



Physicians self selection of a payment mechanism: Capitation versus fee-for-service

Marie Allard, Izabela Jelovac, Pierre-Thomas Léger

► To cite this version:

Marie Allard, Izabela Jelovac, Pierre-Thomas Léger. Physicians self selection of a payment mechanism: Capitation versus fee-for-service. 2010. halshs-00523370

HAL Id: halshs-00523370

<https://shs.hal.science/halshs-00523370>

Submitted on 5 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

W P 1024

**Physicians self selection of a payment mechanism:
Capitation versus fee-for-service**

Marie Allard, Izabela Jelovac, Pierre-Thomas Léger

Octobre 2010

GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne

93, chemin des Mouilles 69130 Ecully – France

Tel. +33 (0)4 72 86 60 60

Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France

Tel. +33 (0)4 77 42 19 60

Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : gate@gate.cnrs.fr

Téléchargement / Download : <http://www.gate.cnrs.fr> – Publications / Working Papers

Physicians self selection of a payment mechanism: Capitation versus fee-for-service^{*}

Marie Allard^a

Izabela Jelovac^b

Pierre-Thomas Léger^c

Octobre 2010

Abstract

The main question raised in this paper is whether GPs should self select their payment mechanism or not. To answer it, we model GPs' behavior under the most common payment schemes (capitation and fee-for-service) and when GPs can select one among those. Our analysis considers GPs heterogeneity in terms of both ability and sense of professional duty. We conclude that when savings on specialists costs are the main concern of a regulator, GPs should be paid on a fee-for-service basis. Instead, when failures to identify severe conditions are the main concern, then payment self selection by GPs can be optimal.

Key words: GPs, gatekeeping, payment scheme, self selection, ability, professional duty.

JEL codes: I18, J33

^{*} Léger thanks HEC Montréal, FQRSC and SSHRC for funding. Jelovac thanks Banque Nationale Belge for funding. We thank Ken Eggleston, Burkhard Hehenkamp, Sverre Grepperud and seminar and conference participants at the University of Lausanne, University of Lund, University of Liège, GATE (Lyon) and European Health Economics Worskhop (Bergen) for helpful comments and suggestions. Previously titled: Physician Payment Mechanisms: Dynamics, Diagnostic Ability and Altruism

^a HEC Montréal

^b CNRS-GATE, Lyon

^c HEC Montréal, CIRANO and CIRPEE

1. Introduction

The main question raised in this paper is whether GPs should be allowed to self select their payment mechanism or not. We model GPs' behavior under the most common payment schemes (capitation and fee-for-service) and when GPs can select one among those. Our analysis recognizes that GPs, as other professionals, are heterogeneous in terms of both ability and sense of professional duty. This diversity of profiles between GPs allows us to compare in an innovative fashion the most common payment schemes with a situation in which GPs can voluntarily select between capitation and fee-for-service (FFS henceforth).

The focus on GPs' selection of a payment scheme is motivated by a new policy in France giving the possibility to some physicians to choose between FFS and some incentive payment. Selection between FFS and capitation also exists in the US where physicians can either work in an HMO, often characterized by capitation payments, or in a more traditional setting under FFS. In Quebec, a 1999 reform also allowed physicians to select the contract under which they preferred to work among FFS and a mixed-remuneration contract giving physicians a base wage and a reduced FFS. From the wide literature on adverse selection, we can easily anticipate that the physicians opting for one type of organization/payment scheme need not have the same profile as the physicians opting for another one.

Many authors have written about the incentive properties of the most traditional payment schemes (see McGuire, 2000, for a survey on physician agency). On the one hand, FFS might encourage physicians to provide too many medical services to maximize their revenue. On the other hand, capitated physicians become residual claimant, which may lead to a limitation of either the amount or the quality of the medical services they provide. To sum up, FFS can be responsible for inefficiently high costs while capitation can be responsible for inefficiently low quality/amount of care.

These incentive properties are likely to hold whenever physicians are not allowed to refer patients to other levels of medical care. However, as shown in a companion paper by Allard *et al* (2009) that compares capitation, FFS and fundholding, the above-mentioned incentive properties can be reversed in the case of GPs able to refer patients to specialty care. Indeed, capitated GPs can save on personal costs by simply referring their patients to expensive specialized care. GPs paid on a FFS basis earn more revenue when treating the patients on their own rather than referring them to a specialist. In that case, FFS would lead to lower total costs and quality compared to capitation, if we suppose that the costs and the quality of specialty care are higher than those of GPs care and we abstract from professional duty considerations.

To our knowledge, only Fortin *et al* (2008) and Dumont *et al* (2008) address the issue of physicians voluntary self selection of a payment mechanism. With regard to the 1999 reform in the province of Quebec, they provide, respectively, a theoretical and an empirical analysis of the consequences of physicians selection into one or another payment scheme. In their analyses, they do not explicitly recognize physicians heterogeneity so that self selection into one payment scheme cannot depend on physicians characteristics. However, empirical literature on personnel economics has shown that agents self selection into a payment mechanism can depend on various personal characteristics such as skill levels, gender, risk aversion or intrinsic motivation (see Teyssier, 2008, and the references therein).

Among the latter, we focus on two sorts of GPs heterogeneities: (1) Skill levels, illustrated in our analysis by the ability to establish more or less accurate diagnoses; (2) Intrinsic motivation, illustrated by GPs concern for their patients health (also called altruism or sense of professional duty). Our model also recognizes that GPs' ability and sense of professional duty are their own private information. So is the information acquired during GPs' diagnosis.

The papers by Ma (2007), by Choné and Ma (2010) and by Allard *et al* (2009) are the only ones so far to recognize that physicians can be heterogeneous in their degree of

concern for their patients' health, and that this concern is physicians' private information. Other papers have analyzed GPs contracting while focusing on their gatekeeping role (Garcia-Mariñoso and Jelovac, 2003; Malcomson, 2004; Levaggi and Rochaix, 2007; Brekke *et al*, 2008; Gonzalez, 2010).

In Garcia-Mariñoso and Jelovac (2003) and in Malcomson (2004), the information acquired by GPs during diagnosis is endogenous as it depends on their costly effort, which itself is influenced by the payment contract. On the contrary, diagnosis accuracy in the present paper is exogenous as it depends on GPs' ability. It is the choice of a payment contract that is endogenous here, and this choice can be influenced by the exogenous quality of the diagnosis. We do not claim that the quality of a GP's diagnosis is exogenous rather than endogenous. We believe that the actual precision of a medical diagnosis relates to both GPs' endogenous effort and exogenous ability. We rather aim at providing a set of results assuming an exogenous quality of diagnosis, as a complement to the existing results in the afore-mentioned papers. This collection of results will provide a basis to empirically sort the question on the endogenous or exogenous nature of the diagnosis quality: "Is it the incentive contract that influences the quality of a diagnosis? Or is it the quality of a GP to perform a diagnosis that influences his selection of a payment scheme?"

To understand our results, it is important to establish the two concerns that are prevalent in the paper. The first concern is about the costs of specialty care, supposedly higher than GPs treatment costs. Therefore, one should avoid referrals of patients with a minor condition who do not need specialty care. Such wasteful referrals will be referred to as type-I errors. The second concern is about the imperfectness of GPs' diagnosis. One should avoid situations where GPs fail to refer patients with a severe condition who actually need specialty care. Such failures to adequately refer patients to specialty care will be considered as type-II errors.

We show that when given the choice, the GPs with a high sense of professional duty and a low ability choose a capitation scheme. These GPs exert a defensive type of medicine,

i.e. systematic referrals to specialists. The remaining GPs rather opt for FFS, but not all for the same reasons. Among those GPs, only those with a high ability are likely to recommend the most adequate course of action for their patients. Instead, the GPs with low professional duty tend to systematically treat their patients on their own before referring to a specialist when needed. We provide a graphical analysis to show that when savings on specialists costs (type-I errors) are the main concern of a regulator, GPs should be paid on a FFS basis. Instead, when failures to identify severe conditions (type-II errors) are the main concern, then payment self selection by GPs can be optimal.

The remaining of the paper is organized as follows. The model is described in Section 2. Section 3 reports some results already obtained in Allard *et al* (2009) about the GPs behavior under capitation and under FFS. Section 4 allows for GPs self selection between capitation and FFS. Section 5 provides a graphical comparison to sort whether a regulator should let GPs self select their payment scheme or not. Section 6 concludes.

2. The model

Without loss of generality, we consider here a simplified version of the model used in Allard *et al* (2009). A patient is assumed to suffer from either a minor condition $\underline{m}$ with probability P , or a severe condition $\bar{m}$ with probability $1-P$. The treatments that adequately cure these conditions are $\underline{t}$ and $\bar{t}$, respectively. Both the GP and the specialist can provide the treatment $\underline{t}$. Only a specialist can provide the treatment $\bar{t}$. The resulting health status, denoted h , depends on the initial condition and on the treatment provided. We reasonably assume that:

$$h_1 = h(\underline{m}, \underline{t}) \geq h_2 = h(\bar{m}, \bar{t}) > h_3 = h(\bar{m}, \underline{t}).$$

We assume that the cost of a treatment does not depend on the treatment itself, but on who provides the treatment: It is c_{GP} (respectively, c_{SP}) when it is provided by a GP (respectively, by a specialist). We reasonably assume that $c_{GP} < c_{SP}$.

The timing of the game is the following. Once the patient suffers from a condition that is unknown to him, he seeks advice and eventually care from a gatekeeping GP.¹ Once the GP has identified the condition with more or less accuracy, he decides whether to provide a treatment on his own (the treatment $\underline{t}$ *a fortiori*) or to refer the patient to a specialist. If the GP has treated his patient adequately with $\underline{t}$ or in case of referral to a specialist, the game ends. However, if the GP has provided $\underline{t}$ to a patient with a severe condition, the patient is ultimately referred to a specialist to receive treatment $\bar{t}$ as well. This is the situation in which the resulting health status is the worst one and it is denoted $h_3 = h(\bar{m}, \underline{t}, \bar{t})$. The costs in this case are the sum of GP's and specialist costs: $c_{GP} + c_{SP}$.

We assume that the accuracy of a GP's diagnosis is directly proportional to his ability. When performing a diagnosis, the GP obtains a signal corresponding to either type of illness: $\underline{s}$ or $\bar{s}$. We assume that the probability that the signal corresponds to the actual illness equals a measure of this ability, denoted a :²

$$\Pr(\underline{s}|\underline{m}) = \Pr(\bar{s}|\bar{m}) = a \in (1/2, 1).$$

Ability a varies across GPs. Bayesian updating allows us to write the probability of a condition for each received signal as a function of a GP's ability:

$$\Pr(\underline{m}|\underline{s}) = 1 - \Pr(\bar{m}|\underline{s}) = \frac{Pa}{Pa + (1-P)(1-a)};$$

$$\Pr(\bar{m}|\bar{s}) = 1 - \Pr(\underline{m}|\bar{s}) = \frac{(1-P)a}{(1-P)a + P(1-a)}.$$

As can be expected, the probability of a correct diagnosis by the GP (either $\Pr(\underline{m}|\underline{s})$ or $\Pr(\bar{m}|\bar{s})$) is increasing in GP's ability a , while the probability of a wrong diagnosis

¹ In Section 4, we go one step backwards and let the GP select his payment scheme in the very first stage.

² By doing so, we rule out diagnoses that are straightforward for all GPs. This is done without loss of generality. We could instead introduce some illnesses having a straightforward diagnosis and occurring with some probability. It would certainly make the analysis more realistic, although at the cost of a more cumbersome notation and without adding anything to the core of the analysis.

(either $\Pr(\bar{m}|\bar{s})$ or $\Pr(\underline{m}|\bar{s})$) is decreasing in ability. We also assume that a specialized diagnosis is always made without error.

The GP's utility depends on both his revenue and his patients' health. His revenue depends on whether he is paid on a capitation or on a FFS basis. If paid by capitation, he receives a payment K no matter whether he provides a treatment or not. Therefore, if he provides a treatment, he earns $K - c_{GP}$. Otherwise, he earns K . If paid on a FFS basis, he receives a fee R only when he provides a treatment. Therefore, he earns $R - c_{GP}$ if he provides a treatment, nothing otherwise.³ To reflect GP's sense of professional duty, we assume that a GP's concern for his patients' health is weighted by a parameter β , that varies across GPs. Therefore, a GP's utility can be written as:

$$U = \text{Net payment} + \beta \cdot h_i,$$

where $\text{Net payment} \in \{K; K - c_{GP}; R - c_{GP}; 0\}$ and $i = 1, 2, 3$.

3. GPs behavior under capitation and fee-for-service

This section heavily draws on the companion paper by Allard *et al* (2009).

We first confirm the result in Allard *et al* (2009) that GPs paid on a capitation basis systematically refer their patients to a specialist, whatever the outcome of the diagnosis. Indeed, the GP faces no trade off in this case: When referring a patient to a specialist, he earns a net payment K and the resulting patient's health is either h_1 or h_2 , depending on her initial condition. When treating the patient on his own with $\underline{t}$, he earns a net payment $K - c_{GP}$ and the resulting patient's health is either h_1 or h_3 , depending on her initial condition. Therefore, referring a patient is the dominant strategy for a GP paid on a capitation basis, whatever the diagnostic signal, the level of GP's ability a and professional duty β . In other words, a GP paid with capitation exerts a defensive type of

³ We omit here a payment for performing the diagnosis. This payment is normalized to zero without loss of generality.

medicine by referring patients to more capacitated specialists, and by doing so they maximize not only their patients health but also their net income. Their resulting expected utility is:

$$EU(Capitation, s.SP) = K + \beta \{P h_1 + (1-P) h_2\}.^4$$

On the other hand, the behavior of a GP paid on a FFS basis does depend on his ability and sense of professional duty, as well as on the outcome of the diagnosis. To understand that, it is useful to write in Table 1 the GP's expected utility for each possible signal and each possible behavior (treat and refer) under FFS:

Table 1. GP's expected utility under FFS, according to signal and behavior

		BEHAVIOR	
		<i>treat</i>	<i>refer</i>
SIGNAL	$\underline{s}$	$R - c_{GP} + \beta \{ \Pr(\underline{m}, \underline{s}) h_1 + (1 - \Pr(\underline{m}, \underline{s})) h_3 \}$	$\beta \{ \Pr(\underline{m}, \underline{s}) h_1 + (1 - \Pr(\underline{m}, \underline{s})) h_2 \}$
	$\bar{s}$	$R - c_{GP} + \beta \{ (1 - \Pr(\bar{m}, \bar{s})) h_1 + \Pr(\bar{m}, \bar{s}) h_3 \}$	$\beta \{ (1 - \Pr(\bar{m}, \bar{s})) h_1 + \Pr(\bar{m}, \bar{s}) h_2 \}$

Therefore, when the signal is $\underline{s}$, the GP's best reply is to treat (respectively, to refer) the patient on his own if and only if

$$\beta < \frac{R - c_{GP}}{\Pr(\bar{m}, \underline{s})(h_2 - h_3)} \quad (\text{respectively, } \beta > \frac{R - c_{GP}}{\Pr(\bar{m}, \underline{s})(h_2 - h_3)}).$$

Similarly, when the signal is $\bar{s}$, the GP's best reply is to treat (respectively, to refer) the patient on his own if and only if

$$\beta < \frac{R - c_{GP}}{\Pr(\bar{m}, \bar{s})(h_2 - h_3)} \quad (\text{respectively, } \beta > \frac{R - c_{GP}}{\Pr(\bar{m}, \bar{s})(h_2 - h_3)}).$$

⁴ The notation "s.SP" refers to the strategy of systematically referring patients to a specialist.

Combining the above comparisons and reasoning in expected terms (i.e. for each possible signal), a GP can adopt any of the following strategies, depending on his ability and sense of professional duty:

- A “follow signal” strategy, denoted “FS”, according to which the GP provides t if the signal $\underline{s}$ is observed, and refers the patient to a specialist if the signal $\bar{s}$ is observed. His expected utility under such a strategy is:

$$EU(FFS, FS) = \{Pa + (1-P)(1-a)\} (R - c_{GP}) + \beta \{P h_1 + (1-P) a h_2 + (1-P)(1-a) h_3\}.$$

- A “systematic GP’s treatment” strategy, denoted “s.GP”, according to which the GP provides treatment t , no matter whether the outcome of the diagnosis is the observation of a signal $\underline{s}$ or $\bar{s}$. His expected utility under such a strategy is:

$$EU(FFS, s.GP) = R - c_{GP} + \beta \{P h_1 + (1-P) h_3\}.$$

- A “systematic referral” strategy, denoted “s.SP”, according to which the GP refers patients to a specialist, whatever is the signal resulting from the diagnosis. His expected utility under such a strategy is:

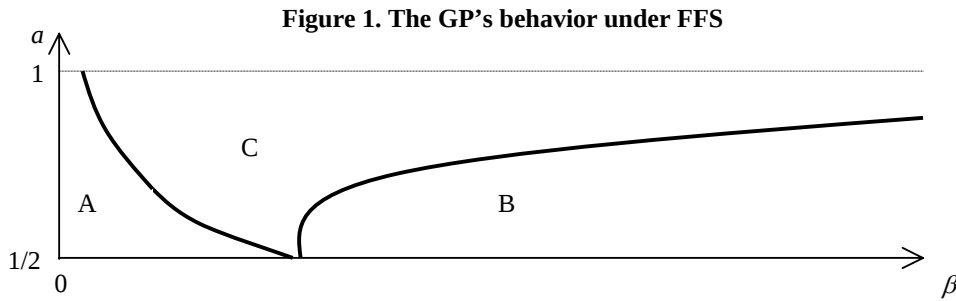
$$EU(FFS, s.SP) = \beta \{P h_1 + (1-P) h_2\}.$$

Comparing his expected utility under each available strategy, a GP’s best reply is to adopt either of these strategies under the following conditions:

- A “s.GP” strategy iff $\beta < \frac{R - c_{GP}}{\Pr(\bar{m}, \bar{s})(h_2 - h_3)}$;

- A “FS” strategy iff $\frac{R - c_{GP}}{\Pr(\bar{m}, \bar{s})(h_2 - h_3)} < \beta < \frac{R - c_{GP}}{\Pr(\bar{m}, \underline{s})(h_2 - h_3)}$;
- A “s.SP” strategy iff $\frac{R - c_{GP}}{\Pr(\bar{m}, \underline{s})(h_2 - h_3)} < \beta$.

We now interpret these conditions using Figure 1 where we plot them in the GPs profile space (a, β) . This allows us to discuss the importance of a GP’s profile in the decision to refer a patient to a specialist or not.



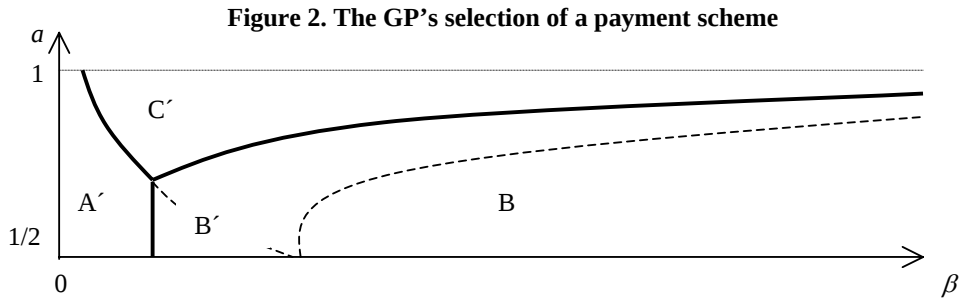
Under a FFS payment scheme, treating a patient on his own is more lucrative for a GP than referring her to a specialist. However, referring a patient to a specialist results in a better expected health for the patient, especially if the GP does not trust his own judgment because of a low ability a . Therefore, GPs characterized by a very poor sense of professional duty (area A in Figure 1) go for a “systematic GP’s treatment” strategy because they put a very high value on their own net income relatively to the health status of their patients. Moving towards higher degrees of professional duty, we find either GPs with a low ability who choose to systematically refer their patients (area B in Figure 1) or GPs with a higher ability who decide a course of action according to the diagnosis outcome (the “FS” strategy in area C of Figure 1). Indeed, the latter GPs are able enough so that trusting their own judgment satisfy their appeal for both income and patients’ health. The same reasoning holds for GPs with a higher sense of professional duty provided they are able enough. However, GPs with a higher sense of professional duty but a limited ability prefer to systematically refer their patients to a specialist. Indeed,

those GPs are very much concerned by their patients' health and they know that specialists are way more able than they are to adequately treat the patients.

4. GPs self selection of a payment mechanism

In a health system where both FFS payments and capitation payments coexist, the GPs can choose to work either under FFS (anticipating their behavior in terms of “s.GP”, “FS” and “s.SP” strategies) or under capitation (anticipating that they will always refer in that case). Whether a GP will choose FFS or capitation depends on his profile (α , β) and on the relative level of the capitation and FFS payments. In particular, comparing the expected GP's utilities, we have the following 3 cases:

- When the capitation payment is very high relatively to the FFS net payment, i.e. $R - c_{GP} < K$, all GPs trivially choose to work under capitation and therefore they all systematically refer their patients to specialists.
- Decreasing the capitation payment or increasing the FFS such that $P(R - c_{GP}) < K < R - c_{GP}$, only the GPs with a poor sense of professional duty ($\beta < (R - c_{GP} - K) / (1 - P)(h_2 - h_3)$) move to FFS and decide to systematically treat their patients on their own. The remaining GPs keep the capitation payment and they exert a defensive type of medicine consisting in systematically referring their patients to specialists.
- Decreasing the capitation payment or increasing the FFS even more until $K < P(R - c_{GP})$, we find the most interesting case depicted in Figure 2.



Area B in Figure 2 coincides with area B in Figure 1. It corresponds to the profiles of GPs making systematic referrals no matter whether they are paid under FFS or capitation. Therefore, when given the choice, they go for capitation to earn the payment K rather than earning nothing under FFS. Those GPs are characterized by relatively low levels of ability and high concern for their patients health.

Moving towards higher abilities and/or lower concern for patients health, we find profiles of GPs who would systematically refer patients under capitation but who would recommend a course of action according to their diagnosis if paid by FFS (area C in Figure 1, equivalent to areas B' and C' in Figure 2). When given the choice, only the most altruistic and least able among those GPs (area B' in Figure 2) would choose capitation rather than FFS to better conciliate their concern for patients' health and a certain income. The GPs in area C' with a better diagnosis ability and/or a weaker concern for patients' health choose a FFS payment and trust their judgment to recommend a course of action for their patients (the "follow signal" strategy). Indeed, FFS provide them with higher earnings while their "follow signal" strategy is relatively beneficial for their patients given their relatively high diagnosis ability.

The remaining GPs whose profiles are represented in area A' are characterized by a very weak interest for their patients health, and FFS is the payment that will secure the highest earnings to them provided they decide to systematically treat their patients on their own first.

5. Should GPs self select their payment mechanism?

To answer this question, we first present the first-best situation in which a regulator maximizes patients' expected health net of the treatment costs. This situation is equivalent to one in which GPs would never behave strategically and would be paid exactly their reservation utility, assuming the same zero-reservation utility for all. We consider that the regulator can tell GPs to follow one of the afore-mentioned strategies, i.e. either to recommend a course of action for the patient according to the diagnosis, or to systematically treat the patients first, or to systematically refer patients to a specialist. We provide in the next table the regulator's objective function corresponding to either of these strategies. They are derived in a way that is similar to the derivation of GP's expected utilities in Section 3.

Table 2. Patients' expected health net of treatment costs, according to GP's strategy

GP's strategy	Regulator's objective
"FS"	$Pa(h_1 - c_{GP}) + P(1 - a)(h_1 - c_{SP})$ $+ (1 - P)a(h_2 - c_{SP}) + (1 - P)(1 - a)(h_3 - c_{GP} - c_{SP})$
"s.GP"	$P(h_1 - c_{GP}) + (1 - P)(h_3 - c_{GP} - c_{SP})$
"s.SP"	$P(h_1 - c_{SP}) + (1 - P)(h_2 - c_{SP})$

Comparing the regulator's objective function under each GP's strategy, we obtain the following first-best rule:

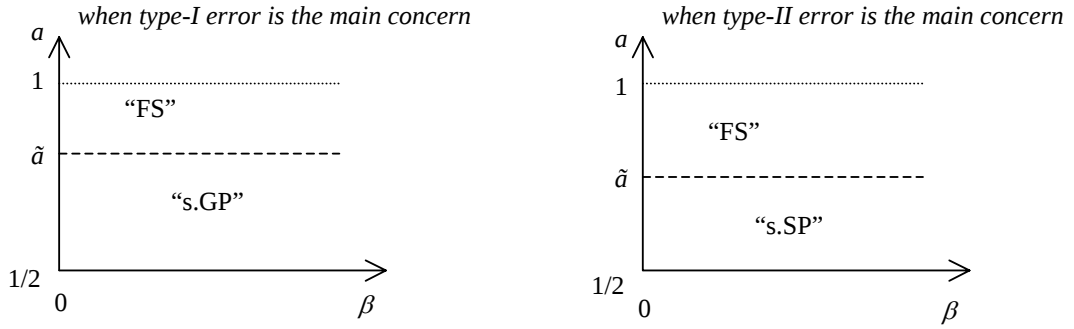
Table 3. First-best rule

Optimal strategy	
"FS"	$\Leftrightarrow a > \tilde{a} = \frac{\text{Max}\{P(c_{SP} - c_{GP}); (1 - P)(h_2 - h_3 + c_{GP})\}}{P(c_{SP} - c_{GP}) + (1 - P)(h_2 - h_3 + c_{GP})}$
"s.GP"	$\Leftrightarrow \frac{P(c_{SP} - c_{GP})}{P(c_{SP} - c_{GP}) + (1 - P)(h_2 - h_3 + c_{GP})} > \text{Max}\left\{a; \frac{(1 - P)(h_2 - h_3 + c_{GP})}{P(c_{SP} - c_{GP}) + (1 - P)(h_2 - h_3 + c_{GP})}\right\}$
"s.SP"	$\Leftrightarrow \frac{(1 - P)(h_2 - h_3 + c_{GP})}{P(c_{SP} - c_{GP}) + (1 - P)(h_2 - h_3 + c_{GP})} > \text{Max}\left\{a; \frac{P(c_{SP} - c_{GP})}{P(c_{SP} - c_{GP}) + (1 - P)(h_2 - h_3 + c_{GP})}\right\}$

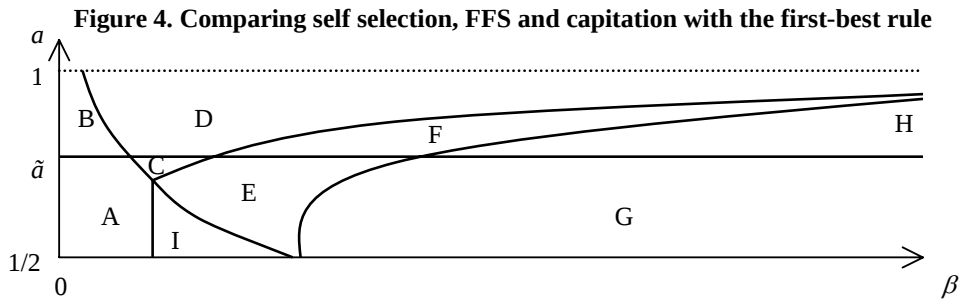
Intuitively, this first-best rule can be interpreted as follows. First, notice that the terms $P(c_{SP} - c_{GP})$ and $(1 - P)(h_2 - h_3 + c_{GP})$ appear repeatedly in the first-best rule. The term $P(c_{SP} - c_{GP})$ reflects the cost of a type-I error, i.e. the cost of referring a patient with a minor condition who does not need specialty care. In turn, the term $(1 - P)(h_2 - h_3 + c_{GP})$ reflects the loss associated to a type-II error, i.e. the loss caused by not referring a patient who actually needs specialty care because of a severe condition. The inequality $P(c_{SP} - c_{GP}) > (1 - P)(h_2 - h_3 + c_{GP})$ characterizes a situation where the costs of specialized care are so high that economizing on specialists costs is a stronger concern for the regulator than failures at referring a patient with a severe illness to a specialist. Instead, the opposite inequality $P(c_{SP} - c_{GP}) < (1 - P)(h_2 - h_3 + c_{GP})$ characterizes a situation where GPs failures at adequately referring patients is the strongest concern of a regulator. The afore-mentioned first-best rule tells us that when high specialists costs are the main concern, then low-ability GPs should systematically treat their patients before eventually referring them to a specialist. On the contrary, when GPs failures at referring a patient to a specialist when needed are the strongest concern, then low-ability GPs should systematically refer their patients to specialists. Last, high-ability GPs should always follow the result of their diagnosis (“FS” strategy) independently on the type of error that is relevant.

We next plot the first-best rule in the profile space (a, β) as in Figures 1 and 2, with the aim of comparing it with the different outcomes previously analyzed; That is, concerning capitated GPs, GPs paid on a FFS basis, and GPs selecting their payment mechanism among capitation and FFS.

Figure 3. The first-best rule



In Figure 4, we superpose Figures 1, 2 and 3 on each other to perform the overall comparison.



The outcome of the overall comparison depends on whether the main concern is either type-I or type-II error. We first focus on the situation where the important regulator's concern is economizing on high specialists costs, that is:

$$P(C_{SP} - C_{GP}) > (1 - P)(h_2 - h_3 + C_{GP}).$$

In this case, imposing a capitation payment for all GPs cannot be optimal since it would result in systematic referrals to specialists while the first-best rule rules out such strategy because of the high specialists costs. More precisely, for high-ability GPs ($a > \tilde{a}$, corresponding to areas B, D, F and H), the first-best rule recommends the “follow signal” strategy while the dominant strategy of capitated GPs is to systematically refer patients to a specialist. There are thus too many referrals compared to the first-best rule. The

situation is even worse for capitated low-ability GPs ($a < \tilde{a}$, that is, areas A, C, E, G and I) since the first-best strategy for low-ability GPs is to systematically treat their patients first. Therefore, there are way too many referrals in that case.

Furthermore, the graphical analysis reveals that FFS is closer to the first-best rule than self selection of payment schemes is. FFS and self selection only differ for GPs profiles in areas F, I and E. The outcome of FFS in areas F and I coincide with the first-best rule, while too many referrals are the outcome of self selection in those areas F and I. In area E, even if different from the first-best rule, FFS results in an outcome that is closer to the first-best than the outcome of self selection is. Indeed, in area E, self selection induces more referrals than FFS, which itself induces more referrals than in the first-best rule.

Therefore, we conclude that imposing a FFS payment scheme to all GPs is optimal when the important regulator's concern is economizing on high specialists costs. Letting GPs the option to choose capitation can only increase the number of referrals, which cannot be interesting when specialists costs are too high. Table 2 summarizes the arguments leading to this conclusion.

Table 2. Comparing self selection with FFS and capitation when type-I errors are the main concern

	First Best	Over refer	Greatly over refer	Under refer
Only FFS	A + D + F + I	C + E + H	G	B
Only capitation	–	B + D + F + H	A + C + E + G + I	–
Self selection	A + D	C + F + H	I + E + G	B

We next focus on the situation where the important regulator's concern is avoiding failures at referring a patient with a severe illness to the specialist, because the losses in terms of patients' health are too high:

$$P(C_{SP} - C_{GP}) < (1 - P)(h_2 - h_3 + C_{GP}).$$

No clear conclusion can be drawn for that case without making hazardous assumptions on the distribution of GPs in the profile space (a, β) . Indeed, for GPs with a profile corresponding to area A, C, E, G or I, the outcome of capitation coincides with the first-best rule. For GPs with a profile in area D, F or G, fee for service coincides with the first-best rule. Last, GPs with a profile as in area D, E, G or I, self selection of a payment contract is optimal. Table 3 summarizes this comparison.

Table 3. Comparing self selection with FFS and capitation when type-II errors are the main concern

	First Best	Over refer	Under refer	Greatly under refer
Only FFS	D + F + G	H	B + C + E	A + I
Only capitation	A + C + E + G + I	B + D + F + H	ϕ	ϕ
Self selection	D + E + G + I	F + H	B + C	A

6. Conclusion

The main contribution of this paper is to answer the question of whether GPs should be allowed to self select their payment mechanism or not. To deal with this issue, we recognize that GPs are heterogeneous in their ability and in their sense of professional duty. We limit our attention to the payment schemes that are the most commonly used, i.e. capitation and fee-for-service. We also limit our attention to gatekeeping GPs who have the choice between treating patients on their own or referring them to a specialist.

We first confirm some results already obtained in a companion paper by Allard *et al* (2009), in which GPs do not choose their payment mechanism. On the one hand, capitated GPs are shown to systematically refer patients to a specialist, whatever the result of the diagnosis and the GP's profile. On the other hand, the behavior of GPs paid on a FFS basis differ according to their profile and eventually according to the outcome of diagnoses. In particular, GPs with a high ability to perform accurate diagnoses and a relatively high concern for patients health recommend the most adequate course of action for their patients: They either provide a treatment on their own or they refer a patient to a specialist according to the outcome of the diagnosis. Instead, less able GPs with a high

concern for patients health systematically refer their patients to a specialist, anticipating a better outcome in terms of health. Last, GPs with little concern for patients health systematically treat the patients first.

When given the choice of a payment mechanism, only the most altruistic GPs with a moderate to low ability choose the capitation scheme and systematically refer patients to specialists. The remaining GPs opt for FFS, although not all for the same reason. Among those GPs, only the ones with higher ability and higher concern for health provide the most adequate health services, i.e. according to their relatively accurate diagnoses. The GPs characterized by the lowest concern for health choose FFS to maximize their revenue only by systematically treating their patients before eventually referring them if needed.

Whether a regulator should allow GPs to self select their payment mechanism or not depends on the priority he faces. If the main concern is to save on the costs of specialized care because those are too high, then it is optimal to pay all GPs on a FFS basis and not let them choose their payment scheme. If the main concern is to limit medical errors that can possibly be made by GPs, then the answer to our main question depends on the distribution of GPs' profiles. For example, it can be optimal to allow GPs to self select a payment mechanism if all GPs are sufficiently concerned about health and they are either very able or very unable at diagnosing patients.

References

Allard, M., I. Jelovac and P.T. Léger (2009). Treatment and referral decisions under different payment mechanisms. Mimeo.

Brekke, K.R., R. Nuscheler and O.R. Straume (2007). Gatekeeping in health care. *Journal of Health Economics* 26, 149–170.

Choné, P. and A. Ma (2010). Optimal health care contracts under physician agency. *Annales d'Economie et de Statistiques*, forthcoming.

Dumont, E., B. Fortin, N. Jacquemet and B. Shearer (2008). Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of Health Economics* 27 (6), 1436–1450.

Fortin, B., N. Jacquemet and B. Shearer (2008). Policy Analysis in the health-services market: accounting for quality and quantity. *Annales d'Economie et de Statistiques*, 91-92, 293-319.

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

Gonzalez, P. (2010). Gatekeeping versus direct access when patient information matters. *Health Economics*, forthcoming.

Levaggi, R. and L. Rochaix (2007). Exit, choice or loyalty: Patient driven competition in primary care. *Annals of Public and Cooperative Economics* 78 (4), 501–535.

Ma, A. (2007). Altruism and incentives in public and private health care, in Finance and incentives of the health care system, Proceedings of the 50th Anniversay Symposium of

the Yrjö Jahnsson Foundation, edited by Antti Suvanto and Hannu Vartiainen, Government Institute for Economic Research, Helsinki, 79–104.

Malcomson, J. (2004). Health Service Gatekeepers. *RAND Journal of Economics* 35 (2), 401–421.

McGuire, T.G. (2000). Physician agency. In: A.J. Culyer and J.P. Newhouse, Editors, *Handbook of Health Economics* vol. 1A, North-Holland, Amsterdam 461–536.

Teyssier, S. 2008. Experimental Evidence on Inequity Aversion and Self-Selection between Incentive Contracts. GATE working paper W.P. 08–21.