Getting the GP to make the right diagnosis: the role of economic incentives.*

Carine Franc † Gérard de Pouvourville ‡

September 30, 2005

Abstract

We study the effects of the reimbursement design of General Practitioners on both their recourse to some expensive tests to acquire information on the type of illness and their effort that improves several dimensions of the quality of the treatment. The GP runs the diagnosis procedure and according to the result, she prescribes the treatment taking into account her financial incentives. We analyze the optimal reimbursement scheme that allows the government to decentralize the social optimal solution by using different tools combining parameters associated to fee-for-service, capitation payment and a cost-sharing parameter concerning the tests.

Keywords: Physicians' payment; quality; diagnosis procedure; public policy.

JEL classification: I18; D82;

*We would like to thank the participants at the 4^{th} European Conference on Health Economics, Paris 2002, at the "Séminaire d'économie de la santé et protection sociale", Paris 2002 and at the Cregas, Paris 2003.

†Corresponding author: CREGAS, INSERM U537/ CNRS UMR 8052, 80 rue du Général Leclerc. 94276 Le Kremlin-Bicêtre cedex. France. Email: franc@kb.inserm.fr

‡CREGAS, INSERM U537/ CNRS UMR 8052, 80 rue du Général Leclerc 94276 Le Kremlin-Bicêtre cedex, France.

1 Introduction

This paper deals with the diagnosis process of a General Practitioner (GP). In particular, we focus on the impact of payment schemes on the decision to prescribe or not a complementary test, following the initial clinical observation. Actual payment schemes may be quite different from one country to another, and it is plausible that they make a difference in physician behavior, if they have an impact on her income. In France, GPs perform very little testing themselves, be they lab tests, imagery diagnostic procedures or other tests, like electrocardiograms. The tests are performed by an other professional. Thus, in principle, the GP has no direct economic incentive to prescribe tests, except in the case of collusion with another professional. The GP does not bear the cost of the test either. In Germany, GPs perform lab tests at their office, and are paid fee-for-service. Thus, testing is an important part of their income. In the UK, GPs hold a budget for diagnostic procedures and lab tests for patients on their list, and thus have a direct incentive to monitor their decisions. In managed care organizations in the USA, insurers design guidelines and financial incentives to control for the cost of testing.

Complementary tests may represent a substantial induced cost in primary care. In France, in 2002, lab tests prescribed by private-based practitioners represented 20% of physician fees. Thus, payers should be concerned by the efficiency of the diagnostic process. The choice of the test must be adequate, which means that the prior clinical observation of the patient must be thorough enough to narrow the set of possible diagnoses: according to Bayes' theorem, if the true occurrence of a disease is very low, even a very effective test (with a high sensitivity and specificity) will yield a great number of false positive cases, thus increasing the number of unnecessary treatments. This suggests that testing must not be an alternative to the physician's initial observation, since this observation will allow an initial screening of risk factors that will improve the performance of complementary diagnostic tools. Again, economic incentives may have an impact on physician choices. In a capitation system, where the GP gets a fixed amount for each patient who is registered at her office for a given period, there is an incentive for the physician not to provide an important effort at each visit, and eventually to avoid repeated visits for the same episode of illness. Thus, one could argue that she will be more inclined to prescribe testing or to refer the patient to a specialist. With a fee-for-service payment with no cap on volume, the effort should be proportioned to the amount of the fee, but the GP has the possibility to perform more visits for the same episode, to exert some surveillance on the evolution of the health problem of her patient.

Because GPs are first line providers, they are particularly concerned by the trade-off between their effort and complementary testing. They are most often confronted to health conditions with a good prognosis, which improve rapidly with adequate treatment and/or reassurance from the Doctor. Moreover, it has been demonstrated (SFMG, 2000) that in 60 to 70% of all cases, the GP will not be able to identify precisely the cause of the condition, either because it is unknown considering the state of medical knowledge, or because repeated experience has demonstrated that systematic investigations lead to negative results and that the condition improves spontaneously. Such conditions are identified by symptoms, and could thus be considered as not requiring much effort from the GP. But if in the great majority of cases, the prognosis is actually good, the probability that such symptoms may be predictive of a severe disease is low but not equal to zero. For example, an abdominal pain may signal an ulcer or a potential heart failure. In general, after a thorough examination, the GP can dismiss such risks and help to the resolution of the problem by the prescription of an adequate symptomatic treatment and by reassuring the patient. This is why some countries have adopted the gatekeeper model. In such situations, Bayes' theorem finds a direct application: because the probability of a severe disease is low, the positive predictive value of testing is also low. Specialists benefit from the first screening done by the first line and patients referred to them have a higher probability of suffering from a complex or more severe illness.

In this paper, we attempt to model the effects of different payment schemes on GP behavior, considering the specificities of her practice and decision process. Specifically, we focus on the decision to prescribe a complementary test to set a diagnosis and thus improve treatment relevance. Tests and effort are considered as substitutes: tests may yield information that could be obtained from a detailed medical examination. We consider only substitution because of the context of general practice: the majority of cases will be benign, the probability of occurrence of a serious disease will be low. This choice excludes two types of cases. In the first type, symptoms are clearly predictive of a disease, benign or severe, and complementary testing will be used mainly to adjust the treatment or to screen for other non clinical risk factors. This may be the case for moderate hypertension observed in a smoking male patient over 50, with low physical exercise and overweight: the Doctor may look for cholesterol, or more severe coronary or heart disease. Testing is then a complement to effort. The second type is the case of symptoms clearly predictive of a serious disease, and the Doctor may directly refer the patient to a specialist or to a hospital. Testing will then be performed by the secondary level of care.

The aim is then to define what is the optimal payment scheme from the perspective of the payer, that ensures the maximum benefit for the patient at the lowest cost. We have adopted the most general case, which combines three different modes of payment: a capitation fee, a fee-for-service and cost-sharing for supplementary investigations.

2 The model

In the economy there are different types of agents: general practitioners GP and other care providers providing among other lab tests, patients and the government playing the role of the insurer.

2.1 The patient

At the beginning of the period, at the time of a potential registration, the representative patient holds a gross wealth W and his initial state of health provides him the welfare H_0. The preferences are represented by a separable utility function in consumption and health such that:

$$U(C, H) = u(C) + H$$

where C denotes the consumption and H the welfare resulting from the final state of health. The function $u(.)$ has the standard properties to represent risk aversion: $u'(.) > 0$ and $u''(.) < 0$.

With the probability p, the individual falls ill and visits his GP[1]. The illness may be

[1] According to the capitation system, individuals are registered by a physician at the beginning of the

either serious requiring a high level of treatment (probability μ) or benign (probability $1-\mu$). The probability μ called "true" probability is assumed to be exogenous (there is no *ex ante* moral hazard) and unknown by both the patient and the physician. In general practice, μ that can also be viewed as the prevalence of the serious disease, is assumed to be very small. Depending on the type of illness, the welfare resulting from the final state of health without treatment is such that $H_s < H_b < H_0$ (where the subscripts indicate the severity of the illness), H_s being obviously the worst in case of a serious illness.

2.2 The General Practitioner

2.2.1 The diagnosis process

If p and μ are only determined by nature, during the first visit of the overall treatment process the GP runs a detailed examination that allows her to establish an initial diagnosis, her own perception of the risk. This diagnosis is characterized by the standard properties of the sensibility S_e and the specificity S_p. In screening for a disease, S_e corresponds to the ability for the GP to detect the serious cases among the patients having the serious disease; S_e is the proportion who are correctly identified by the diagnosis process among serious cases. S_p corresponds to the probability for the GP to detect the benign cases among the patients having a benign illness; in other words, S_p is the proportion without a serious disease correctly identified.

During the visit that costs c to the GP (unitary constant cost), she has to decide wether or not it is useful to run some supplementary tests X to improve her perception of the risk and thus the quality of the diagnosis. She may also exert an effort e by running for instance a more thorough and longer medical examination. The cost of the effort is directly born by her in terms of working time and the cost is ce. On the opposite, as the supplementary tests are provided by some other providers, the GP may only share a part of the cost of the tests if the public insurer implements such a mechanism. The unitary price of X is a constant denoted $\pi(>1)$. In other words, we call supplementary tests the part of tests that give some information that could be obtained from a detailed medical examination[2]. The

period.

[2] The effort could also be viewed as the effort exerted in keeping the information on the patient after a

components e and X allow the GP to better evaluate the risk and to improve the quality of the diagnosis process such that the final sensitivity and specificity are respectively higher[3]:

$$\begin{aligned} f(S_e; e, X) &= S_e [1 + \theta(e, X)] \\ f(S_p; e, X) &= S_p [1 + \theta(e, X)] \end{aligned}$$

where $\theta(e, X)$ represents the adjustment parameter such that $\theta(0,0) = 0$, $\theta'_e(e, X) > 0$, $\theta''_{ee}(e, X) < 0$, $\theta'_X(e, X) > 0$, and $\theta''_{XX}(e, X) < 0$.

As μ is the prevalence of the serious illness in the population of patients (or $p\mu$ for all the population) and according to the definitions of S_e and S_p, we can infer the probability of the different states of nature:

$\Pr TP = \mu f(S_e; e, X)$	$\Pr FN = \mu [1 - f(S_e; e, X)]$
$\Pr FP = (1-\mu)[1 - f(S_p; e, X)]$	$\Pr TN = (1-\mu) f(S_p; e, X)$

Note p_w the probability to realize a wrong diagnosis: $p_w = \Pr FP + \Pr FN$ and p_t the probability of the right diagnosis respectiveley such that:

$$\begin{aligned} p_w &= 1 - [1 + \theta(e, X)] [(1-\mu) S_p + \mu S_e] \\ p_t &= [1 + \theta(e, X)] [(1-\mu) S_p + \mu S_e] \end{aligned}$$

Note that p_w and p_t are obviously always positive and lower than 1. According to the definition of $\theta(.)$, we deduce that the effort e or supplementary tests X allows the GP to reduce the probability of a wrong diagnosis p_w as we get that the two derivatives are negative. Inversely, the effort e and the tests X allow the GP to increase p_t:

$$\begin{aligned} p'_{we} &= -\theta'_e(e, X) [(1-\mu) S_p + \mu S_e] = -p'_{te} \\ p'_{wX} &= -\theta'_X(e, X) [(1-\mu) S_p + \mu S_e] = -p'_{tX} \end{aligned}$$

Indeed, the effort and /or the supplementary tests allow the GP to come closer to μ when estimating the probability of occurrence of a serious illness.

medical process for the next one. And thus, during the second medical process, the tests should be avoided, as the GP is supposed to have the information.

[3] For simplicity, we assume that the respective impact of e and X on S_e and S_p is symmetric.

2.2.2 Information and treatment

Supplementary tests provide in the serious case extra information on the type of illness allowing the GP to reduce the cost of the treatment q_s such that q_s becomes $q_s\left[1-a\left(X\right)\right]$ where $a\left(.\right)$ represents the intrinsic value function of the information such that $a'\left(.\right)\geq 0$ and $a''\left(.\right)\leq 0$. However, in the benign case, there is no benefit of any supplementary tests on the cost of the treatment that remains $q_{b.}$ When the GP does not diagnose a severe disease early and thus prescribes q_b, the patient requires obviously a costly supplementary treatment $q_s' > q_s$. In the latter case, the patient may decide to switch the GP. This decision depends on the relationship between the patient and his GP, which is linked to the GP's level of effort. By definition, the effort is unobservable and not contractible and thus it cannot be used as a basis for any payment. However, it is important to consider because it is the mechanism by which the GP will be assessed by her patient and then rewarded or not for a high level of effort. We assume that in case of an error the probability for a patient to switch his GP is a decreasing function of the effort $\varepsilon\left(e\right)$ such that $\varepsilon'\left(e\right)\leq 0$ and $\varepsilon''\left(e\right)\geq 0$. Indeed it may happen that even if the GP failed to prescribe the suitable treatment, the patient decides to keep the same GP. According to the patient's choice, q_s' in case of False Negative (or q_b in case of False Positive) is fully delivered by an other provider (exit decision)[4]. Thus, if the loss for the GP is normalized to $\overline{K}$, the expected loss of a patient due to the exit decision is given by:

$$\begin{aligned}\left[E\overline{K}\right] &= p\varepsilon\left(e\right)\overline{K}\left\{1-\left[1+\theta\left(e,X\right)\right]\left[\left(1-\mu\right)S_p+\mu S_e\right]\right\}\\ \left[E\overline{K}\right] &= p\varepsilon\left(e\right)\overline{K}p_w\end{aligned}$$

All assumptions are summarized on Figure (1).

The GP acts as a price-taker: for instance the fee in a FFS is fixed by the government through the full public health insurance[5]. Thus the GP's market power comes from

[4] The patient would be referred to a specialist or admitted to a hospital.

[5] We assume that full public insurance is the optimal coverage as there is neither inequality accross patients nor moral hazard on the demand side: the patient is supposed to act as a perfect agent. Thus as shown by Rochet (1991), full insurance is the optimal solution and this is still the case as long as redistribution accross agents through risks and wealths are both going in the same way (negative correlation between inequalities parameters) see Cremer Pestieau (1996).

Figure 1:

her ability to determine both the quantity (the expected quantity) and the quality of the treatment provided. At this stage, the physician can modify the quality of the diagnosis by prescribing some supplementary tests that improve the quality of the diagnosis or by exerting an effort.

According to our notations and in order to simplify the calculations, we define $[E_n]$ the expected number of visits provided by the representative GP for an individual medical process. We assume that a wrong diagnosis may reduce the GP's long term medical activity (through a potential loss of a patient) but may also increase the number of visits: in case of an error, the patient comes for a second visit, but with probability $\varepsilon(e)$, he does not trust his GP anymore and thus decides to switch.

$$\begin{aligned} [E_n] &= 2p - p\left[1+\theta(e,X)\right]\left[(1-\mu)S_p + \mu S_e\right] \\ [E_n] &= p(2-p_t) \end{aligned}$$

We characterize the cost of the treatment in each state of nature and the expected cost of the curative treatment received by the patient (whatever his potential exit decision) is denoted $[E_q]$

$$\begin{aligned}[E_q] &= pq_b + pq_s\left[1-a\left(X\right)\right] + p\mu\left\{q_s' - q_s\left[1-a\left(X\right)\right]\right\} \\ &\quad -p\mu f\left(S_e; e, X\right)\left\{q_b + q_s' - q_s\left[1-a\left(X\right)\right]\right\} - pq_s\left[1-a\left(X\right)\right]\left(1-\mu\right) f\left(S_p; e, X\right)\end{aligned}$$

2.2.3 Medical activity and risk perception

Both the effort e and the quantity of supplementary tests X on the probability of the right diagnosis have a crucial impact on the GP's expected medical activity

$$\frac{\partial E_n}{\partial e} = -p\theta_e'\left(e, X\right)\left[\left(1-\mu\right)S_p + \mu S_e\right] \tag{1}$$

$$\frac{\partial E_n}{\partial X} = -p\theta_X'\left(e, X\right)\left[\left(1-\mu\right)S_p + \mu S_e\right] \tag{2}$$

$$\begin{aligned}\frac{\partial E\overline{K}}{\partial e} &= p\varepsilon'\left(e\right)\overline{K}\left\{1-\left[1+\theta\left(e, X\right)\right]\left[\left(1-\mu\right)S_p + \mu S_e\right]\right\} \\ &\quad -p\varepsilon\left(e\right)\overline{K}\theta_e'\left(e, X\right)\left[\left(1-\mu\right)S_p + \mu S_e\right]\end{aligned} \tag{3}$$

$$\frac{\partial E\overline{K}}{\partial X} = -p\varepsilon\left(e\right)\overline{K}\theta_X'\left(e, X\right)\left[\left(1-\mu\right)S_p + \mu S_e\right] \tag{4}$$

Both the effort e and the supplementary tests X help the GP in reducing the probability to perform a wrong diagnosis p_w. Both e and X allow to reduce the expected number of visits required $[E_n]$ but also the expected financial loss born by the GP due to the risk to lose a patient $\left[E\overline{K}\right]$. Through $[E_n]$ and $[E\overline{K}]$, the effort e has an indirect impact on the medical activity. The effort e increases the probability to perform a right diagnosis and thus a right diagnosis increases the probability to keep the patients, increasing indirectly the expected long term activity through $\frac{\partial E\overline{K}}{\partial e} < 0$; simultaneously it decreases the expected short term activity $\frac{\partial E_n}{\partial e} < 0$. Thus the effort e has two indirect effects on the medical activity: firstly the effort increases the expected long term activity by reducing p_w but also by reducing the probability to lose a patient in case of an error (the two effects act in the same way). Secondly, the effort reduces the expected short term activity by reducing the probability of a second visit.

$$\frac{\partial E_q}{\partial e} = -p\theta'_e(e,X)\left\{\mu S_e\left[q_b + q'_s - q_s(1-a(X))\right] + (1-\mu)S_p q_s[1-a(X)]\right\} \quad (5)$$

$$\frac{\partial E_q}{\partial X} = -p\theta'_X(e,X)\left\{\mu S_e\left[q_b + q'_s - q_s(1-a(X))\right] + (1-\mu)S_p q_s[1-a(X)]\right\}$$
$$+pa'(X)\,q_s\{\mu[1-f(S_e;e,X)] + (1-\mu)f(S_p;e,X) - 1\} \quad (6)$$

The sign of the two derivatives (5) and (6) is unambiguously negative meaning that the effort e and the supplementary tests X both allow the GP to decrease the cost of the treatment. This corresponds to the idea that a better diagnosis leads to a cheaper treatment. Note that the second term of the derivative (6) is also negative due to to the intrinsec value of the tests represented by $a'(.) < 0$. Indeed the term into brackets corresponds to the proportion of patients for whom the diagnosis process gives a"severe illness"result.

2.3 The public insurer

Through the public health insurance and the GP's reimbursement mechanism, the government is responsible for organizing the ambulatory care sector. In the context of a representative patient[6], there is no purpose for some income redistribution and we assume that the government deducts a lump sum transfer F to finance the health system. There is no private insurance market, and we could show that the optimal coverage in this context of a perfect patient is a full insurance coverage[7]. To organize the GP's payment scheme, the public insurer implements a mix of different instruments: K, a *per-capita* transfer to the GP distributed at the time of registration, δ the *fee-for-service* part to pay for visits and finally λ the part of πX born by the GP. In other words, if $K = 0$, and $\lambda = 0$ the payment is a complete fee-for-service payment; on the contrary, if $\lambda = 0$ and $\delta = 0$, the payment is a pure capitation mechanism. All intermediary combinations (K, δ, λ) are supposed to be available.

[6] By assuming that a GP is no able to recover by herself, she is considered as an individual like all other ones.

[7] Among others Rochet (1991) or Cremer Pestieau (1996) analyze the conditions required to get a full optimal public coverage: a perfect agent (no moral hazard and negative correlation between health status and wealth).

The social objective function is the expected utility of the patient[8]. Note that, at the end of any medical process (whatever the sequence of the timing), the patient recover

s fully and the final state of health is the same. All consequences of GP's decisions are represented in terms of costs. Thus the public insurer maximizes the expected utility function subject to the means constraint in the economy and verifies that the GP participation constraint holds. We first study the social optimum and then by trying to decentralize the optimal solution, we deduce the optimal mix of K, δ, λ in order to study the role of economic incentives

$$W = C + [E_n]\, c + [E_q] + pce + p\pi X + \left[E\overline{K}\right] \tag{7}$$

The right hand side term sums up all expenses: individual expenses in terms of consumption C, the expected cost of the visits, the expected cost of the treatment, the expected cost of the GP's effort, the expected cost of supplementary tests and finally the potential cost for the GP to lose a patient.

Maximizing the patient's expected welfare, the programme is then:

$$\begin{array}{ll} \underset{e,X}{Max} & u\,(C) \\ s/t & (7) \end{array} \Leftrightarrow \underset{e,X}{Max} \quad u\left[W - [E_n]\, c - [E_q] - pce - p\pi X - \left[E\overline{K}\right]\right] \tag{P1}$$

The optimal effort is given by the first order condition $\dfrac{\partial u\,(C)}{\partial e} = 0$:

$$e^{*}/ -\frac{\partial\,[E_n]}{\partial e}c - \frac{\partial\,[E_q]}{\partial e} - pc - \frac{\partial\left[E\overline{K}\right]}{\partial e} = 0 \tag{8}$$

$$\begin{aligned} & p\theta_e'\,(e,X)\left\{\mu S_e\left[c + q_b + q_s' - q_s\,(1 - a\,(X)) + \varepsilon\,(e)\,\overline{K}\right]\right. \\ & \left. + (1-\mu)\, S_p\left[c + q_s\,(1 - a\,(X)) + \varepsilon\,(e)\,\overline{K}\right]\right\} \\ = \; & p\left\{c + 1 - [1 + \theta\,(e,X)]\,[\mu S_e + (1-\mu)\, S_p]\right\} \end{aligned}$$

and for the optimal level of tests $\dfrac{\partial u\,(C)}{\partial X} = 0$:

$$X^{*}/ -\frac{\partial\,[E_n]}{\partial X}c - \frac{\partial\,[E_q]}{\partial X} - p\pi - \frac{\partial\left[E\overline{K}\right]}{\partial X} = 0 \tag{9}$$

[8] On the patient's point of view, there is no cost when the patient switches his physician. As long as there is an other physician to prescribe the complementary treatment q_s' in case of a serious illness.

$$\begin{aligned} & p\theta'_X(e,X)\left\{\mu S_e\left[c+q_b+q'_s-q_s(1-a(X))+\varepsilon(e)\overline{K}\right]\right. \\ & \left.+(1-\mu)S_p\left[c+q_s(1-a(X))+\varepsilon(e)\overline{K}\right]\right\} \\ = \;& p\left\{\pi+a'(X)q_s\left[\mu\left[1-f(S_e;e,X)\right]+(1-\mu)f(S_p;e,X)-1\right]\right\} \end{aligned}$$

The two first order conditions (8) and (9) represent the trade-off between marginal cost and marginal benefit for respectively e and X. For the two endogenous parameters of the diagnosis process, the marginal cost is only represented by the direct cost of the e and X (respectively pc and $p\pi$). As well as the cost of e and X, all other terms that sum up the different benefits, are weighted by the marginal utility of consumption. This is because the social health insurance is financed by individuals through taxes constraining the individual available consumption. Thus, all terms (excepted the direct cost) represent the respective benefits of e or of X on the expected number of visits, on the expected cost of the treatment and on the expected cost of the loss of a patient. Note that the benefit of X on the expected cost of the treatment also depends on the information productivity of X that increases the benefit.

3 Physician's trade-off

In our setting, the GP first invests c that yields the initial diagnostic characterized by S_e and S_p. The diagnosis process may provide more information by running some supplementary tests X or by exerting an individual effort e. According to the result of the diagnosis, the GP prescribes the treatment.

Thus, by maximizing her profit, the GP solves the following programme:

$$\begin{array}{ll} \underset{X,e}{Max} E[\Pi] = & K+(\delta-c)[E_n]-pce-\lambda p\pi X-\left[E\overline{K}\right] \\ S/t & E[\Pi]\geq 0 \end{array} \tag{P2}$$

The optimal level of effort is such that $\dfrac{\partial E[\Pi]}{\partial e}=0$:

$$e^{GP}/\,(\delta-c)\frac{\partial[E_n]}{\partial e}-pc-\frac{\partial\left[E\overline{K}\right]}{\partial e}=0 \tag{10}$$

$$-\left(\delta-c\right)p\theta_e'\left(e,X\right)\left[\mu S_e+\left(1-\mu\right)S_p\right]-pc$$
$$-p\varepsilon'\left(e\right)\overline{K}\left[1-\left[1+\theta\left(e,X\right)\right]\left[\left(1-\mu\right)S_p+\mu S_e\right]\right]$$
$$=\quad -p\varepsilon\left(e\right)\overline{K}\theta_e'\left(e,X\right)\left[\mu S_e+\left(1-\mu\right)S_p\right]$$

According to the GP, the optimal level of supplementary tests is such that $\frac{\partial E\left[\Pi\right]}{\partial X}=0$:

$$X^{GP}/\left(\delta-c\right)\frac{\partial\left[E_n\right]}{\partial X}-\lambda p\pi-\frac{\partial\left[E\overline{K}\right]}{\partial X}=0 \tag{11}$$

$$-\left(\delta-c\right)p\theta_X'\left(e,X\right)\left[\mu S_e+\left(1-\mu\right)S_p\right]-\lambda p\pi$$
$$=\quad -p\varepsilon\left(e\right)\overline{K}\theta_X'\left(e,X\right)\left[\mu S_e+\left(1-\mu\right)S_p\right]$$

4 Optimal payment mechanism

To induce the GP to determine her level of effort and complementary tests according to the social optimum, first order conditions must respectively coincide (8) with (10) and (9) with (11). Combining (8) and (10), we deduce the optimal fee for service that ensure that $e^*=e^{GP}$ that is δ^* such that

$$\delta^*\frac{\partial\left[E_n\right]}{\partial e}=-\frac{\partial\left[E_q\right]}{\partial e} \tag{12}$$
$$\delta^*=-\frac{\mu S_e\left[q_b+q_s'-q_s\left(1-a\left(X\right)\right)\right]+\left(1-\mu\right)S_pq_s\left[1-a\left(X\right)\right]}{\left[\left(1-\mu\right)S_p+\mu S_e\right]}$$

The optimal fee for service is negative! Running some effort reduces the expected error and thus the expected cost of the treatment. For the GP, the effort reduces the probability of a second visit. Thus the negative fee for service reduces significantly the opportunity for the GP to provide a second visit, she is indirectly but highly incited to reduce the probability to realize a wrong diagnosis.

A negative fee for service to incite the GP to exert an effort may appear strange, it comes from the fact that the wrong diagnosis may be profitable for the GP when the fee for service is positive (and when $\varepsilon(e)$ is low).

Assuming that the public insurer implements the optimal fee for service δ^* $(e^*=e^{GP})$, to induce the GP to opt for the optimal quantity of complementary tests $X^*=X^{GP}$, the two first order conditions (9) and (11) must also coincide and

$$\delta\frac{\partial\left[E_n\right]}{\partial X}+\frac{\partial\left[E_q\right]}{\partial X}=p\pi\left(\lambda-1\right) \tag{13}$$

$$\lambda^* = 1 - \frac{1}{\pi} a'(X)\, q_s \left\{ \mu f(S_e; e, X) + (1-\mu)\left[1 - f(S_p; e, X)\right] \right\}$$

The participation parameter to the cost of X is non null[9]. Through this expression, the GP is incited only to pick up the high productive tests ($a'(.) > 0$). Note that this result allows the public insurer to promote the use of very high productive tests by providing a subsidy to the GP for the prescription when $\lambda^* < 0$; this could be the case for tests recommended in the context of a targeted preventive policy. This result illustrates the role of primary care in the implementation of a targeted preventive policy as the preventive tests is only efficient for a type of individuals, requiring the GP's information.

Lastly, to ensure the GP's participation to the health system, the capitation payment has to, at least, compensate the GP for her cost. By assuming that her reserve income is normalized to 0 corresponding to the GP's participation constraint of $(P2)$:

$$K^* = -(\delta^* - c)\left[E_n\right] + pce + \lambda^* p\pi X + \left[E\overline{K}\right]$$

5 Restrictive tools

Let us assume that the public insurer is not able for several reasons to implement a negative fee for service. The lower value available is such that $\widetilde{\delta} = 0$, corresponding to the situation where the two residual tools to reimburse the GP are $\widetilde{K}$ and $\widetilde{\lambda}$. The public insurer is still able to incite the GP to opt for the optimal level of tests $X^* = X^{GP}$. Using (13), we deduce that the quantity of supplementary tests is optimal as long as $X^* = X^{GP}$:

$$\begin{aligned} \widetilde{\lambda} &= 1 - \frac{1}{p\pi}\left[\frac{\partial E_q}{\partial X}\right] \\ \widetilde{\lambda} &= \lambda^* - \frac{p\theta'_X(e,X)}{p\pi}\left\{\mu S_e\left[q_b + q'_s - q_s(1 - a(X))\right] + (1-\mu) S_p q_s\left[1 - a(X)\right]\right\} \end{aligned}$$

The second term on the right side is negative ($\theta'_X(e,X) > 0$) such that $\widetilde{\lambda} < \lambda^*$. The overall mechanism is less incitative. Indeed, we show that according to the restriction and as $\widetilde{\delta} = 0$, the effort exerting by the GP is lower.

[9] The expression directly obtained is $\lambda^* = 1 - \frac{1}{\pi} a'(X)\, q_s \left\{1 - \mu\left[1 - f(S_e; e, X)\right] - (1-\mu) f(S_p; e, X)\right\}$; the term into brakets corresponds to 1 minus the probabilities of False Negative patients and of True Negative patients.

To understand how the optimal value of effort for the GP varies, we have to determine the sign of $\frac{\partial e^{GP}}{\partial \delta}$. First we derive the two first order conditions (10) and (11) by the fee for service parameter δ, and we get a system of two equations equations where the two unknown variables are $\frac{\partial e^{GP}}{\partial \delta}$ and $\frac{\partial X^{GP}}{\partial \delta}$. To determine the sign of the two derivatives we use Cramer's theorem. Note that we need two simplifications in order to conclude unambiguously.[10]

First, we specify $\theta(e, X)$ by a function that has all properties required according to our assumptions:

$$\theta(e, X) = \sqrt{e + X}$$

Thus $\theta(e, X) > 0$ as log as the GP overestimates the high risk; and $\theta(0, 0) = 0$. The derivatives are symmetric $\theta'_e(.) = \theta'_X(.) < 0$ and finally $\theta''_{ee}(.) = \theta''_{eX}(.) = \theta''_{XX}(.) > 0$.

Secondly, we assume that the probability for the GP to lose a patient is linear: $\varepsilon(e)$ is such that $\varepsilon'(.) < 0$ and $\varepsilon''(.) = 0$.

Using the two previous simplifications, we get that:

$$\frac{\partial e^{GP}}{\partial \delta} < 0 \quad \text{and} \quad \frac{\partial X^{GP}}{\partial \delta} > 0$$

Thus $\widetilde{e}^{GP} < e^{GP} = e^*$. When the public insurer does not implement a fee-for-service system ($\delta = 0$), the GP reduces his own level of effort. The probability to realize a wrong diagnosis is less reduced than in the optimal case, because $\widetilde{e}^{GP}$, the effort exerted is lower and the GP tends to prescribe a higher quantity of supplementary tests. But the public insurer is still able to induce the GP to prescribe the optimal level that is X^{GP}, but the participation to the cost has to be adjusted:

$$\widetilde{\lambda} = \lambda^* - \frac{p\theta'_X(e, X)}{p\pi} \left\{\mu S_e\left[q_b + q'_s - q_s(1 - a(X))\right] + (1 - \mu) S_p q_s [1 - a(X)]\right\}$$

This is why $\widetilde{\lambda} < \lambda^*$. Indeed, a lower value of λ represents a lower cost for the GP when she chooses to run some tests, and that allows her to compensate the reduction in e.

$$\begin{aligned} K^* &= -(\delta^* - c)[E_n] + pce + \lambda^* p\pi X + \left[E\overline{K}\right] \\ \widetilde{K} &= -\left(\widetilde{\delta} - c\right)[E_n] + pce + \widetilde{\lambda} p\pi X + \left[E\overline{K}\right] \end{aligned}$$

$$K^* - \widetilde{K} > 0$$

[10] See computations in annex 1.

6 Conclusion

We have attempted to design a model that takes into account the specificities of general practice, i.e. the predominance of benign or moderate problems that may not require extensive testing to be diagnosed and treated, but the physician still has to deal with a low probability of high risks for each case. She can manage this uncertainty both by prescribing tests to acquire information and to eliminate the risk of an error, and/or through her effort to conduct a thorough examination or exert a follow-up. In the French system, the GP does not bear the cost of testing, which could incite her to make a systematic trade-off in favor of testing versus increasing her effort. This may lead to inefficiencies from the insurer's perspective, thus we have explored wether cost-sharing mechanisms could lead to a better allocation of resources. On the patient's side, we have introduced the idea that extra testing and induced demand may create some disutility for the patient, for example associated to time losses or to the inconveniencies of some tests. Altogether, we have strived to reach a more realistic model representing the interaction between the patient and the GP, and to analyze the consequences for the insurer. Our main findings are as following.

Pure payment mechanisms prevent from any decentralization of the social optimum. In our first setting, the residual probability to miss a serious risk which is a characteristic of the GP, may only be reduced by running some supplementary tests. In that context, a per-capita payment, that prevents from any induced demand, should be completed by a cost-sharing or subsidy parameter to induce the optimal quality of the diagnosis process. Depending on the weight of private incentives born by the GP, and on the positive predictive value of the test related to its cost, the public insurer will have to favour or to contain the prescription of supplementary tests.

When the GP may also exert an effort which has several consequences illustrating several dimensions of the quality of the service, the analysis is more complicated. Mind that the GP's effort reduces the probability to miss a serious risk, improves directly the benefit of the treatment and thirdly reinforces customer loyalty. Even if the public insurer will always be able to induce the optimal quantity of supplementary tests, the first best solution is no more reachable. There is a trade-off between the two main objectives, high quality and cost containment. When the priority is to improve quality, the public insurer has to incite the

GP to exert the optimal level of effort. To induce the GP in exerting a higher effort than the one she would prefer without incentive, the public insurer has to determine a fee-for-service parameter that offers a rent to her for any service produced. Obviously this mechanism does not prevent from induced demand anymore as the GP is incited to produce as much she can. These incentives are strong enough to incite her in running all supplementary test required and even more (in order to increase her chance to detect a serious illness). Thus the optimal cost-sharing parameter takes this into account such that the GP bears the full cost of test. Depending on all parameters the lump-sum transfer may be either positive or negative. When it is positive it could be interpreted as a transfer from the public insurer to the GP to reimburse the fixed investments and the cost of the effort and the prospective cost of supplementary tests required. When it is negative, it could be a payment from the GP to the public insurer as a participation fee as she will get a high rent from all services provided.

On the opposite when the priority is to contain costs, the public insurer acts as there were no moral hazard, as if the probability to miss a serious risk were only dependant on the supplementary tests level. The payment mechanism implemented is made up by both a per-capita lump sum transfer and a cost-sharing (or a subsidy) parameter associated to the cost of supplementary tests. Indeed depending on the relative weight of private incentives born by the GP, the public insurer will have to favour or to refrain the recourse of an expensive diagnosis process.

Part of our results come from the way we introduce the induced demand which is like a discrete variable 0 or 1 of an ex-post moral hazard. This implies some extreme results and this is even more true as the patient is assumed to be a perfect agent who prefers only the minimal treatment required. The simplification allows us to discuss the relative impacts of payment mechanisms in terms of incentives, otherwise the results would be even more ambiguous.

Another bias of our model comes from the fact that if both the effort and the test increase the sensitivity and the specificity of the diagnostic process of the GP, the test also brings an additionnal informational benefit and reduces the cost of treating severe cases, by a better adjustment. It may be that this makes too strong a case for combining effort and testing. Moreover, the function we have assumed a equal marginal benefit of e and X on sensitivity and on specificity. In real life, it is plausible that the effort made by the GP has a high impact on specificity, while testing would increase sensitivity. If it is the case that μ, the probability of a severe disease, is very small, than avoiding to treat uselessly too many false positive patients may be less costly to the payer than missing a small number of severe patients. The fact that patients' health is restored equally wether severe or benign probably also induces a bias in favor of testing, since false positive patients have no disutility when getting useless treatments.

The model can also be improved by introducing different attitudes of Doctors vis-à-vis risk. We are not alluding here strictly speaking to the conventionnal concept of risk aversion. We suggest that some Doctors may tend to overestimate the probability the probability of the occurrence of a severe disease because of its expected consequences. She may then systematically prescribe complementary testing because of lack of self-confidence or anxiety. This assumption is grounded on empirical and theoretical research work done first by Quiggin, and further developed mainly by Tvarsky and Hanheman on one hand, and Wakker on the other hand, leading to the formulation of the Rank Dependent Utility Theory and of Prospect Theory. Doctor may also underestimate the probability of a severe disease, either because of intrinsic self-confidence or through repeated past experience.

As a result of our setting, we notice that whatever the final patient's state of health, the optimal payment mechanism depends on the initial probability of a visit p and on the probability p_s which denotes the sum of all serious risks associated to the observed symptoms. Thus the optimal payment scheme depends on the couple of risks (p, p_s). This last remark may lead to an other important impact that may change the results. What would change if GPs had different attitudes towards high risk and if some of them could anticipate a higher probability of a serious illness (for instance, because of the individual practice experiment such as previous mistakes)? In a next step, we will take into account GPs' different attitudes towards risks. Even if we finally get some "combined" optimal mechanism, we could allow, as Van Barneveld and al. (2001), some risk sharing as a complement to a capitation system. In other words, we may consider a non linear fee-for-service payment depending on the type of illness, such that δ_s differs from δ_b, where each parameter may also be 0.

7 Annex 1

$\frac{\partial E[\Pi]}{\partial e} = 0$ thus $\frac{\partial^2 E[\Pi]}{\partial e \partial \delta} = 0$ and $\frac{\partial E[\Pi]}{\partial X} = 0$ thus $\frac{\partial^2 E[\Pi]}{\partial X \partial \delta} = 0$: this leads to the system of two equations:

$$\frac{\partial e}{\partial \delta} \left\{ p\theta''_{ee}(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] + 2p\theta'_e(.) \varepsilon'(e) \overline{K} - p\varepsilon''(e) \overline{K} \left[\frac{1}{[\mu S_e + (1-\mu) S_p]} - [1 + \theta(.)] \right] \right\}$$
$$+ \frac{\partial X}{\partial \delta} \left\{ p\theta''_{eX}(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] + p\theta'_X(.) \varepsilon'(e) \overline{K} \right\} = p\theta'_e(.)$$

$$\frac{\partial e}{\partial \delta} \left\{ p\theta''_{Xe}(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] + p\theta'_X(.) \varepsilon'(e) \overline{K} \right\}$$
$$+ \frac{\partial X}{\partial \delta} \left\{ p\theta''_{XX}(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] \right\} = p\theta'_X(.)$$

To solve the system, we use two simplifications: first, we specify $\theta(e, X)$, the adjustment parameter knowing that

$$\begin{aligned} f(S_e; e, X) &= S_e [1 + \theta(e, X)] \\ f(S_p; e, X) &= S_p [1 + \theta(e, X)] \end{aligned}$$

$\theta(0,0) = 0$, $\theta'_e(e, X) > 0$, $\theta''_{ee}(e, X) < 0$, $\theta'_X(e, X) > 0$, and $\theta''_{XX}(e, X) < 0$ such that

$$\theta(e, X) = \sqrt{e + X}$$

Second, we assume that the probability to lose a patient $\varepsilon(e)$ is a decreasing linear function such that $\varepsilon'(.) < 0$ and $\varepsilon''(.) = 0$. Under the two assumptions, the system of the two equations can be rewrite:

$$\frac{\partial e}{\partial \delta} \left\{ p\theta''(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] + 2p\theta'(.) \varepsilon'(e) \overline{K} \right\}$$
$$+ \frac{\partial X}{\partial \delta} \left\{ p\theta''(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] + p\theta'(.) \varepsilon'(e) \overline{K} \right\} = p\theta'(.)$$

$$\frac{\partial e}{\partial \delta} \left\{ p\theta''(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] + p\theta'(.) \varepsilon'(e) \overline{K} \right\}$$
$$+ \frac{\partial X}{\partial \delta} \left\{ p\theta''(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right] \right\} = p\theta'(.)$$

By substituting $p\theta''(.) \left[-\delta + c + \varepsilon(e) \overline{K} \right]$ by A, we get:

$$\frac{\partial e}{\partial \delta} \left\{ A + 2p\theta'(.) \varepsilon'(e) \overline{K} \right\} + \frac{\partial X}{\partial \delta} \left\{ A + p\theta'(.) \varepsilon'(e) \overline{K} \right\} = p\theta'(.)$$

$$\frac{\partial e}{\partial \delta} \left\{ A + p\theta'(.) \varepsilon'(e) \overline{K} \right\} + \frac{\partial X}{\partial \delta} \left\{ A \right\} = p\theta'(.)$$

By using the Envelope Theorem, we can deduce $\frac{\partial e}{\partial \delta}$ and $\frac{\partial X}{\partial \delta}$ such that:

$$\begin{aligned}
\frac{\partial e}{\partial \delta} &= \frac{p\theta'(.)\,A - \left[A + p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right] p\theta'(.)}{\left[A + 2p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right] A - \left[A + p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right]^2} \\
\frac{\partial e}{\partial \delta} &= \frac{-p\theta'(.)\,\varepsilon'(e)\,\overline{K}p\theta'(.)}{A^2 + 2p\theta'(.)\,\varepsilon'(e)\,\overline{K}A - A^2 - 2Ap\theta'(.)\,\varepsilon'(e)\,\overline{K} - \left[p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right]^2} \\
\frac{\partial e}{\partial \delta} &= \frac{1}{\varepsilon'(e)\,\overline{K}} < 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial X}{\partial \delta} &= \frac{\left[A + 2p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right] p\theta'(.) - p\theta'(.)\left[A + p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right]}{\left[A + 2p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right] A - \left[A + p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right]^2} \\
\frac{\partial X}{\partial \delta} &= \frac{p\theta'(.)\,\varepsilon'(e)\,\overline{K}p\theta'(.)}{A^2 + 2p\theta'(.)\,\varepsilon'(e)\,\overline{K}A - A^2 - 2Ap\theta'(.)\,\varepsilon'(e)\,\overline{K} - \left[p\theta'(.)\,\varepsilon'(e)\,\overline{K}\right]^2} \\
\frac{\partial X}{\partial \delta} &= \frac{1}{-\varepsilon'(e)\,\overline{K}} > 0
\end{aligned}$$