treatment may still be provided but only under the specific circumstances which have been showed in the previous section.

More difficult is to compare the two level of effort. While a payment fee higher than the average cost of the minimum treatment is always a sufficient condition for ensuring a positive effort under a FFS contract, it is not under a capitation contract. However, it is not possible to conclude under which type of contract reputation induces the GP to exert the highest level of effort. This depends strongly on the relative level of the FFS and the capitation fee.

6. Concluding remarks

This paper derives three main conclusions concerning the effects of reputation in primary care. First, we contrast the findings of some authors that, because of the GP opportunistic behaviour, capitation always leads to exert the minimum effort in diagnosis, while FFS always induces to oversupply treatment. On the contrary, by modelling the patient-GP relationship as a potential repeated relationship, we show that, when reputational effects are strong enough, the GP may find convenient to restrain his moral hazard, independently on the way he is paid.

Second, we prove that, under specific circumstances, reputation can make the capitation and the FFS contracts equivalent with respect to their effects on the GP's

choices of effort and treatment. Therefore, the superiority of one of the two contractual arrangements over the other is no more a fact, at least from the point of view of the quality of the care provided by the GP.

Third, in terms of policy implications, the results leave an important role to be played by the public health authority. Whatever payment scheme is adopted, the amount of the GP's remuneration contributes to reinforce the reputational effects. *Ceteris paribus*, an increase in the payment fee may encourage the GP both to increase the effort in diagnosis and to provide a less intensive treatment. Moreover, the health authority may act on other two variables: the sensitivity of patients to variations in the level of both effort and treatment provided by the GP. However, the effort exerted by the GP is not directly observable by patients, who can judge it only through the improvements in their health status. Thus, the task of the policy maker becomes that of making the patients' less tolerant to their illness status and to increase the GP's productivity of effort[8]. All these goals may be achieved by promoting a greater information on patients' rights and on the available treatment opportunities, and by financing training courses for the GPs.

This work represents a first step towards going into the reputation topic in more depth. Obviously, it presents some weaknesses, first of all that of assuming *a-priori* not only the relationship between GP's choices and patients' demand but also the nature of this relationship. In order to refine these first results, it might be interesting to empirically study the rationale of this demand's behaviour, and, more generally, of the GP-patient relationship. Another logical extension of this paper could bring to design the optimal remuneration contract by the public insurer, so as to induce GPs to pursue both the objectives of quality enhancement and cost containment ("multitask agency").

The Appendices may be provided on request

References

Brasseur, C. (2000), *Competition between general practitioners and specialists in the primary health care market,* Working Paper, Institut de Recherches Economiques and Sociales (IRES), Université catholique de Louvain, Belgium.

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

Chalkley, M. and J.M. Malcomson (2000), "Government purchasing of health services", in A.J. Culyer and J.P. Newhouse (eds.), *Handbook of Health Economics*, chapter 15:848-889, Elsevier, Amsterdam.

Dionne, G. and A. Contandriopoulos (1985), "Doctors and their workshops", *Journal of Health Economics* 4 :21-33.

[8] This instrument is particularly important under a FFS system, since the GP's decisions are not influenced by the patients' sensitivity to variations in both effort and treatment.

Ellis, R.P. and T. McGuire (1990), "Optimal payment systems for health services", *Journal of Health Economics* 9:375-396.

Ellis, R.P. and T. McGuire (1993), "Supply-side and demand-side cost sharing in health care", *Journal of Economic Perspectives* 7(4):135-151.

Garcia Mariñoso B. and I. Jelovac (2003), "GPs payment contracts and their referral practice", *Journal of Health Economics*, 22(4):617-635.

González, P. (2002), "Should physicians' dual practice be limited? An incentive approach", Manuscript, Departamento de Fundamentos del Análisis Económico,Universidad de Alicante, Spain.

Jelovac, I. (2001), "Physicians' payment contracts, treatment decisions and diagnosis accuracy", *Health Economics* 10:9-25.

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

Ma, C.A. and T.G. McGuire (1997), "Optimal health insurance and provider payment", *American Economic Review* 87:685-704.

Montefiori, M. (2001), *Reputazione come credibilità per comportamenti passati nei contratti tra acquirente e fornitore di servizi sanitari*, Working Paper n. 7/2001, DISEFIN, Dipartimento di Economia Pubblica, Università di Genova.

Montefiori, M. (2002), *Libertà di scelta e contratti prospettici: l'asimmetria informativa nel mercato delle cure sanitarie ospedaliere*, Working Paper n. 29, Dipartimento di Politiche Pubbliche e Scelte Collettive, POLIS, Università degli Studi del Piemonte Orientale, Alessandria.