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

Reimbursement schemes for physicians:
risk sharing, asymmetry of information and patients' choice

Rosella Levaggi and Lise Rochaix

Dipartimento di Scienze Economiche, Università di Brescia, Via S.Faustino 74/b, Italy, *levaggi@eco.unibs.it*, corresponding author

GREQAM-IDEP Centre de la Vieille Charité, 2 rue de la Charité, 13002 Marseille, France, *rochaix@ehess.cnrs-mrs.fr*

Reimbursement schemes for physicians: risk sharing, asymmetry of information and patients' choice

Rosella Levaggi and Lise Rochaix

Abstract

Primary health care is experiencing a process of radical change as concerns, finance, organisation and the role the patient is playing in determining the quantity and the type of health care to be supplied. As concerns financial issues, asymmetry of information does not allow to reach a first-best allocation of resources. For this reason, physicians' payment schemes are constantly being reformed to enhance their own performance and to ensure a more optimal allocation of scarce medical resources. On the other hand, the relationship between the patient and his physician is changing. The patient is (or thinks to be) informed about his need for health care and makes a critical assessment of the medical advice he receives. In this paper we present an optimal payment scheme for physician that combines the rules for incentive compatible mechanisms with the exploitation of the competition among doctors that patients might be able to generate.

In our model we focus on the possible competition between physicians and specialist and we show that even when patients do not have priors on their medical conditions, the competition they are able to generate reduces the information rent that doctors derives from their ability to observe the true medical state of each patient before making their effort.

1-Introduction

Public health care systems are under a great deal of pressure to enhance their efficiency in delivering health care. The reforms proposed have been mainly focussed on hospital production where the introduction of internal markets for health care has represented a very important step in the innovative process aimed at introducing competition in a sector that was dominated by the rules of bureaucratic production. With some important exceptions[1], primary care has not been so widely reformed. However GP's play a very important role in the determination of health care for their twofold function of gate keepers for Central Government and agents for the patients. It is for these reasons that the system of payment for GP are being widely revised to provide better instruments to control the quality and the quantity of services. A primary health care system is based on two elements: the payment scheme and the access rules to specialist care. The remuneration options for GP's are numerous and a number of studies have been proposed to explain the reactions of doctors to alternative systems[2]. As per access, there are two basic methods: a *vertical* organisation where the patients have to go and see a GP first or an *horizontal* structure where the patient can choose which provider to consult. The literature dealing with primary care has not always taken into due account an important aspect related to the supply of health care, i.e. uncertainty. In the provision of primary health care uncertainty has at least three important dimensions: the health status, the severity of illness and patients' personal ability to take advantage of medical care. These variables can be observed at national level and are expressed by the aggregate probability of being ill and of recovering, but cannot be observed by both parties. The GP, by being closer to the patients, can predict the probability of the patient being ill and of recovering. In this process the patient is assumed to play a passive role: he is considered to be so ignorant about his health status that he cannot translate his need for health care into the demand for a specific service. Consequently, he is assumed to delegate choice to a physician and to strictly comply with the recommended course of action, a type of behaviour parallel to Hirschman[3] notion of loyalty. Several papers show that this ignorance does not imply a passive role. Gravelle and Masiero (2002) use patients expectations to improve the quality of health care while Levaggi and Rochaix (2003b) show that patients choices can be used to reduce the scope for opportunistic behaviour in an asymmetry of information framework.

In this paper we propose a payment system that takes explicit account of asymmetry of information

[1] In this context it is interesting to remind the introduction of GP fundholders in Britain. The experience has not been positive and in Britain a new organisation for primary care has been recently introduced .
[2] See Gravelle (1999), Gravelle and Masiero (2000), Franc (2002)
[3] cf. Hirschman A.O., (1970), 'Exit, Voice and Loyalty', Harvard Univ. Press, Cambridge, Mass..

and patients choices. The theoretical foundations of our model are presented in Levaggi and Rochaix (2003a,b). The paper will be organised as follows: in the following section we describe the environment in which health care is delivered and the payment options that can be adopted, in section three the model is outline; in section four the payment system when information is symmetric is presented; section five presents the results under information asymmetry and finally section six concludes the paper.

2 – Primary care systems

Primary care systems are characterised by two important elements: the payment scheme, i.e. the system chosen to pay for physicians services and the organisation of the service, i.e. the definition of the access routes to specialist care. The properties of the different systems and the role of patients choice are discussed at length in Levaggi and Rochaix (2003a); in this article we extend that analysis in two directions: we explicitly model the risk sharing properties of the payment schemes and we introduce diagnostic.

Commentaire [L1]: Has to be written better.

2.1 The payment schemes

There are three basic methods for compensating clinicians and they are salary, capitation and fee-for-service (FFS). Each method presents strengths and weaknesses and has different performances in the presence of asymmetry of information[4]. The first method, salary, foresees the payment of a reward to the doctor based on the number of hours he agrees to "work". It is important to note that the concept of working is not strictly related to the concept of producing health care. The dycothomy is due to the presence of uncertainty on the probability of the patient being ill. The salaried physician, in fact, is asked to spend a predetermined amount of hours in his office where he will take care of his patients. In the event of nobody going to his clinic, he waste his effort, in terms of hours, because he is paid without being able to produce a service. This method is quite effective as concerns cost control and it leaves very little to cheating under asymmetry of information. On the other hand, its application does not allow to reach a technical efficient position since too much effort is wasted in the waiting process. From an applied point of view, one has also to consider that in the salary system fixed costs are quite high. It is only in theory that the optimal effort can be expressed in terms on units per patient. In its actual implementation, the system foresees the payment of the doctor for a predetermined number of hours of surgery. Even in a world of perfect information on the productivity of the effort, Central Government should assign to each GP a number of patients to match the optimal effort and this might not be possible, given the distribution

[4] See Levaggi and Rochaix (2003a) for a review

of the population and/or their health requirements. For these reasons, we think that this type of payment is more appropriate for hospital health care and it will not be considered in our analysis.

The second method for paying physicians involves capitation. Pure capitation pays a lump-sum per persons in a physician's list of "clients" and offers health care when needed. This is the system by which GP's are paid in Italy and the U.K., with some corrections. The physician has to produce a minimum amount of work represented by the number of hours of surgery he has to offer and after this he will be able to modulate his effort to the real need of each patient for health care. This system has the advantage, as the previous one, to control cost and in an environment without any type of information asymmetry, it would be superior to the previous scheme since it allows the GP to reduce the number of hours he spends in his surgery without producing health care. However, in the presence of asymmetry of information the doctor might have an incentive to refer patients to specialist care or to choose for his list those patients with a lower probability of being ill.

The third important category of payment is that of fee-for-service (FFS). This was the predominant system for paying for health care, but it is now declining. Under fee-for-service, the payment depends on the unit of work the physician produces, measured in terms of services that are actually offered to the patients. In this case it might seem that the effort of the doctor is even more related to the patient's health condition. In a world where information is symmetric, this would be the case and it is for this reason that for a long time this method was the preferred one to reimburse doctors. However, in an asymmetric world, this system produces perverse effects to the behaviour of doctors. The doctor has, as in the first case, an incentive in reducing the effort if the patient is going to recover quickly, but he would instead choose patients that are going to be ill to improve his reward (Jaegher and Jeger, 2000).

2.2 The organisation of the service

The organisation of the service defines the set of rules that governs access to specialist and hospital care. Two are the main models: direct access to specialists (hereafter called the horizontal structure) versus indirect access through referral from GPs (hereafter the vertical structure). In an *horizontal structure* patients are allowed to choose any GP or any specialist they like according to their preferences. If the system is instead organised in a *vertical structure*, patients are asked to go and see their GP first; on the basis of this first visit, the physician decides to treat the patient or to refer him to a specialist.

3 - The model

The model proposed here is designed to explain the relationship between patients, GPs and specialists. In particular, we want to focus on the incentives to opportunistic behaviour caused by asymmetry of information and on the corrections to the payment scheme and access rules that can be introduce to reduce these problems.

3.1- The environment

The model focuses on a limited part of the decision process that a Central Government faces regarding the provision of health care to the population. The environment in which such decisions are taken is one of uncertainty and information asymmetry. Uncertainty mainly depends on two factors: Central Government cannot observe the health status of each individual; for those falling ill neither the severity of illness nor their ability to take advantage of health care can be observed. Each event has however a known probability that is common knowledge.

Each individual is endowed with a stock of health H. Illness implies a given health loss $\overline{H}$ so that the patient is left either with $H-\overline{H}$ with probability (1-p_h) or with H with probability p_h. The loss of health might derive from two diseases labelled *A* and *B*. The patient is assumed to be ignorant about his health status; in particular we will assume that while any individual falling ill will go and see his doctor, also a fraction ρ of the ones not being ill will go and consult a physician to be reassured about their health status. If medical care is appropriate (*MC*) the system allows the patient to restore his health completely. Each illness can be either mild or severe. If the patient has a severe condition, it is assumed that the GP cannot treat the case and has to refer it to a specialist. In this case, the GP has to make an effort e^* which corresponds to the time for referral. For a mild disease, GP's treatment is appropriate, but the level of care that has to be used depends on the ability of the patient to recover. Medical care is the throughput or intermediate output of the GP's effort and is obtained through the following production process:

$$MC = f(e) \qquad (1)$$

Uncertainty prevailing in the production of health is introduced in a multiplicative form in the production function: $H = \sigma MC$. This random parameter reflects the patient's reaction to the treatment, itself a function of factors such as education, or psychological / physical abilities (as in the Grossman approach). We assume for simplicity that it only takes two values, namely σ_L (for the low productivity patient) and σ_H (for the high productivity patient) with known probability *z* for the former and (1-z) for the latter type of patient. To simplify the analysis we assume that the probability distribution of *z* and *p* are independent. This allows to simplify the exposition without altering the results of our model.

When the patient falls ill, the health care system allows him to restore his health completely, i.e. there is no trade off in this model between cost for health care and health recovered[5]. The effort of the physician is uniformly distributed across the time necessary for the therapy. Care is then defined by a set number of consultations established on the basis of protocols. This assumption allows us to transform the effort of the physician in number of consultations (v). As a result of these assumptions, we can summarise the health status and the effort by the GP as follows:

Status	Probability	Health level before	Health level after	Effort	Number of visits
Not ill /no consultation	$p_h(1-\rho)=p_0$	$\overline{H}$	$\overline{H}$	0	0
Not ill/ easy diagnosis	$p_h\rho\delta=p_1$	$\overline{H}$	$\overline{H}$	e^*	$1=v_1$
Not ill/ difficult diagnosis	$p_h\rho(1-\delta)=p_2$	$\overline{H}$	$\overline{H}$	e_d	$v_2=\frac{e_d}{e^*}$
Mild ill /high recovery	$p_m z=p_3$	$\overline{H}-H$	$\overline{H}$	$e_{imH}=f^{-1}\left(\frac{\overline{H}}{\sigma_H}\right)$	$v_3=\frac{e_{imh}}{e^*}$
Mild ill / slow recovery	$p_m(1-z)=p_4$	$\overline{H}-H$	$\overline{H}$	$e_{imL}=f^{-1}\left(\frac{\overline{H}}{\sigma_L}\right)$	$v_4=\frac{e_{imL}}{e^*}$
Severely ill	$p_s=p_5$	$\overline{H}-H$	$\overline{H}$	e^*	1; v_5♣

♣ effort made by the specialist

Tab. 1: Health status, associated probabilities and physicians' effort

where $v_3>v_2$ and $v_4>v_2$. The specialist is disease-specific, i.e. he can treat any degree of severity of the condition he is specialised in, but he cannot cure any type of the other ailment. To simplify the exposition it will be assumed that to treat a severe case of either *A* or *B*, v_5 visits are necessary. Both specialists are paid on a FFS basis at a rate equal to γ_{SP}; patients falling ill are uniformly distributed among the two diseases so that the probability of having a mild (severe) attach of ailment *A* is equal to ½ p_m (½ p_s) and so on.

[5] The trade off between these variables has been extensively studied in Levaggi and Rochaix (2003a)

3.2 The actors of the game

In this model there are three active parts: the physicians, the patients and CG each of them having specific objectives that might be in contrast with one another. Some of these conflicts can be used to improve the performance of the system as it will be shown in this paper.

3.2.1 The physicians

In this model we assume that physicians are utility maximisers over two distinct levels. At the first level, they decide their participation in the productive process. The utility of the physician is given by the difference between the payment received and the effort made:

$$EU_{GP} = EU(R(v) - d(e,v)) \tag{2}$$

where *e* is the effort and *R(v)* is the state contingent payment. For capitation, it will be equal to a fixed amount *CP*; for FFS *R(v)* depends on the number of consultations that have to be made. The function is assumed to be separable and additive in the two components. To concentrate on risk sharing issues, it is assumed that each consultation made by the physician, whether it is to determine if the patient is ill, to refer him or to administer care, requires a fixed effort e^* and, if the payment is made on a FFS basis a fixed remuneration γ_{GP} is foreseen for each consultation. The disutility is increasing in the effort $(\frac{\partial d}{\partial e} > 0; \frac{\partial^2 d}{\partial e^2} > 0)$, and linear in the number of consultations $(\frac{\partial d}{\partial v} > 0; \frac{\partial^2 d}{\partial v^2} = 0)$ so that the expected utility of the physician can be written in terms of the care needed by each patient as follows:

$$EU_{GP} = EU\left[R(v) - E(v)d(e^*)\right] \tag{2a}$$

The physician accepts the contract if the utility he expects to receive is greater than his reservation utility $\overline{U}$ that is agreed at central level by the Government and the medical profession representatives. To simplify the analysis we will assume $\overline{U}$ =0.

If the contract satisfies the participation constraint, the physician then maximises his net income through the following function:

$$V_{GP} = V(nER) - Eg(n) \tag{3}$$

where *n* is the number of patients treated by the GP and *Eg(n)* is the expected cost to enrol patients: it comprises administrative and marketing costs. *ER* is the expected income the physician gets from treating a patient.

3.2.2 The patient

Each patient has a fixed income Y_i and is maximising a utility function which depends on his private consumption, on the time spent to be reassured on his health status and on the expected time needed to get back to his former health level[6]. The system entitles each patient to free treatment until full recovery, which implies that the only cost considered here is the opportunity cost of the time spent in the health care system. The time spent for being reassured it is assumed that it does not produce any disutility to the patient since the loss is compensated by the reassurance; the time spent to recover his health level produces instead a disutility that is assumed to be proportional to the time spent to get back to full health. Utility is assumed to be additive in these elements:

$$EU = U(Y_i) - E(t) \quad (4)$$

When care is organised according to an horizontal structure, the patient can choose whether to see a GP or a specialist; in a vertical system he will instead be forced to go and see a GP first. The first system allows the patient to choose the solution that maximises his expected utility; in the second case CG determines through his choice the level of expected utility of the patient.

3.2.3 Central Government

Central Government's objective is to restore health to those that have lost it because of illness at the minimum expected cost. Cost for ambulatory care has three main components: the remuneration for the GP, the payment to the other physicians and other costs related to drugs, test etc. In this article we focus on the first two categories. Specialists are paid on a FFS basis at a rate γ_{SP} for each consultation made. As per the GP, the two main systems are capitation and FFS[7].

We assume CG's preferences to be lexicographic over the following three objectives: maximising the patient's health outcome, minimising expected cost per patient and minimising the patient's opportunity cost of time spent seeking care. The first objective is met by the assumption that each patient is entitled to free care, up to full recovery . The second objective will drive the comparison between alternative structures. CG will only consider the third objective if the second is fulfilled and if patients preferences are such as to value choice per se. In formal terms, CG's objective can be written as:

$$Min \qquad p_o\phi(R_o)p_1\phi(R_1) + p_1\phi(R_2) + p_3\phi(R_3) + p_4\phi(R_4) + p_5\phi(R_5) \quad (5)$$

[6] ~~Write note on fiscal illusion~~The utility of the patient should also depend on the cost for health care. Even if it is free at the point of use, the cost will have to be borne through taxation so that the patient has an interest in keeping it as low as possible. However, the impact on the tax rate of each single treatment is marginal and for this reason it has not been considered

[7] See Levaggi and Rochaix (2003a, 2003b) for a review

Once the payment scheme has been determined, CG will have to choose the organisation (horizontal or vertical) that allows cost minimisation and finally it will take account of patients preferences.

Commentaire [L2]: 1Assess whether to introduce the section The rules of the game

4 – Payment systems when information is symmetric

In this section we examine the optimal payment system that Central Government might offer to GPs in a context characterised by symmetry of information. The time flow of the contract is presented in figure one.

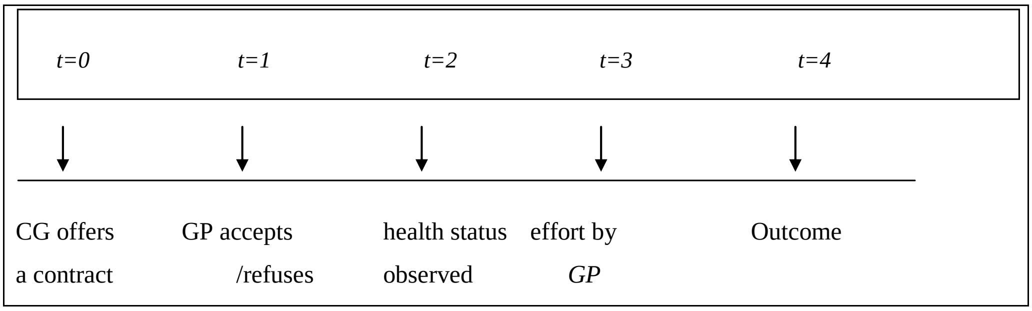

Figure one: The timing of the contract when information is symmetric

When the contract is signed the health status of each patient is not known; this information will be available at some later stage and it will be observed by both agents. In this environment, the form of the contract depends on the attitude towards risk of the two parties which in turn determines the risk sharing rules.

4.1 Payment schemes

The payment schemes we analyse in this paper ranges from pure capitation to FFS. They can be summarised in the following formula:

$$R_i = a + b\gamma_{GP} v_i^R \qquad (6)$$

where *a* and *b* are two parameters that determine the risk sharing properties of the contract. The risk shared is the cost to restore health care to each patient falling ill which is in turn determined by the uncertainty on the health status and the ability of taking advantage of health care.

v_i^R is the number of consultations to be reimbursed; in a symmetric world, this is equal to the number of visits made by the GP as one might expect. When information is asymmetric, v_i^R becomes a strategic variable to avoid an opportunistic behaviour by GP.

For $b=0$ and $a>0$, we have a pure capitation scheme and the provider bearing all the risk; if $a=0$

and $b=1$ the payment is FFS and the purchaser takes on the risk. Finally if $a>0$ and $0<b<1$ a mixed payment scheme is to be implemented.

The optimal payment scheme is derived from the solution of a principal-agent problem in which CG acts as the principal. In formal terms we can write it as:

$$
\begin{aligned}
Min \quad & p_o\phi(a) + (p_1+p_5)\phi(a+b\gamma_{GP}) + p_2\phi(a+b\gamma_{GP}v_2^R) + p_3\phi(a+b\gamma_{GP}v_3^R) + p_4\phi(a+b\gamma_{GP}v_4^R) \\
s.t \quad & p_0\left[U(a)\right] + (p_1+p_2)\left[U(a+bv_1^R\gamma_{GP} - d(e))\right] + p_2\left[U(a+bv_2^R\gamma_{GP} - v_2 d(e))\right] \\
& + p_3\left[U(a+bv_3^R\gamma_{GP} - v_3 d(e))\right] + p_4\left[U(a+bv_4^R\gamma_{GP} - v_4 d(e))\right] \geq 0 \qquad (7) \\
& a \geq 0 \\
& b \geq 0 \\
& b \leq 1
\end{aligned}
$$

The first constraint represents GP's expected utility and defines his minimum reservation utility. This constraint represents the imperfection of the market for health care. The presence of occupational licensure allows GP's to have a monopoly power which means that they are not paid a price based on the marginal productivity of their effort alone. In the context of the GP - Central Government relationship it might be set by an agreement between the representatives of the two parties, much in the same way as trade union agreements. The results of the model are presented in appendix one and are summarised in table two.

Tab. 2: Payment schemes and risk sharing

Physician / Central Government	Risk averse	Risk neutral
Risk averse	Any (Mixed)	Capitation
Risk neutral	FFS	Any (Capitation)

Commentaire [L3]: To be done: is it possible to define conditions for U and ϕ that allows to say that the choice will be a mixed payment scheme?

If Central Government is risk averse $(\frac{\partial\phi}{\partial R} > 0; \frac{\partial^2\phi}{\partial R^2} > 0$ and the physician is risk neutral $(\frac{\partial U}{\partial R} > 0; \frac{\partial^2 U}{\partial R^2} = 0)$, the optimal system is capitation:

$$
\begin{aligned}
& a = CAP \\
& b = 0 \qquad (8) \\
& CAP = E(v)d(e) = E(v)\gamma_{GP}
\end{aligned}
$$

where $E(v)$ is the expected number of consultations. The payment made to the GP is independent of any uncertain event such as the health status and the ability of the patient to recover health. The GP is in fact paid a fixed amount based on the expected number of consultations that the protocol defines as appropriate for each ailment. The GP will have to make a state contingent effort that depends on the actual health status of each patient.

Capitation might be the optimal solution also for a system where CG and GP are risk neutral. In this case, any solution is optimal, but we might think that public choice issues in this case will make CG opt for capitation[8].

If Central Government is risk neutral and the physician is risk averse, it is optimal to choose a FFS that allows the physician to receive his reservation utility in any case. The payment will beequal to:

$$\begin{aligned} &a = 0 \\ &b = 1 \\ &\gamma_{GP} = d(e) \\ &v_i^R = v_i \end{aligned} \tag{9}$$

When both agents are risk averse, any solution is compatible with (7) and the choice depends on the relative degree of risk aversion. The most likely solution is a mixed payment system where the physician is paid a fixed amount for each patient in his list and a reimbursement for each visit he makes that is lower than its actual cost in terms of utility.

The second level of maximisation, i.e. the number of patients to enrol is not affected by the reimbursement scheme. In this case, in fact, because information is symmetric, the GP cannot cheat on the number of consultations or on the severity of illness, hence he will choose the number *n* of patients so as to maximise his utility. The FOC can be written as:

$$\frac{\partial V(E(R))}{\partial n} = \frac{\partial g}{\partial n} \tag{10}$$

4. 2 Access rules

If the type of organisation chosen is horizontal, the patient can choose the type of care he prefers.

Three possibilities are opened to the patient: consulting the GP's or seeking care by one of the two specialists. In either case, the patient has a given probability of being right and each action has an associated time cost. Various treatment paths can be identified[9]; if the patient has sought the right provider, he will be treated, otherwise he will be referred back to another provider. Note that for a mild condition the time cost for the patient associated with specialist or GP care is identical, provided he has chosen the right specialist. In this case, in fact, the specialist takes care of the patient instead of referring him to the GP. From the patient point of view, this is the most efficient outcome and it allows the specialist to increase his income. Finally we will assume that the choice

[8] Risk neutrality means that there is no trade off here between expected payments and thier variance. It is however reasonable to assume that when confronted with the choice of a certain payment and an uncertain one having the same mean,CG will chose the certain one. If the payment is fixed, this means that its finance is less complicated since budgets can be defined in advance without any risk of overspending

[9] See Levaggi and Rochaix (2003b)

of seeing a specialist or a GP is just related to health states that requires care. The patients that are not ill and still go and consult a physician to be reassured on their health status are assumed to go and see a GP. This assumption can be justified in several ways: one might think that the patient that is not sure about his health status would go and see a GP because would not know which specialist to choose; for our analysis another consideration is more important: patients that need to be reassured about their health status might receive positive utility from the time spent in care. This means that equation (4) should have a specific form according to the health status. For our analysis it is however sufficient to assume that patients choose their provider only if they are ill. We will get the same results in terms of choice and competition among providers and a much simpler model. We assume that to recover from a severe attack of either A or B, v_5 visits made by the specialist are required. Table 3 shows the time cost of each action and the associated probability of incurring such a cost.

Tab 3: Time cost for the different alternatives

Action	Time cost	Probability
Going to GP	$v_3;v_4;1+v_5$	$p_3;p_4;p_5$
Going to SP_A	$v_3;v_1+v_3;v_4;v_1+v_4;v_5;1+v_5$	½ p_3; ½ p_3;½ $p_{4;}$ ½ $p_{4;}$½ $p_{5;}$ ½ p_5
Going to SP_B	$v_1+v_3;v_3;v_1+v_4;v_4;1+v_{5;}v_5$	½ p_3; ½ p_3;½ $p_{4;}$ ½ $p_{4;}$½ $p_{5;}$ ½ p_5

The patient will choose the alternative that allows him to maximise his utility. The utility level for the three alternatives can be written as:

$$
\begin{aligned}
EU_{GP} &= U(Y) - [p_3 v_3 + p_4 v_4 + p_5(1+v_5)] \\
EU_{SPa} &= U(Y) - [\frac{1}{2} p_3 v_3 + \frac{1}{2} p_3(v_1+v_3) + \frac{1}{2} p_4 v_4 + \frac{1}{2} p_4(v_1+v_4) + \frac{1}{2} p_5(v_5) + \frac{1}{2} p_5(1+v_5)] \qquad (11) \\
EU_{SPb} &= U(Y) - [\frac{1}{2} p_3(v_1+v_3) + \frac{1}{2} p_3 v_3 + \frac{1}{2} p_4(v_1+v_4) + \frac{1}{2} p_4 v_4 + \frac{1}{2} p_5(1+v_5) + \frac{1}{2} p_5(v_5)]
\end{aligned}
$$

If the utility is linear in the time spent in care and all the events have the same probability, the patient is indifferent between going to the GP or to the specialist[10]. From a policy point of view, this means that choosing the provider has no value for the patient in this context and Central Government can choose the type of organisation of the service that allows him to minimise the cost of provision.

[10] The more general case is presented in Levaggi (2003b)

4.3 The organisation of the system

In an environment characterised by uncertainty and symmetric information the decision on the type of organisation can be separated from the choice of the payment scheme. In particular, after choosing the payment scheme on the basis of the degree of risk aversion of the two main actors (CG and the physicians), the rules of the access to the service can be laid down. CG will of course choose the system that minimise his expected costs.

For a vertical organisation, the total expected cost depends on the probability of each event; if the system is organised in an horizontal way it also depends on the choice of the patients. In the model presented here we assume a linear disutility in the time spent to get health care and uniform probabilities on the type of disease so that the patients are indifferent between the two providers; we can then assume that in an horizontal system the probability of going to specialist or a GP is equal to ½ . The costs for the different alternatives are presented in table four.

Tab. 4: Expected total cost for patient

Payment scheme	Expected cost for each patient	
	Vertical organisation	Horizontal organisation
Capitation	$CAP + p_5\gamma_{SP}v_5$	$\frac{1}{2}(CAP + p_5\gamma_{SP}v_5) + \frac{1}{2}\Big[CAP + \frac{1}{2}(p_3\gamma_{SP}v_3 + p_3\gamma_{SP})) + \frac{1}{2}(p_4\gamma_{SP}v_4 + p_4) + \frac{1}{2}(p_5\gamma_{SP}v$ $= CAP + p_5\gamma_{SP}v_5 + \frac{1}{4}\gamma_{SP}\big[p_3(v_3+1) + p_4(v_4+1) + p_5\big]$
FFS	$p_0 0 + p_1\gamma_{GP} + p_2\gamma_{GP}v_2 + p_3\gamma_{GP}v_3 + p_4\gamma_{GP}v_4 + p_5\gamma_{GP} + p_5\gamma_{SP}v_5$ $= CAP + p_5\gamma_{SP}v_5$	$\frac{1}{2}\{p_0 0 + p_1\gamma_{GP} + p_2\gamma_{GP}v_2 + p_3\gamma_{GP}v_3 + p_4\gamma_{GP}v_4 + p_5\gamma_{GP} + p_5\gamma_{SP}v_5\}$ $\frac{1}{2}\left\{ p_0 0 + p_1\gamma_{GP} + p_2\gamma_{GP}v_2 + \frac{1}{2}p_3\big[v_3(\gamma_{GP}+\gamma_{SP}) + \gamma_{SP}\big] + \frac{1}{2}p_4\big[v_4(\gamma_{GP}+\gamma_{SP}) + \gamma_{SP}\big] + p_5\gamma_{SP}v_5 + \frac{1}{2}\gamma_{SP} \right\}$ $= p_0 0 + p_1\gamma_{GP} + p_2\gamma_{GP}v_2 + p_3v_3\left(\frac{1}{2}\gamma_{GP} + \frac{1}{2}\bar{\gamma}\right) + p_4v_4\left(\frac{1}{2}\gamma_{GP} + \frac{1}{2}\bar{\gamma}\right)$ $+ p_5\left(\frac{1}{2}\gamma_{GP} + \frac{1}{4}\gamma_{SP}\right) + p_5\gamma_{SP}v_5$

If CG chooses a capitation system, the choice of a vertical organisation is necessary because the payment made to the GP is independent of the prevailing state of the world. The cost of an horizontal organisation, where the patient chooses to go straight to the specialist and might make a mistake, is always higher.

FFS is in theory compatible with both system; the choice depends on the relative cost of the GP's and the specialist effort and on the distribution of probabilities among the different states of nature. In our case where the probability is uniform and given that specialists fees are usually higher than GPs', a vertical organisation is less costly.

The same considerations can be made for priors of patients; in our system patients do not have priors on neither on the severity nor on the type of disease. If this assumption is relaxed and if their priors are not biased, it might once more possible that an horizontal system is more efficient than a vertical one at reducing cost.

In general, we can conclude that a vertical system allows a stricter control of expenditure and might be preferred for planning reasons: the variation in expenditure for patient depends only on the probability of having a severe disease. The system allows to reduce the risk that wrong priors by patients increases expenditure by making the wrong choice of going to the specialist (either because it is the right one or because it was not appropriate). However, this system does not allow to take advantage of right priors of the patient and, as we will see in the following section, it does not allow to use competition among providers.

5 - Information asymmetry

In an uncertain world, information is a resource that has, as any other commodity, a price. In the context of our model, we can assume that, although when the contract is stipulated both parties cannot observe σ or the severity of the patient, this information can be acquired later by the physicians (GPs and specialists) that can use it to their own advantage. The time flow for information in this environment is presented in figure two.

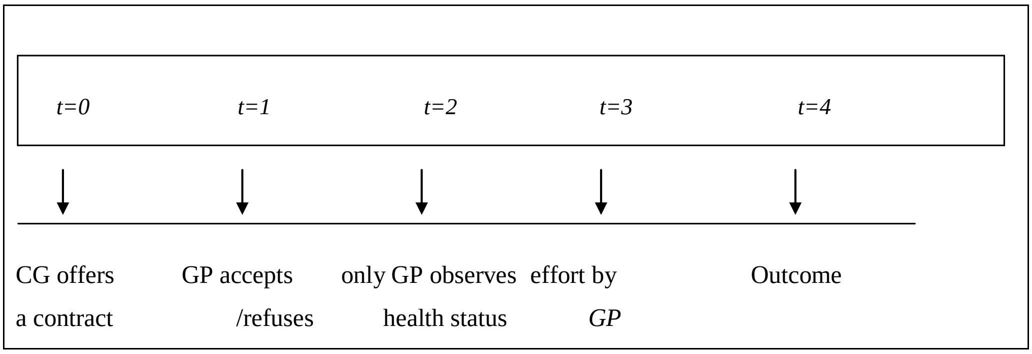

Fig.2 : Contract schedule: asymmetry of information

There are two areas of uncertainty in the model which correspond to two separate possibilities for the GP to set up a strategic behaviour to his own advantage. The first possibility is that of reporting a false σ if the patient is ill, a behaviour that can be classified as moral hazard. The schemes proposed offer different comparative advantages to these policies and present different solution to the problem of stopping GP's to use this type of strategic behaviour.

If capitation is the system chosen, GPs do not have any incentive in reporting a false σ because the payment they receive is not correlated to state contingent effort. In FFS, the payment is related to the number of consultations, if they cheat on σ, they can make more visits and reduce the effort. For an high recovery patient, the GP should make the effort $e_{imh} = \frac{\overline{H}}{\sigma_H} = \frac{v_3 e*}{v_3} = e*$ but, if he reports that the patient is a slow recovery, he is entitled to make v_4 consultations, so that the unit effort is equal to $e_{cLH}^{Cheat} = \frac{v_3 e}{v_4} < e*$. If the patient is slow recovery, the GP has not got any incentive to cheat. The total effort to treat him is in fact equal to $e_{iml} = \frac{\overline{H}}{\sigma_L} = v_4 e*$ hence the effort for each visit is equal to $e_{cLH}^{Cheat} = \frac{v_4 e*}{v_3} > e*$. Given that the disutility is increasing in the effort, the physician will cheat in the first case and will not have an incentive in reporting a level lower than the actual one.

A second way of cheating consists misreporting the level of severity of the patient. This behaviour could be classified as adverse selection and it is usually referred in the literature as cream skimming or dumping[11]. In this case, the study of the incentives is more complex.

In a system of capitation, the GP has an interest in minimising the number of consultations he has to make to the patient. Given the assumptions of our model, this would mean that the GP is indifferent between reporting that the patient is not ill or that he is severely ill. In both cases, in fact he will make one consultation. For ethical reasons it seems reasonable to assume that the GP will follow the second alternative. The outcome of the opportunistic behaviour in a capitation scheme is presented in table five

Tab. 5: Opportunistic behaviour in a capitation scheme

True state	Reported state	Conditions for cheating
Not ill	Not ill	
Not ill /difficult diagnosis	Refer to specialist	$v_2>1$
Mild ill /high recovery	Severely ill	$v_3>1$
Mild ill / slow recovery	Severely ill	$v_4>1$
Severely ill	Severely ill	

[11] See Ellis (1997), Lewis and Sappington (1999)

The incentive for the GP to an opportunistic behaviour depends on the number of consultations he has to make in each case. Since the health status of the patient can always be verified ex post, the GP cannot cheat by offering treatments that will make the patient not recover his full health. Referring when it is not necessary is a possibility allowed in this model by the assumption that specialists are paid, as it usually happens, on a FFS basis hence they do not have any incentive in referring the patient back to the GP.

If the system is pure FFS, physicians have then an incentive in maximising the number of consultations they make in order to maximise their net income. Their net income is in fact given by the volume of activity minus the cost which, in our model, depends on the number of patients.

The study of the possible ways of cheating depends on the assumptions about the ethics of the physicians and the environment in which they supply health care. In our model we assume that physicians do not administer health care to healthy patients. In this case they are likely to induce demand through more diagnostic tests than necessary before reassuring the patient on his state of health.

The second important principle is that the patient has to recover his full health; to secure this principle we assume that the doctors that treat the patient for a severity level lower than appropriate can subsequently – but within the period considered – refer the patient to the specialist. Table six records the outcome of an opportunistic behaviour by physicians under FFS.

Tab. 6: Opportunistic behaviour in a FFS scheme

True state	Reported state	Conditions for cheating
Not ill/easy	Not ill difficult	$v_2>1$
Not ill/difficult	Not ill difficult	
Mild ill /high recovery	Mild ill / slow recovery	$v_4>v_3$
Mild ill / slow recovery	Mild ill / slow recovery	
Severely ill	Mild ill / slow recovery	$v_4>1$

For a system in which the physicians are paid on a FFS basis, the incentive to opportunistic behaviour comes from moral hazard as concerns the patients' ability to recover and on the income maximising properties of inflating the number of consultations.

In the presence of asymmetry of information, the process for choosing the organisation of primary health care is no longer sequential as in the case where information is symmetric. The organisation of the service and of the payment scheme have to be combined to find the best institutional arrangement. In this context risk sharing among the two actors plays a different role and attitudes towards risk can only be used for the appraisal of the expected cost of the various alternative systems that can be implemented. The choice of the optimal payment scheme for physicians is quite different from the previous case:

- it is not possible to define an incentive compatible contract only based on risk sharing attitudes
- CG cannot minimise cost through a straight minimisation of the expected cost. The constraints of the problem restrict the set of the option to few well defined contracts and CG will choose the one that minimises expected cost as in the real option literature[12].

5.1 Payment schemes under asymmetry of information: pure risk sharing contracts

To start with, we will assume that the system has a vertical organisation so that each patient has to see his GP before going and see a specialist, if needed. In other words the choice of the patients is not taken into account.

The price for health care is assumed to be equal to the disutility of the effort, i.e. $\gamma_{GP}=d(e)$ as in the previous model. In a symmetric context we showed that the number of visits to be reimbursed, v_i^R is equal, if the system is FFS, to the number of consultations made by the GP, v_i. In the ICC model this is the variable to be used to give the physician the incentive to avoid an opportunistic behaviour. *a* and *b* will be still used to define the variance in the payment system. The problem faced by Central Government can be written as:

$$
\begin{aligned}
Min\quad & p_o\phi(a)+(p_1+p_5)\phi(a+b\gamma_{GP}v_1^R)+p_2\phi(a+b\gamma_{GP}v_2^R)+p_3\phi(a+b\gamma_{GP}v_3^R)+p_4\phi(a+b\gamma_{GP}v_4^R)\\
s.t\quad & p_0\left[U(a)\right]+(p_1+p_5)\left[U(a+bv_1^R\gamma_{GP}-d(e))\right]+p_2\left[U(a+bv_2^R\gamma_{GP}-v_2d(e))\right]\\
& +p_3\left[U(a+bv_3^R\gamma_{GP}-v_3d(e))\right]+p_4\left[U(a+bv_4^R\gamma_{GP}-v_4d(e))\right]\geq 0\\
\text{(II)}\quad & U(a+v_3^Rb\gamma_{GP}-v_3d(e))\geq U(a+v_4^Rb\gamma_{GP}-v_4d(e_{LH}^{cheat}))\\
\text{(III)}\quad & U(a+v_2^Rb\gamma_{GP}-v_2d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e))\\
\text{(IV)}\quad & U(a+v_3^Rb\gamma_{GP}-v_3d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e))\\
\text{(V)}\quad & U(a+v_4^Rb\gamma_{GP}-v_4d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e))\\
\text{(VI)}\quad & V(b\gamma_{GP}(v_2^R-v_1^R))\leq V(E(R)(\Delta n))-\left[g(n+\Delta n)-g(n)\right]\\
\text{(VII)}\quad & V(b\gamma_{GP}(v_3^R-v_1^R))\leq V(E(R)(\Delta n))-\left[g(n+\Delta n)-g(n)\right]\\
\text{(VIII)}\quad & V(b\gamma_{GP}(v_4^R-v_1^R))\leq V(E(R)(\Delta n))-\left[g(n+\Delta n)-g(n)\right]\\
\text{(IX)}\quad & V(b\gamma_{GP}(v_4^R-v_3^R))\leq V(E(R)(\Delta n))-\left[g(n+\Delta n)-g(n)\right]
\end{aligned}
\qquad (12)
$$

(II) represents the incentive compatibility constraint to avoid moral hazard. (III)-(IX) avoid the misrepresentation of the patient's state of illness; the first four are typical of a pure capitation system the other might emerge in a FFS system. Their aim is to stop referring patients that should be treated by the physician or to avoid inappropriate treatment.

The problem presented in equation (12) cannot be solved. The constraints (VI) to (IX) are compatible only with a pure capitation system. They state that to avoid an opportunistic behaviour the increase in income received through cheating should be not greater than the increase in net income that can be derived by enrolling a number (fraction) Δn of new patients. This condition is satisfied only if the marginal cost in enrolling a patient is zero ($g'(n)=0$) or when income deriving from opportunistic behaviour is zero ($b=0;\ v_i^R = \bar{v}^R$). However, to satisfy the constraints (III) to (V) it is necessary to make payment depend on the number of consultations made: the two solutions are then incompatible.

In this environment, the alternative for Central Government is to use the competition among specialists and GP's for patients using an horizontal organisation. Levaggi and Rochaix (2003b) show that under specific conditions, competition reduces the scope for opportunistic behaviour if the system is organised on a FFS basis, or at least if a part of the payment is made on this basis.

We will apply the same model of competition to our problem.

5.2 Access rules to improve the contract

Before showing the choice of the patient in this context, we need to define an opportunistic behaviour by the specialist. The specialist, being paid on a FFS basis, treats disease specific patients as if they were severely ill and refers the one that he cannot treat to the other specialist.

As in the previous case, it is assumed that competition among physicians and specialists can only be developed for treating patients, i.e. for those cases where the patient is ill. If the patient needs reassurance on his health status, we assume that he goes and see the GP.

Given these assumptions, if the patients feels ill, he will go and see the physician that minimises his expected time for being treated. Table seven records the time alternatives for the different options:

Tab. 7: Time cost for the different alternatives

Action	Time cost	Probability
Going to GP	$v_4;v_4;v_4+v_5$	$p_3;p_4;p_5$
Going to SP_A	$v_5;1+v_5;v_5;1+v_5;v_5;1+v_5$	½ p_3; ½ p_3;½ $p_{4;}$ ½ p_4;½ $p_{5;}$ ½ p_5
Going to SP_B	$1+v_5;v_5;1+v_5;v_5;1+v_{5;}v_5$	½ p_3; ½ p_3;½ $p_{4;}$ ½ p_4;½ $p_{5;}$ ½ p_5

[12] See Trigenorgis for a thorough explanation.

The patient will choose the alternative that allows him to maximise his utility. The utility for the three alternatives is equal to:

$$
\begin{aligned}
EU_{GP} &= U(Y) - [p_3 v_4 + p_4 v_4 + p_5 (v_4 + v_5)] \\
&= U(Y) - (p_3 + p_4 + p_5) v_4 - p_5 v_5 \\
EU_{SPa} &= U(Y) - [\frac{1}{2} p_3 v_5 + \frac{1}{2} p_3 (1 + v_5) + \frac{1}{2} p_4 v_5 + \frac{1}{2} p_4 (1 + v_5) + \frac{1}{2} p_5 (v_5) + \frac{1}{2} p_5 (1 + v_5)] \\
&= U(Y) - (p_3 + p_4 + p_5)(v_5 + \frac{1}{2}) \\
EU_{SPb} &= U(Y) - [\frac{1}{2} p_3 (1 + v_5) + \frac{1}{2} p_3 v_5 + \frac{1}{2} p_4 (1 + v_5) + \frac{1}{2} p_4 v_5 + \frac{1}{2} p_5 (1 + v_5) + \frac{1}{2} p_5 (v_5)] \\
&= U(Y) - (p_3 + p_4 + p_5)(v_5 + \frac{1}{2})
\end{aligned}
\tag{13}
$$

It can be shown that if $v_5 < \frac{p_3 + p_4 + p_5}{p_3 + p_4}(v_4 - \frac{1}{2})$ the patient goes and see a specialist, otherwise he will go and see the GP. The specialist (GP) can react to this loss of patients by treating only a fraction z (q) of patient in an opportunistic way. In this case the expected time for patients will be equal to:

$$
\begin{aligned}
EU_{GP} &= U(Y) - q\,[(p_3 + p_4 + p_5) v_4 - p_5 v_5] - (1 - q)[p_3 v_3 + p_4 v_4 + p_5 (1 + v_5)] \\
EU_{SP} &= U(Y) - z\left[(p_3 + p_4 + p_5)(v_5 + \frac{1}{2})\right] - (1 - z)[p_3 v_3 + p_4 v_4 + p_5 (1 + v_5)]
\end{aligned}
\tag{14}
$$

where z^* and q^* represent the optimal choice by the specialist and the GP for the probability of not cheating. In appendix two it is shown that an equilibrium is found for $z=q=1$, i.e. for a situation where the GP and the specialist behave as perfect agents for the patient. In our case, competition can work for mild and severely ill patients. The inflation in visits that can arise from diagnostic procedures cannot be avoided using just the patients choice.

5.3 Payment systems in an asymmetry of information framework

The inflation in the number of consultation for diagnosis can be reduced by using a mixed payment scheme that foresees FFS and horizontal organisation if the patient is ill and capitation/FFS/a mixed payment for diagnosis. The choice of the payment scheme for the diagnostic part depends on the attitude toward risk of the two agents. The problem for Central Government can be written as:

$$
\begin{aligned}
Min \quad & p_o\phi(a)+(p_1+p_5)\phi(a+b\gamma_{GP}v_1^R)+p_2\phi(a+b\gamma_{GP}v_2^R)+p_3\phi(a+b\gamma_{GP}v_3^R)+p_4\phi(a+b\gamma_{GP}v_4^R) \\
s.t. \quad & p_0\left[U(a)\right]+(p_1+p_5)\left[U(a+bv_1^R\gamma_{GP}-d(e))\right]+p_2\left[U(a+bv_2^R\gamma_{GP}-v_2d(e))\right] \\
& +p_3\left[U(a+bv_3^R\gamma_{GP}-v_3d(e))\right]+p_4\left[U(a+bv_4^R\gamma_{GP}-v_4d(e))\right]\geq 0
\end{aligned}
$$

$$
\begin{aligned}
&(II) \quad U(a+v_2^Rb\gamma_{GP}-v_2d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e)) \\
&(III) \quad U(a+v_3^Rb\gamma_{GP}-v_3d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e)) \\
&(IV) \quad U(a+v_4^Rb\gamma_{GP}-v_4d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e)) \\
&(V) \quad V(b\gamma_{GP}(v_2^R-v_1^R))\leq V(E(R)(\Delta n))-\left[g(n+\Delta n)-g(n)\right]
\end{aligned}
\tag{15}
$$

The problem written in these terms cannot be solved since constraint (II) and (V) are incompatible. (V) implies that $v_2^R=v_1^R$ while the second constraint implies that $v_2^R\geq\frac{v_2-v_1}{b}+v_1^R$.

In this case, Central Government has to choose between inflation on the consultations made by GP or having patients not ill, but for whom the diagnosis is difficult to be referred to the specialist. The choice between each solution depends on the environment in which Central Government works and on his attitude towards risk. From the point of view of our analysis, both solutions have to be identified.

If Central Government wants to avoid the risk of referral to specialist, the problem can be written as:

$$
\begin{aligned}
Min \quad & p_o\phi(a)+(p_1+p_5)\phi(a+b\gamma_{GP}v_1^R)+p_2\phi(a+b\gamma_{GP}v_2^R)+p_3\phi(a+b\gamma_{GP}v_3^R)+p_4\phi(a+b\gamma_{GP}v_4^R) \\
s.t. \quad & p_0\left[U(a)\right]+(p_1+p_5)\left[U(a+bv_1^R\gamma_{GP}-d(e))\right]+p_2\left[U(a+bv_2^R\gamma_{GP}-v_2d(e))\right] \\
& +p_3\left[U(a+bv_3^R\gamma_{GP}-v_3d(e))\right]+p_4\left[U(a+bv_4^R\gamma_{GP}-v_4d(e))\right]\geq 0
\end{aligned}
$$

$$
\begin{aligned}
&(II) \quad U(a+v_2^Rb\gamma_{GP}-v_2d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e)) \\
&(III) \quad U(a+v_3^Rb\gamma_{GP}-v_3d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e)) \\
&(IV) \quad U(a+v_4^Rb\gamma_{GP}-v_4d(e))\geq U(a+v_1^Rb\gamma_{GP}-d(e)
\end{aligned}
\tag{16}
$$

The solution is presented in appendix 3. The incentive compatible constraints in the problems means that the consultations in excess of v_1 have to be paid using a FFS. For CG the choice is then to use a pure FFS or a mixed payment system where a fixed fee is paid to the GP for any patient in his list and only the consultations in excess of v_1 can be reimbursed. The choice depends on the risk sharing attitudes of the two actors as shown in table 8.

Tab. 8: The optimal incentive system

Central Government \ Physician	Risk averse	Risk neutral
Risk averse	Any	Mixed $a = (1-p_0)d(e) = (1-p_0)\gamma_{GP}$
Risk neutral	FFS $a = 0;$ $\quad$ $b = 1$	Any

It is interesting to note that in this case, risk sharing is limited to the diagnostic part of the care delivered. For anything that implies delivering health care, a FFS system has to be devised to make use of the competition among physician generated by patients behaviour.

If Central Government wants to avoid the opportunistic behaviour deriving from inflation in the diagnostic tests, he has to foresee the same reimbursement for all the consultations involving diagnostics. The problem can be written as:

$$\begin{aligned}
Min \quad & p_o\phi(a) + (p_1+p_5)\phi(a+b\gamma_{GP}v_1^R) + p_2\phi(a+b\gamma_{GP}v_2^R) + p_3\phi(a+b\gamma_{GP}v_3^R) + p_4\phi(a+b\gamma_{GP}v_4^R) \\
s.t. \quad & p_0\left[U(a)\right] + (p_1+p_5)\left[U(a+bv_1^R\gamma_{GP}-d(e))\right] + p_2\left[U(a+bv_2^R\gamma_{GP}-v_2d(e))\right] \\
& + p_3\left[U(a+bv_3^R\gamma_{GP}-v_3d(e))\right] + p_4\left[U(a+bv_4^R\gamma_{GP}-v_4d(e))\right] \geq 0 \\
& \text{(III)} \quad U(a+v_3^Rb\gamma_{GP}-v_3d(e)) \geq U(a+v_1^Rb\gamma_{GP}-d(e)) \qquad (17) \\
& \text{(IV)} \quad U(a+v_4^Rb\gamma_{GP}-v_4d(e)) \geq U(a+v_1^Rb\gamma_{GP}-d(e)) \\
& \text{(V)} \quad V(b\gamma_{GP}(v_2^R-v_1^R)) \leq V(E(R)(\Delta n)) - \left[g(n+\Delta n)-g(n)\right]
\end{aligned}$$

which implies that $v_2^R = v_1^R$ The optimal solution is presented in appendix three and it is similar to the one just presented for the previous case.

Tab. 9: Optimal incentive scheme: referall inflation

CG \ Physician	Risk averse	Risk neutral
Risk averse	Any	Mixed $a = (1-p_0)v_2d(e) = (1-p_0))v_2\gamma_{GP}$ $v_1^R > v_2$
Risk neutral	FFS $a = 0;$ $\quad$ $b = 1$ $v_1^R > \frac{(1-p_0-p_2)v_1 + p_2v_2}{1-p_0} \geq v_1 + \frac{p_2}{1-p_o}(v_2-v_1)$	Any (Mixed?)

If CG is risk averse and the GP risk neutral, a mixed system which foresee the payment of a capitation fee for each patient in the list and a FFS for any consultation in excess of v_2 is chosen. If Central Government is risk neutral and the GP is risk averse, a pure FFS will be chosen, but the GP is paid a nominal number of consultations v_1^R for diagnostic.

5.4 Comparing the two solutions

At the beginning of this section we argue that CG's choice of the optimal system to organise primary care is limited by asymmetry of information. These limitations have at least effects:

- to avoid demand inducement on treatment, a FFS system with horizontal organisation where the patient can choose his physician has to be implemented;
- no first best optimal solution can be implemented: CG has to choose between a system with demand inducement through diagnosis or referral inducement for the same reason.

CG chooses the second best system by a real option appraisal; i.e. rather than minimising his utility function over a continuous set of options, he has to identify a finite number of solutions and choose the one that minimise his cost function.

The result of this process cannot be predicted in general since it depends on the utility function of CG, on the attitude towards risk of the GP and on the cost of having a patient referred to the specialist for diagnosis.

For some cases, where the costs can be easily defined such a comparison is possible. To show the general method, we will see how a CG that is risk neutral makes its choice. We will assume that $\gamma_{GP}=d(e)=1$. If CG allows GP's to induce expenditure through diagnosis, the expected cost for health care will be equal to:

$$(p_1 + p_2)v_2 + FFS_H^T$$

where FFS_H^T is the expected cost for treating patients under an horizontal FFS system. If instead CG allows them to over-refer to the specialists the costs will be:

$$(p_1 + p_2)v_R^1 + FFS_H^T + p_2\gamma_{SP}v_{SP}$$

where v_{SP} is the number of consultations that the specialist has to make to for a difficult diagnosis.

If we assume that the visits made by the specialist for a difficult diagnosis are the same number as those made by the GP ($v_{SP}=v_2$) and that $\gamma_{SP}=\gamma_{GP}(1+\mu)$, we can see that a pure FFS is preferred.

If CG is risk averse, a general solution cannot be derived since the choice depends on the functional form of the expenditure function defined for this tier.

6.Conclusions

This paper defines the strategy CG should follow in defining the payment system for primary care in an asymmetry of information context where citizens can be ill with a different level of severity and recovery with known probabilities. It is shown that in a symmetric context the choice can be sequential: CG chooses the payment scheme according to his attitude towards risk and to GP's risk aversion. Given this choice, the organisation of the service and access to medical care will be defined. In this context a vertical system is usually less expensive than an horizontal organisation where patients are allowed to choose.

For capitation, this is a general rule implicit in the design of the payment scheme. For FFS it depends on the probability attached to each health status and on the priors patients have about their illness.

In a context of asymmetry of information on the health status of the population if this element can be observed only by the physicians (GPs and specialists), a first best solution cannot be designed. Demand inducement can be partially reduced through the competition for patients between GP's and specialists, but demand inducement for diagnosis cannot be avoided. CG can in fact control either demand inducement by GPs or the referral process to the specialist, but in both cases there is an extra cost for the system. In an asymmetry of information framework, the choice of the optimal system cannot be done using a straightforward maximisation process; the constraints imposed by the incentive compatible structure of the problem means that the set of viable contracts is very limited an real option approach has to be used to define the best solution.

In this context, demand inducement can be reduced only using other instruments such as overall budgets for consultations or on the number of diagnostic tests. Both instruments have not however been introduced in our model. Another way to avoid demand inducement is to relax the assumption that each patient has to recover his full health. The combined use of effort and payment as contingent variables might avoid demand inducement[13].

The model presented here can be developed further in several directions: the priors of patients should be introduced in order to see if our results, in particular the superiority of FFS and horizontal organisation, is robust to a change in this assumption; time preferences could be made dependent on the severity of illness as in Levaggi and Rochaix (2003b) and finally…

[13] Levaggi and Rochaix (2003a) presents a model in this line

Appendix one

The problem for Central Government can be written as:

$$Max_{a,b} \quad - \left\{ p_o\phi(a) + (p_1 + p_2)\phi(a + b\gamma_{GP}) + p_2\phi(a + b\gamma_{GP}v_2^R) + p_3\phi(a + b\gamma_{GP}v_3^R) + p_4\phi(a + b\gamma_{GP}v_4^R) \right\}$$

$$\begin{aligned} s.t \quad & p_0 \left[U(a)\right] + (p_1 + p_2)\left[U(a + bv_1^R\gamma_{GP} - d(e))\right] + p_2\left[U(a + bv_2^R\gamma_{GP} - v_2 d(e))\right] \\ & + p_3\left[U(a + bv_3^R\gamma_{GP} - v_3 d(e))\right] + p_4\left[U(a + bv_4^R\gamma_{GP} - v_4 d(e))\right] \geq 0 \\ & a \geq 0 \\ & b \geq 0 \\ & b \leq 1 \end{aligned}$$

The F.O.C. can be written as:

$$\frac{\partial L}{\partial a} :- \left[p_o\phi'(a) + (p_1 + p_5)\phi'(a + b\gamma_{GP}) + p_1\phi'(a + b\gamma_{GP}v_2^R) + p_3\phi'(a + b\gamma_{GP}v_3^R) + p_4\phi'(a + b\gamma_{GP}v_4^R)\right]$$
$$-\lambda \left\{ \begin{array}{l} p_0\left[U'(a)\right] + (p_1 + p_5)\left[U'(a + bv_1^R\gamma_{GP} - d(e))\right] + p_2\left[U'(a + bv_2^R\gamma_{GP} - v_2 d(e))\right] + \\ p_3\left[U'(a + bv_3^R\gamma_{GP} - v_3 d(e))\right] + p_4 U'(a + bv_4^R\gamma_{GP} - v_4^R d(e)) \end{array} \right\}$$

$$\frac{\partial L}{\partial b} :- \left[(p_1 + p_5)\phi'(a + b\gamma_{GP})\gamma_{GP} + p_1\phi'(a + b\gamma_{GP}v_2^R)\gamma_{GP}v_2^R + p_3\phi'(a + b\gamma_{GP}v_3^R)\gamma_{GP}v_3^R + p_4\phi'(a + b\gamma_{GP}v_4^R)\gamma_{GP}v_4^R\right]$$
$$-\lambda \left\{ \begin{array}{l} (p_1 + p_5)\left[U'(a + bv_1^R\gamma_{GP} - d(e))\gamma_{GP}\right] + p_2\left[U'(a + bv_2^R\gamma_{GP} - v_2 d(e))\gamma_{GP}v_2^R\right] + \\ p_3\left[U(a + bv_3^R\gamma_{GP} - v_3 d(e))\gamma_{GP}v_3^R\right] + p_4\left[U(a + bv_4^R\gamma_{GP} - v_4 d(e))\gamma_{GP}v_4^R\right] \end{array} \right\}$$

$$\frac{\partial L}{\partial \lambda} : \; p_0\left[U(a)\right] + (p_1 + p_5)\left[U(a + bv_1^R\gamma_{GP} - d(e))\right] + p_2\left[U(a + bv_2^R\gamma_{GP} - v_2 d(e))\right] + p_3 U(a + bv_3^R\gamma_{GP} - v_3 d(e))$$
$$+ p_4\left[U(a + bv_4^R\gamma_{GP} - v_4^R d(e))\right]$$

The problem can be solved using the Khun Tucker conditions.

Let us start by assuming that CG is risk neutral and GP is risk averse. If we divide the derivative for a by the derivative for b we obtain:

$$\frac{1}{\gamma_{GP}E(v)(1 - p_o)} = \frac{E(U_a^{'})}{E(U_b^{'})\gamma_{GP}E(v)} = \frac{E(U_a^{'})}{\left[E(U_a^{'}) - p_o U_a^{'}\right]\gamma_{GP}E(v)}$$

which can never be satisfied. This means that a=0. Solving for the second and third constraint we obtain b=1.

If GP is risk neutral and CG is risk averse we obtain:

$$\frac{E(\phi_a^{'})}{E(\phi_b^{'})\gamma_{GP}E(v)} = \frac{E(\varphi_a^{'})}{\left[E(\phi_a^{'}) - p_o\phi_a^{'}\right]\gamma_{GP}E(v)} = \frac{1}{\gamma_{GP}E(v)(1 - p_o)}$$

In this case, b=0 and a=$E(v)\gamma_{GP}$

When both are risk averse, any solution is compatible since:

$$\frac{E(\phi_a^{'})}{E(\phi_b^{'})\gamma_{GP}E(v)} = \frac{E(U_a^{'})}{\left[E(U_a^{'}) - p_o U_a^{'}\right]\gamma_{GP}E(v)}$$

when both are risk neutral:

$$\frac{1}{\gamma_{GP}E(v)(1 - p_o)} = \frac{1}{\gamma_{GP}E(v)(1 - p_o)}$$

Appendix 2:

If, as in the previous examples, we assume that there are no priors on the probabilities so that p_i=1/6, the expected utility can be written as:

$$EU_{GP} = U(Y) - q\left[(p_3 + p_4 + p_5)v_4 + p_5 v_5\right] - (1-q)\left[p_3 v_3 + p_4 v_4 + p_5(1+v_5)\right]$$
$$EU_{SP} = U(Y) - z\left[(p_3 + p_4 + p_5)(v_5 + \frac{1}{2})\right] - (1-z)\left[p_3 v_3 + p_4 v_4 + p_5(1+v_5)\right]$$

If we observe the derivatives of the two expressions:

$$\frac{\partial EU_{GP}}{\partial q} = -\left[(p_3 + p_4 + p_5)v_4 + p_5 v_5 - p_3 v_3 - p_4 v_4 - p_5(1+v_5)\right] < 0$$
$$\frac{\partial EU_{SP}}{\partial z} = -\left[(p_3 + p_4 + p_5)(v_5 + \frac{1}{2}) - p_3 v_3 - p_4 v_4 - p_5(1+v_5)\right] < 0 \qquad \text{(a2)}$$

which are linear in z and q. The relative magnitude of the derivatives depends on the value of vi and pi, but there is a negative correlation between the growth rate of the function and its initial value. This means that if for values of q and z equal to one, going to the GP (specialist) is the best alternative, a reduction at the same rate of the cheating will cause a bigger (smaller) increase in the expected utility for going and see the specialist. In this way the difference between the two levels of utility will shrink.

From equation (a2) we can write the reaction functions of the GP's and the specialist as follows:

$$z = q\frac{p_3(v_4 - v_3) - p_5(v_4 - 1)}{(p_3 + p_4 + p_5)(v_5 + \frac{1}{2}) - p_3 v_3 - p_4 v_4 - p_5(1+v_5)}$$
$$q = z\frac{(p_3 + p_4 + p_5)(v_5 + \frac{1}{2}) - p_3 v_3 - p_4 v_4 - p_5(1+v_5)}{p_3(v_4 - v_3) - p_5(v_4 - 1)}.$$

If we assume that the GP and the specialist cannot play any co-operative game, they both will try to maximise their probability of getting patients in the worst event. For both of them the worst situation is when the other part does not cheat, i.e. when z=0 and/or q=0.

Appendix 3

$$Max_{a,b} \quad - \{p_o\phi(a) + (p_1+p_2)\phi(a+b\gamma_{GP}) + p_2\phi(a+b\gamma_{GP}v_2^R) + p_3\phi(a+b\gamma_{GP}v_3^R) + p_4\phi(a+b\gamma_{GP}v_4^R)$$

$$sp_0 \, [U(a)] + (p_1+p_2)\,[U(a+bv_1^R\gamma_{GP} - d(e))] + p_2\,[U(a+bv_2^R\gamma_{GP} - v_2 d(e))]$$
$$+ p_3\,[U(a+bv_3^R\gamma_{GP} - v_3 d(e))] + p_4\,[U(a+bv_4^R\gamma_{GP} - v_4 d(e))] \geq 0$$

(II) $\quad U(a+v_2^R b\gamma_{GP} - v_2 d(e)) \geq U(a+v_1^R b\gamma_{GP} - d(e))$

(III) $\quad U(a+v_3^R b\gamma_{GP} - v_3 d(e)) \geq U(a+v_1^R b\gamma_{GP} - d(e))$

(IV) $\quad U(a+v_4^R b\gamma_{GP} - v_4 d(e)) \geq U(a+v_1^R b\gamma_{GP} - d(e))$

To simplify the algebra we assume γ=d(e)=1. The system of equations can be written as:

(II) $\quad v_2^R \geq \frac{v_2 - v_1}{b} + v_1^R$

(III) $\quad v_3^R \geq \frac{v_3 - v_1}{b} + v_1^R$

(IV) $\quad v_4^R \geq \frac{v_4 - v_1}{b} + v_1^R$

Since b is always multiplied by v, we take v as benchmark. The problem can be written as:

$$Min \quad p_o\phi(a) + (p_1+p_5)\phi(a+bv_1) + p_2\phi(a+v_2+(b-1)v_1) + p_3\phi(a+v_3+(b-1)v_1) + p_4\phi(a+v_3+(b-1)v_1)$$

$$s \quad p_0 U(a) + (1-p_0)U(a+(b-1)v_1)$$

$$a \geq 0 \qquad b \geq 0 \qquad b \leq 1$$

The F.O.C. can be written as:

$$\frac{\partial L}{\partial a} :- [p_o\phi'(a) + (p_1+p_5)\phi'(a+bv_1) + p_2\phi'(a+v_2+(b-1)v_1) + p_3\phi'(a+v_3+(b-1)v_1) + p_4\phi'(a+v_4+(b-1)v_1)]$$
$$-\lambda\,\{p_0\,[U'(a)] + (1-p_0)\,[U'(a+(b-1)v_1]\}$$

$$\frac{\partial L}{\partial b} :- [v_1\,[(p_1+p_5)\phi'(a+bv_1) + p_2\phi'(a+v_2+(b-1)v_1) + p_3\phi'(a+v_3+(b-1)v_1) + p_4\phi'(a+v_4+(b-1)v_1)]]$$
$$-\lambda(1-p_0)\,[U'(a+(b-1)v_1]$$

$$\frac{\partial L}{\partial \lambda} : \quad p_o\phi(a) + (p_1+p_5)\phi(a+bv_1) + p_2\phi(a+v_2+(b-1)v_1) + p_3\phi(a+v_3+(b-1)v_1) + p_4\phi(a+v_3+(b-1)$$

Commentaire [L4]: 27

The solution can be obtained using the method described in appendix one.

Appendix 4

$$Max_{a,b} \quad -\left\{p_o\phi(a) + (p_1+p_2)\phi(a+b\gamma_{GP}) + p_2\phi(a+b\gamma_{GP}v_2^R) + p_3\phi(a+b\gamma_{GP}v_3^R) + p_4\phi(a+b\gamma_{GP}v_4^R)\right\}$$

$$sp_0\left[U(a)\right] + (p_1+p_2)\left[U(a+bv_1^R\gamma_{GP} - d(e))\right] + p_2\left[U(a+bv_2^R\gamma_{GP} - v_2 d(e))\right]$$
$$+ p_3\left[U(a+bv_3^R\gamma_{GP} - v_3 d(e))\right] + p_4\left[U(a+bv_4^R\gamma_{GP} - v_4 d(e))\right] \geq 0$$

(III) $\quad U(a+v_3^R b\gamma_{GP} - v_3 d(e)) \geq U(a+v_1^R b\gamma_{GP} - d(e))$

(IV) $\quad U(a+v_4^R b\gamma_{GP} - v_4 d(e)) \geq U(a+v_1^R b\gamma_{GP} - d(e))$

(V) $\quad (b\gamma_{GP}(v_2^R - v_1^R)) \leq V(E(R)(\Delta n)) - \left[g(n+\Delta n) - g(n)\right]$

To simplify the algebra we assume γ=d(e)=1. The system of equations can be written as:

(III) $\quad v_3^R \geq \frac{v_3 - v_1}{b} + v_1^R$

(IV) $\quad v_4^R \geq \frac{v_4 - v_1}{b} + v_1^R$

(V) $\quad v_2^R = v_1^R$

$$Min \quad p_o\phi(a) + (p_1+p_2+p_5)\phi(a+bv_1^R) + p_3\phi(a+v_3+(b-1)v_1^R) + p_4\phi(a+v_3+(b-1)v_1^R)$$

$$s \quad p_0U(a) + (1-p_0-p_2)U(a+(b-1)v_1) + p_2U(a+bv_1^R - v_2)$$

$$a \geq 0 \qquad b \geq 0 \qquad b \leq 1$$

The F.O.C. can be written as:

$$\frac{\partial L}{\partial a}:- \left[p_o\phi'(a) + (p_1+p_2+p_5)\phi'(a+bv_1^R) + p_3\phi'(a+v_3+(b-1)v_1^R) + p_4\phi'(a+v_4+(b-1)v_1^R)\right]$$
$$-\lambda\left\{p_0\left[U'(a)\right] + (1-p_0-p_2)\left[U'(a+(b-1)v_1^R\right] + p_2\left[U'(a+bv_1^R - v_2\right]\right\}$$

$$\frac{\partial L}{\partial b}:- \left[v_1^R\left[(p_1+p_2+p_5)\phi'(a+bv_1^R) + p_3\phi'(a+v_3+(b-1)v_1^R) + p_4\phi'(a+v_4+(b-1)v_1^R)\right]\right]$$
$$-\lambda\left\{(1-p_0-p_2)\left[U'(a+(b-1)v_1^R\right]v_1^R + p_2\left[U'(a+bv_1^R - v_2\right]v_1^R\right\}$$

$$\frac{\partial L}{\partial \lambda}: \quad p_o\phi(a) + (p_1+p_2+p_5)\phi(a+bv_1^R) + p_3\phi(a+v_3+(b-1)v_1^R) + p_4\phi(a+v_3+(b-1)v_1^R)$$

The solution can be obtained using the method described in appendix one.

References

Chalkley M., 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.

Coscelli, Andrea. " The Importance of Doctors' and Patients' Preferences in the Prescription Decision " Journal of Industrial Economics, Sep 01, 2000, Vol. 48, Issue 3, p 349.

Culyer, A. J., A.K. Maynard and A. Williams (1981) Alternative systems of health care provision: an essay on motes and beams, in Olson , M. (ed) A new approach to the Economics of Health Care, Washington, American Enterprise Institute

Delnoij, D .M.J., (1994), « Physician payment systems and cost control », NIVEL.

Elliott, R. F., (1991), Labor Economics: a comparative analysis, McGraw Hill, 89-92.

Ellis R., Mc Guire T., (1986), Provider Behaviour under Prospective Reimbursement, *Journal of Health Economics*, 5(2):129-51.

Ellis R., Mc Guire T., (1990), Optimal Payment schemes for Health Services, *Journal of Health Economics*, 9(4):376-96.

Evans, R. G. (1974) «Supplier- induced demand: some empirical evidences» in Perlman, M. (ed.) *The economics of Health and Medical Care*, New York, John Wiley

Franc, C. (2000) Rémuneration des médicins: Qualité ou Efficacité?, *mimeo*

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

Gravelle H and Masiero G. (2000) "Quality incentives in a regulated market with imperfect information and switching costs: capitation in general practice ", *Journal of Health Economics* (19) 2000;

Gravelle H and Masiero G. (2002) "Quality incentives under a capitation regime: the role of patients' expectations", In: Lindgren, B (ed) *Individuals Decisions for Health*. London: Routledge,

Hughes, D., Yule B., (1992), The Effect of per-item Fees on the Behaviour of Primary Health Care, *Journal of Health Economics*, 11(4):413-438.

Jaegher, K, Jegers M., (2000), A Model of Physician Behaviour with Demand Inducement, *Journal of Health Economics,* 19:231-258

Kristiansen, I.S. and Mooney, G., (1993), « Remuneration of GP services: time for more explicit objectives? a review of the systems in five industrialised countries », Health Policy, 24, pp. 203-12.

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

Levaggi, R. e L. Rochaix (2003a) Optimal incentive schemes for physicians, *Portuguese Economic Review*, 2, 87-107

Levaggi, R., Rochaix, L., (2003b), Alternative Systems for Organising and Paying Primary Health Care Services: the Role of Patients' Choice, *Paper presented at the 4th ECHE conference*, Paris, 7-10 July 2002.

Lewis T, Sappington D., (1997), Cream skimming, cost sharing and subjective risk adjusting in health care industry, University of Florida, March.

McGuire, T., (2000), "Physician Agency", Handbook of Health Economics, volume 1, chapter 9, eds. A.J. Culyer and P.J. Newhouse, Elsevier Science, 461-536.

McGuire, T., Pauly M., (1991), Physician Response to Fee Changes with Multiple Payers, *Journal of Health Economics* , 10:385-410.

Mossialos

Pauly, MV and Satterthwathe, M.A. (1981) The pricing of primary care physician services: a test for the role of consumer's information, *Bell Journal of Economics*, vol. 12, pp. 488-506

Pauly, MV, (1980), « Doctors and their workshops : economic models of physician behaviour », NBER, University of Chicago Press.

Reinhardt U., (1983), The Theory of Physician-Induced Demand and its Implication for Health Policy, research supported by DHHS/HCFA grant No 95-P97309/2.

Reinhardt, U., (1985), "Editorial - The Theory of Physician-Induced Demand: Reflections after a Decade", Journal of Health Economics, Vol. 4, pp. 187-193.

Rice, T., (1984), "Physician Initiated Demand for Medical Care: New Evidence from the Medicare Program", Advances in Health Economics and Health Services Research, Vol. 5, pp. 129-160.

Rice, T., and Labelle, R.G., (1988), "Do Physicians Induce Demand for Medical Services?", Journal of Health Politics, Policy and Law, Vol. 14, (3), pp. 587-625.

Robinson, J.C., (1993), Payment Mechanisms, Nonprice Incentives, and Organisational Innovation in Health Care , *Inquiry*, (Fall), 30(3):328-33.

Robinson, J.C., (2001), Theory and Practice in the Design of Physician Payment Incentives, *The Millbank Memorial Fund Quarterly*, Volume 79(2) June 2001.

Rochaix, L. (1997) Asymétries d'information et incertitude en santé: les apports de la théorie des contracts, Economie et Prevision, n.129-130, pp.11-24

Rochaix, L., (1993) « Financial Incentives for Physicians : the Québec experience », Health Economics, vo 2, pp. 163-176.

Rochaix, L., (1998a), Performance Tied Payment Systems for Physicians, in Critical Challenges for Health Care Reforms in Europe, Saltman R., Figueras J., Sakellarides C. (eds), Open University Press, 196-218.

Rochaix, L., (1998b), The Physician as Perfect Agent : a Comment, *Social Science and Medicine*, 47(3):355-356.

Scott, A., Shiell A., (1997), Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model, *Journal of Health Economics,* 16:323-342.

Scott, S., J. Hall, (1995), Evaluating the Effects of GP Remuneration: Problems and Prospects , *Health Policy*, 31:183-195.

Stearns, S., Wolfe, B.L., Kindig, D.A., (1992), Physician responses to fee-for-service and capitation payment , *Inquiry*, 29(4):416-25.

Trigenorgis, L. (1998) Real options, The MIT Press, Cambridge, MA