



## On a policy of transferring public patients to private practice

Paula González\*

*Universidad Pablo de Olavide and centra, Sevilla, Spain*

### Summary

We consider an economy where public hospitals are capacity-constrained, and we analyse the willingness of health authorities to reach agreements with private hospitals to have some of their patients treated there. When physicians are dual suppliers, we show that a problem of cream-skimming arises and reduces the incentives of the health authority to undertake such a policy. We argue that the more dispersed are the severities of the patients, the greater the reduction in the incentives will be. We also show that, despite the patient selection problem, when the policy is implemented it is often the case that health authorities decide a more intensive transfer of patients to private practice. Copyright © 2004 John Wiley & Sons, Ltd.

*JEL classification:* I11; I18; L51; H51

**Keywords** public–private health services; physician's incentives; patient selection; waiting-lists

### Introduction

Worldwide public health services are plagued by excess demand and lengthy waiting-lists. This unsatisfactory situation has persisted from the very inception of most public health systems and, far from improving, it seems to get more systematic over the years. In contrast, the private sector is usually characterized by relatively short or even zero waits and idle capacity. As a consequence of this, several national health authorities have turned to private hospitals and clinics for assistance in reducing their waiting-lists.

The purpose of this paper is to study the incentives of health authorities to divert patients from public sector waiting-lists to the private sector. In contrast with other analyses, we restrict attention to a particular regulatory regime in which health authorities fully finance some patients' access to the private sector. There are several examples of the application of such kind of

policies in the real world. At the Spanish regional level, for instance, the Valencian Region has been undertaking a policy of transferring patients to private hospitals over the last years. As a consequence, more than 100 000 social security patients were treated at private hospitals from July 1996 to June 2000, and 4.26 million Euros were spent in the year 2000 to defray the debt to private clinics that have participated in the 'Impact Plan' for reducing surgical waiting-lists. In Galicia 12% of the operations financed by the public sector are performed at private hospitals.<sup>a</sup>

The model in this article aims at pointing out the benefits and drawbacks that health authorities face when implementing this sort of policy. Finding the correct balance between cost-containment and improvements in the provision of health care services will be a major task for public decision-makers.

We study the effects that exploitation of public waiting-lists by third parties may have over

\*Correspondence to: Dpto. Economía y Empresa, Universidad Pablo de Olavide, Carretera de Utrera, km 1. 41013 Sevilla, Spain.  
E-mail: pgonrod@dee.upo.es

governments' costs when a policy of transferring patients to private clinics is undertaken. In particular, we concentrate on doctors' incentives to influence and manage waiting-lists for elective surgery to their own private benefit. This problem worsens if, as is common in countries with public health services and waiting-lists, doctors who work for government hospitals also engage in private practice. In the UK, for instance, most private medical services are provided by physicians whose main commitment is to their public sector duties.<sup>b</sup> Furthermore, there is usually a significant difference between the forms of payment to the doctors in the public and private sectors. In the public sector, it is common for physicians to receive a fixed salary, with limited financial incentives related to workload. In the private sector, on the contrary, the use of tailored incentives seems to be more widespread. In particular, it is common that physicians are partners or shareholders in private hospitals, with their remuneration being linked to the performance of the hospital (some form of profit-sharing).

These two features, doctors acting in both private and public sectors and different remuneration schemes in both sectors, create a basic problem. In a simple model in which the policy-maker reaches agreements with private hospitals to have some patients on the public waiting-lists treated there, we show that patient selection (cream-skimming) by the physicians occurs, i.e. physicians have incentives to strategically divert the easiest cases to their private practice.<sup>c</sup>

We show that the presence of cream-skimming reduces the incentives of the health authority to undertake such a policy in order to shorten public waiting-lists. As the physician only transfers the less severe cases to private practice, the average severity of those patients who remain in the public sector is increased, along with the costs that the health authority faces.

There are two main variables that determine the response of the health authority. First, the relative dispersion of the patients' severities. The higher this dispersion, the more the physician earns from selecting patients and, at the same time, the greater the impact on the costs borne by the health authority. Secondly, the patients' health losses associated with waiting for treatment. The larger these losses are, the more willing is the health authority to resort to private clinics, even in the presence of patient selection.

It is also shown that when the policy of transferring patients is implemented, the health authority's reaction to the problem of patient selection is, in many instances, a more intensive transfer of patients to private hospitals, i.e. fewer patients are kept on the public sector waiting-lists. The intuition behind this result is the following: As the patients who remain in the public sector are the most severe ones, the marginal cost of providing treatment in this sector increases. Hence, more patients are diverted to the private sector, despite this fostering the capacity of the physician to select patients.

Two main assumptions are made in this analysis. First, the public sector is capacity-constrained. Hence, there is a waiting time before treatment is performed. In the private sector, on the contrary, waiting times are assumed to be zero. This assumption is typical in the literature (see, for instance [3]). The private sector does indeed report almost zero waiting times in most western economies. For instance, Bosanquet [4] states that 'at present, there is under-occupation in private hospitals (in the UK), with occupancy rates at 50% or less'.

This paper is related to studies dealing with the interaction between public and private health sectors and waiting-lists. Iversen [5] and Olivella [3] have investigated the optimal decision of the health authority in terms of waiting times, focusing on the effects of the presence of a private alternative. In our model, however, waiting-lists are taken as given and we concentrate on the willingness of governments to turn to the private sector in order to alleviate such lists. We model the physician's behaviour, whereas Iversen and Olivella deal with patients' decisions as individuals choosing between waiting for free treatment in the public sector and paying for immediate treatment in the private sector. As such, Iversen and Olivella clearly rule out the possibility of cream-skimming on the part of the doctors.

In [Barros PP, Olivella P. Waiting lists and patient selection. *J Econ Manage Strat*, forthcoming] the doctor's strategic behaviour plays an important role. However, the different systems of remunerating the physicians in either sector (which is the crucial variable in our model) is not considered in their work. Patient-selection arises mainly from a combination of the rationing policy undertaken by the health authority, the criterion of the private physician regarding which severities he is willing to treat and the decision of the

patients to leave the queue in the public sector and resort to having to pay for private treatment. In our model, patients are fully insured irrespective of the sector where they are treated and no rationing policy is considered. This gives the full power to decide to the physician and, thus, always generates a situation of 'full cream-skimming'.<sup>d</sup>

There are some works where possible contributions from the private sector to reduce public waiting-lists are analysed. Cullis and Jones [6] and Hoel and Sæther [7] argue that a subsidy to private treatments may be an optimal policy instrument to reduce the public sector waiting time. In contrast with these analyses, we restrict attention to a particular regulatory regime in which health authorities purchase private services, by fully financing some patients' access to the private sector.

The rest of the paper is organized as follows: First the model is presented. Next, we compute the optimal policy in the benchmark scenario. Afterwards, we study the behaviour of the physician concerning the selection of patients and the response of the health authority in a general context. Then, we provide a complete characterization of the results for particular functional forms. Finally we conclude. All of the proofs are in the appendix.

## The Model

There are three players in this economy: a continuum of patients, a physician and the health authority. We model their relationship as a simple game in which the health authority maintains some patients waiting for treatment in the public sector, and reaches agreements with private hospitals to have the remaining patients treated there.

Let us now detail the decision variables, parameters, and utility functions that are relevant in the model.

There is a continuum of individuals in need of health care, all of whom require elective treatment. We consider that the size of this population of potential patients is exogenously given and it is normalized to 1. These patients are homogeneous, except for their degree of severity, which is measured by the random variable  $s$ . This variable is distributed according to a continuous and positive density function  $dF(s)$  defined on  $[\underline{s}, \bar{s}]$ .

Let  $\Delta s = \bar{s} - \underline{s}$  be the difference between the extremes of the domain for  $s$ , and  $\hat{s}$  the average severity of the population of patients.

We assume that the public sector is capacity constrained. We do not model this assumption as a maximum threshold on the treatments that can be provided by the public hospital. We consider, instead, that the public sector suffers from congestion and, hence, there is a waiting time before treatment is performed. Let  $x$  denote the number (or percentage) of patients treated in the public sector. We may interpret this variable in the model as a proxy for the length of the public waiting-list. The remaining  $1 - x$  patients will be treated in the private sector.<sup>e</sup> As the private sector is assumed to be operating under capacity, there are no waiting-lists for private sector treatment. Moreover, since we are dealing with public health systems, we assume that patients are fully insured irrespective of the sector where they are eventually treated.

A patient is assumed to obtain a benefit from treatment defined by  $Q$ . When the patient is treated in the public sector, this benefit is undermined due to the existence of waiting-lists, because the patient suffers the costs associated with waiting for treatment. We consider this loss to be higher the fewer patients that are diverted to the private sector (i.e. the higher the congestion in the public system). We define  $\delta > 0$  as the patient's marginal loss associated with waiting for treatment. The patient's benefit from being treated in the public sector is, therefore, given by  $Q^{pb} = Q - \delta x$ . In contrast, since in the private sector patients suffer no loss due to waiting,  $Q^{pv} = Q$ . Under this formulation,  $Q^{pb} < Q^{pv}$  and  $\lim_{x \rightarrow 0} Q^{pb} = Q^{pv}$ . Social expected health benefits derived from treatment are, hence, given by  $xQ^{pb} + (1 - x)Q^{pv} = Q - \delta x^2$ .

Let  $k^{pb}$  and  $k^{pv}$  denote the unit cost of treatment (disregarding physician's compensations) in the public and private sector respectively. We take the cost of the public treatment to be linear to simplify exposition. The qualitative results of the model, however, can be easily extended to other more general cost structures. With regard to the unit cost in the private sector, we consider that it is not too restrictive to assume that is linear if the private sector operates under capacity, as is assumed throughout all our analysis. Still, as is the case for the public unit cost, the analysis is robust to other cost specifications. For the sake of clarity, we define

$\Delta k = k^{pb} - k^{pv}$  as the difference between the unit costs of treatment in the two sectors.

To be consistent with real-life observations, we consider that the health authority makes a constant payment for each operation performed in the private sector ( $w$ ). Although we take the value of  $w$  as constant, one may think that this payment is computed according to the average severity of the whole population of patients. This specification, however, does not alter the qualitative results of the paper.

We model the physicians' behaviour as being that of a single representative agent. As we argued in the Introduction, the fact that the same doctor may work in both private and public practice is a common feature in Europe. We model this by assuming that the doctor who undertakes the operations in the public sector (in the morning, say) also works for a private hospital (in the afternoon). In defining the utility of the physician, therefore, we must not only take his revenues and costs in the public sector into account, but those in his private practice as well.

The physician incurs costs in providing treatment to public and private patients. These costs capture the physician's disutility from treating patients, and depend not only on the number of operations performed in each sector ( $x$  and  $1 - x$ , respectively), but also on the average severity of the patients ( $\hat{s}^{pb}$  and  $\hat{s}^{pv}$ ). As such, the physician's costs in each sector are given by:

$$\Psi^{pb} = \Psi^{pb}(x, \hat{s}^{pb})$$

$$\Psi^{pv} = \Psi^{pv}((1 - x), \hat{s}^{pv}).$$

The functions  $\Psi^i(\cdot)$ , with  $i = pb, pv$  are increasing and convex in both the amount and the severity of the patients treated. Moreover, denoting  $x = x^{pb}$  and  $1 - x = x^{pv}$ , we assume that  $\partial^2 \Psi^i / \partial x^i \partial \hat{s}^i > 0$  and  $\Psi^i(0, \hat{s}^i) = \partial \Psi^i / \partial x^i |_{x^i=0} = 0$ ,  $\forall i = pb, pv$ . The convexity of  $\Psi^i(\cdot)$  reflects the fact that the higher the amount and the greater the severity of patients treated by the physician, the greater his marginal disutility. Moreover, this formulation also captures the idea that physicians' working capacity in both sectors is limited.

As the physician performs two independent (although related) duties from which he receives different payments, we consider a construction with separable cost functions as more appropriate. We should also take into account that the time devoted by the physician to each activity may affect his costs. It may be reasonable to think, for

instance, that performing the same task in a shorter time might be more costly. For this reason, although working times are not explicitly modelled, they are implicitly captured by the fact that the physician's costs functions are different across sectors.

This modellization we have proposed is only appropriate if the working time of the physician in the public sector is fixed and exogenously given. This assumption can be justified as it reflects the actual situation in most European mixed health care systems, where full time physicians' working hours in the public sector range between 35 and 40 h per week.<sup>f</sup>

Concerning physician's revenues, and according to empirical evidence, we consider that they differ across sectors. First, in many public systems, physicians receive a salary conditional upon them providing a (loosely defined) amount of labour. The physician commits to perform  $x$  operations and receives, in exchange, a payment that covers the costs which he has incurred. We assume, hence, that the physician receives in the public hospital a transfer  $T$  that includes a fixed amount of money ( $R$ ) plus a full reimbursement of the costs, i.e.  $T = R + \Psi^{pb}(x, \hat{s}^{pb})$ . It is assumed that the health authority cannot observe the severity of each patient. Hence, the payment received by the physician ( $T$ ) is a function of the physician's total costs (see [8] for a similar set-up).

Secondly, as is often assumed in the literature (see, for instance [9,10] in the US context), we consider that the physician's private income is directly related to hospital profits. In particular, we define physician's private net revenue as a constant share of the private hospital profits ( $\pi$ ).

Physician's total utility is given by the sum of his net revenues in the public and in the private hospital. Hence, we can define the utility function of the physician as follows:

$$U^s = T - \Psi^{pb}(x, \hat{s}^{pb}) + \pi[(w - k^{pv})(1 - x)] - \Psi^{pv}((1 - x), \hat{s}^{pv}). \quad (1)$$

The health authority's expected surplus derived from the care provided is given by the difference between the patient's expected health benefits from treatment and the total costs that the provision of services generates. Hence, the health authority's objective function is given by:

$$H = Q - \delta x^2 - [T + k^{pb}x + w(1 - x)]. \quad (2)$$

The timing of the game is as follows: At a stage prior to the starting-point of our model, the health authority and the private hospital bargain over the value of the private fee  $w$  that the private hospital will receive per operation performed.<sup>g</sup> At the first stage, the health authority decides on two variables: the number of operations that are diverted to private clinics  $(1 - x)$ , and the payment the physician will receive from his public activity ( $R$  plus a reimbursement of the costs). At the second stage, the physician selects the patients he wants to treat in the public sector. The remaining patients will be transferred to the private hospital. Finally, the whole population of patients receives treatment and the payoffs are realized.

We confront two different frameworks. In the first one, it is assumed that the physician can not select patients on the basis of their severities. Stage 2 is, therefore, not active in this initial set-up. In the second scenario we consider that, since the actual severities of the patients can only be known by the physician, the health authority cannot monitor the physician in his selection of patients.

We start by analysing the optimal policy in the first setting.

### Benchmark scenario

In this section, we assume that the health authority can prevent the physician from selecting the patients he wants to treat in each sector. This can be understood as a situation in which the physician can not manipulate the waiting-list and, hence, patients are uniformly distributed between the public and the private sectors.<sup>h</sup> We can thus ensure that the average severity of the operations is the same in both sectors, i.e.  $\hat{s}^{pb} = \hat{s}^{pv} = \hat{s}$ .

In order to guarantee the existence of an interior solution in this framework, we make the following assumption.

*Assumption 1.*  $k^{pb} < w < \frac{\partial \Psi^{pb}(x, \hat{s})}{\partial x} |_{x=1} + 2\delta + k^{pb}$ .

Under this assumption, the fee paid to the private hospital has to be bounded between two values. With this we are only requiring that: On the one hand, the private fee cannot be too low. If this was the case, the health authority could purchase all the health services from the private sector, i.e. it would be trivially optimal to send all patients to the private sector. On the other hand, we also require that the private fee is not so high

that the policy of transferring patients is not undertaken, even in this framework where manipulation is not possible.

Since stage 2 is not active under the benchmark scenario, we move to the health authority optimization problem. The health authority maximizes its objective function, subject to the constraint that the payment made to the physician allows him to attain an exogenously fixed level of utility ( $\underline{U}$ ). This level can be understood as the physician's utility reservation. The optimization problem that the health authority faces is as follows:

$$\begin{aligned} \max_{x,T} H &= Q - \delta x^2 - [T + k^{pb}x + w(1 - x)] \\ \text{s.t. } T &\geq \underline{U} + \Psi^{pb}(x, \hat{s}^{pb}). \end{aligned}$$

The following lemma characterizes the optimal sharing of patients between public and private practice and the associated payment to the physician.<sup>i</sup>

*Lemma 1.* In the benchmark scenario, the optimal number of patients treated in the public sector ( $x^*$ ) and the payment the physician receives ( $T^*$ ) are such that:

$$\Psi_x^{pb}(x^*, \hat{s}) + k^{pb} + 2\delta x^* = w,$$

with  $T^* = \underline{U} + \Psi^{pb}(x^*, \hat{s})$ .

With the above lemma we have computed the optimal policy in the first best scenario. This will be our reference case for comparison with the results in the next section.

As one may expect, the health authority decides that the public sector performs the number of operations that equates the marginal costs of treating patients in both sectors. It is obvious that the optimal level of patients treated in the public sector is increasing in the private fee  $w$ . Moreover, as the fee per operation paid to the private hospital is fixed, the optimal number of patients treated in the public sector is decreasing in the average severity of the population ( $\hat{s}$ ). Finally, the larger the marginal loss associated with waiting for treatment ( $\delta$ ) the more patients will be transferred.

It is worth mentioning that if we allowed the health authority to strategically select which patients to treat, the outcome would be that only the most severe cases were transferred to the private sector. This is given by the fact that the payment to the private sector is a fixed fee per operation, while the public sector costs are increasing in the average severity of the patients.<sup>j</sup>

We proceed now to study the effects of dealing with a physician who can choose strategically which patients are to be treated in each sector. This will allow us to analyse the consequences of this potential strategic behaviour on the willingness of the health authority to undertake the policy.

## Patient selection

Our concern in this section is to analyse whether the results differ when the physician has the ability to select patients on the basis of their severity for treatment in one or other sector.

The actual severity of a patient is often difficult to ascertain. In any case, it is obvious that doctors and hospitals have better information than health authorities on any patient's severity (see, for instance, [11,12]). We consider, then, that the health authority is not able to monitor the severity of each patient treated. Although the overall average severity can be inferred ex-post (through the observability of the physician's costs), the health authority cannot prove whether the physician actually modified the pool of patients that were to be treated, i.e. it can not monitor ex-ante the severities of the patients that are chosen to receive treatment.<sup>k</sup> As the patients will no longer be randomly distributed between the public and private sectors, we cannot ensure, in general, that the average severity of the patients treated in both systems is the same.

To characterize the solution in this framework, we proceed to solve the game by backwards induction. At the second stage, the physician decides which severities he wants to treat in either sector, subject to the restriction that  $x$  operations have to be performed in his public practice. Therefore, he does not have complete freedom in the choice of  $\hat{s}^{pb}$ , as there may be values that are not compatible with the sharing of patients set by the health authority. The physician will choose the value of  $\hat{s}^{pb}$  (and therefore also of  $\hat{s}^{pv}$ ) in order to maximize his total revenue. Since he is a dual supplier, he will consider the effects of his strategic behaviour concerning his two sources of income.

In the following proposition we characterize the physician's behaviour concerning the selection of patients.

*Proposition 1.* For a given sharing of patients between the two sectors ( $x$  and  $1-x$ ), the

physician will transfer the least severe cases to the private sector. Formally:

A patient with severity  $s$  will be treated in the public practice if and only if:

$$s \in (s'(x), \bar{s}), \quad \text{with } s'(x) = F^{-1}(1-x).$$

This proposition shows that the physician wants to treat only the less severe cases in his private practice, leaving the more severe patients for the public sector. This behaviour, known in the literature as 'cream-skimming', is caused primarily by the difference between the physician's remunerations from the two systems. In the public sector, the physician has his costs fully reimbursed, independently of the amount and severity of the cases treated, whereas, in the private sector, he receives a share of the hospital profits. As a consequence, he has incentives to divert the 'easier' cases to his private practice as this will have a direct impact on his revenues. Note that what triggers physician behaviour is the fact that the payment structure in the private sector provides him with incentives for cost-containment (in order to maximize hospital profits), and this is not the case in the public sector.

As a consequence of Proposition 1, it is easy to see that the average severity in the public sector ( $\hat{s}^{pb} = \hat{s}^{pb}(x)$ ) depends negatively on the amount of patients that are kept on the waiting-list, i.e.  $\partial \hat{s}^{pb}(x) / \partial x < 0$ . The more patients that are kept waiting, the smaller the physician's capacity to select patients strategically and, hence, the lower the expected level of severity that the health authority faces.

We shall now study how this problem of cream-skimming affects the decision of the health authority on when to undertake a policy of diverting patients to private practice, and on the amount of patients that should be transferred to the private sector.

In the first stage, when the cost of implementing the policy is considered, the health authority should take into account the fact that the most severe cases will be treated in the public sector.

The maximization problem of the health authority in this scenario is as follows:

$$\begin{aligned} \max_{x,T} \quad & Q - \delta x^2 - [T + k^{pb}x + w(1-x)] \\ \text{s.t.} \quad & \begin{cases} T \geq \underline{U} + \Psi^{pb}(x, \hat{s}^{pb}) \\ \hat{s}^{pb} = \hat{s}^{pb}(x). \end{cases} \end{aligned}$$

The first-order necessary condition for an interior candidate to a solution ( $x^m$ ) is such that:

$$-2\delta x - k^{pb} + w - \frac{\partial \Psi^{pb}(x, \hat{s}^{pb}(x))}{\partial x} \Big|_{x^m} - \frac{\partial \Psi^{pb}(x, \hat{s}^{pb}(x))}{\partial \hat{s}^{pb}} \frac{\partial \hat{s}^{pb}(x)}{\partial x} \Big|_{x^m} = 0.$$

The decision of the health authority, when choosing the fraction of patients that will be diverted to the private sector, is driven by two main effects: On the one hand, there is a direct effect of changes in the number of patients kept in the public sector over the health authority's costs. On the other hand, there is an indirect effect, caused by the problem of patient selection. The health authority takes into account that the physician's capacity to behave strategically links the proportion of patients he has to treat in his public sector with the final distribution of the severities across sectors.

These effects have an immediate and important consequence, presented in the following proposition.

*Proposition 2.* The presence of patient selection may induce the health authority to decide not to implement the policy of transferring patients to private practice.

The presence of patient selection makes the strategy of reducing waiting-lists by transferring patients to the private practice, less attractive for the health authority. By diverting some patients, the health authority incurs increased marginal costs for those who remain in the public sector since, by the selection of the physician, they are the most severe cases.

The next step is to assess the impact of patient selection when the policy is actually undertaken. That is, assuming that the solution to cream-skimming is interior (i.e.  $x^m < 1$ ), will the health authority decide to maintain longer or shorter waiting-lists than in the benchmark scenario?

In order to analyse this issue, the curvature of the health authority's objective function is crucial. The added complexity of the problem means that, in general, we cannot ensure the objective function to be concave in all its domain.<sup>1</sup> Assuming that the first order condition is sufficient (i.e. the program is generically concave), it is straightforward to show the following:

*Proposition 3.* The distortion caused by patient selection on the proportion of patients diverted to private practice cannot be unambiguously ranked.

In particular, the number of patients kept in the public sector waiting-lists will be smaller under patient selection (i.e.  $x^* > x^m$ ) if and only if:

$$\frac{\partial \Psi^{pb}(x, \hat{s}^{pb}(x))}{\partial x} \Big|_{x^*} - \frac{\partial \Psi^{pb}(x, \hat{s})}{\partial x} \Big|_{x^*} > - \frac{\partial \Psi^{pb}(x, \hat{s}^{pb}(x))}{\partial \hat{s}^{pb}} \frac{\partial \hat{s}^{pb}(x)}{\partial x} \Big|_{x^*}.$$

The left-hand side of the inequality above ( $\partial \Psi^{pb}(x, \hat{s}^{pb}(x)) / \partial x|_{x^*} - \partial \Psi^{pb}(x, \hat{s}) / \partial x|_{x^*} > 0$ ) measures the increase in the marginal cost of treatment in the public sector due to cream-skimming. This effect induces the health authority to increase the fraction of patients transferred to the (cheaper) private sector. However, an indirect (strategic) effect emerges, that pushes in the opposite direction. This is reflected by the right hand side of the inequality ( $(\partial \Psi^{pb}(x, \hat{s}^{pb}(x)) / \partial \hat{s}^{pb})(\partial \hat{s}^{pb}(x) / \partial x|_{x^*}) < 0$ ). This term captures how a decrease in the number of patients transferred to the private sector reduces public sector marginal costs, by indirectly affecting the average severity in the public system. As already mentioned, by increasing the number of patients that remain in the public sector, the health authority limits the physician's strategic capacity to select patients.

At this level of generality, the relative importance of the two effects described above cannot be assessed and, therefore, no clear-cut prediction about the optimal transfer of patients to private practice can be made.

Next, we consider a specific physician's cost function, as well as a particular distribution for the severities. This will allow us to fully characterize the effects of the cream-skimming phenomenon. Moreover, it will show how the relative dispersion of the patients' severities, as well as the health losses associated with waiting for treatment, are the two key elements that determine the response of the health authority and its willingness to undertake the policy.

### Patient selection in the uniform-quadratic case

Throughout this section, we assume that the physician's cost of treatment in each sector is

given by a standard quadratic function:

$$\Psi^{\text{pb}} = \frac{1}{2}(x\hat{s}^{\text{pb}})^2 \quad \text{and} \quad \Psi^{\text{pv}} = \frac{1}{2}((1-x)\hat{s}^{\text{pv}})^2$$

On top of this we consider that the severities are distributed uniformly in the interval  $[\underline{s}, \bar{s}]$ . These two assumptions allow us to obtain explicit solutions to the problem with patient selection, and to derive meaningful comparisons to the benchmark scenario.

Before proceeding to analyse the optimal response from the health authority, we need to study the curvature of its objective function. For the sake of notational clarity, let  $d = \Delta s / \hat{s}$  denote the relative dispersion of the patients' severities. The following lemma characterizes the curvature as a function of  $x$ .

*Lemma 2.* The curvature of the health authority's objective function, under patient selection ( $H^m$ ), is as follows:

1. If  $\delta \geq \frac{3}{8}\Delta s^2$ , or if  $\delta < \frac{3}{8}\Delta s^2$  and  $d \leq \tilde{d}(\frac{\delta}{\Delta s^2})$ , then  $H^m$  is always concave.
2. If  $\delta < \frac{3}{8}\Delta s^2$  and  $d > \tilde{d}(\frac{\delta}{\Delta s^2})$ , then:
  - (a) For any  $x \in [0, \bar{x}(d, \frac{\delta}{\Delta s^2})]$ ,  $H^m$  is concave at  $x$ .
  - (b) For any  $x \in (\bar{x}(d, \frac{\delta}{\Delta s^2}), 1]$ ,  $H^m$  is convex at  $x$ .

$$\text{With } \tilde{d}(\frac{\delta}{\Delta s^2}) \equiv \frac{4-2\sqrt{3-8(\delta/\Delta s^2)}}{1+8(\delta/\Delta s^2)}, \quad \text{and } \bar{x}(d, \frac{\delta}{\Delta s^2}) = \left(\frac{1}{d} + \frac{1}{2}\right) - \sqrt{\left(\frac{1}{d} + \frac{1}{2}\right)^2 - \frac{4}{3} \frac{\delta}{\Delta s^2}}.$$

Lemma 2 allows us to study the curvature of the health authority's objective function in terms of the relative dispersion of the severities and the patients' health losses associated with waiting for treatment. If patients' health losses are very high, the function is still concave in all the domain and, therefore, the problem has a unique candidate for an optimum. When the value of such losses is small, then the relative dispersion of the severities turns out to be crucial. In this case, if the severities are sufficiently dispersed, the objective function has a convex section that is bigger the higher the dispersion. As a consequence of this feature, we may have two candidates to optimum: an interior one and the boundary solution (no transfer of patients to private practice).

The following proposition presents the solution to the health authority's maximization problem.

*Proposition 4.* In the presence of patient selection, the health authority decides to undertake the

policy of transferring patients to private practice if the value of the private fee is below a given threshold. The higher the relative dispersion of the patients' severities and the lower the health losses associated with waiting for treatment, the more demanding this condition is. Formally:

1.  $x^m < 1$  if  $w < 2\delta + k^{\text{pb}} + \hat{s}^2 G(d, \frac{\delta}{\Delta s^2})$ .
  2.  $x^m = 1$ , otherwise.
- where

$$G(d, \frac{\delta}{\Delta s^2}) = \begin{cases} -\frac{d}{2} & \text{if } \delta > \frac{3}{8}\Delta s^2, \text{ or if } \delta < \frac{3}{8}\Delta s^2 \text{ and } d \leq \tilde{d}(\frac{\delta}{\Delta s^2}) \\ g(d, \frac{\delta}{\Delta s^2}) & \text{otherwise.} \end{cases}$$

is a continuous function,  $g(d, \frac{\delta}{\Delta s^2})$  is such that  $\partial g(d, \delta/\Delta s^2)/\partial d < 0$  and  $\lim_{d \rightarrow \infty} g(d, \delta/\Delta s^2) = \frac{1}{2}$ .

As was already advanced in the previous section, this proposition shows that the presence of 'cream-skimming' can lead to a situation in which the health authority is no longer willing to transfer patients to the private sector.

The first insight that emerges from the proposition is that the larger the health losses associated with waiting for treatment, the more likely that the health authority decides to transfer patients to private practice. The waiting cost is a great source of inefficiency for the public sector even in the presence of cream-skimming. In addition to this, the threshold of the private fee at which the health authority is willing to carry out the policy is decreasing in the relative dispersion of the patients' severities. The reason is that the strategic behaviour of the physician (in transferring the milder cases to private practice) is fostered by the wider range of severities, since his gains in diverting patients are higher. As a result, when  $d$  is low the health authority does not suffer a substantial problem of cream-skimming (since patients have similar levels of severity). In this case, the policy of transferring patients to private practice is undertaken for a wide range of values of  $w$ . In particular, when  $d \rightarrow 0$ , the condition for undertaking the policy converges to the one required in Assumption 1. However, as the relative dispersion of severities increases, the condition necessary for the health authority to reach agreements with private hospitals becomes more demanding. Since the treatment of some patients in the private sector increases the cost per operation in the public sector, the health authority decides to retain the

size of the waiting-list, unless the private fee is sufficiently low.

The particular functional forms we have assumed allows us to compare the interior solution in this setting with the optimal one in the benchmark case. The result is the following:

*Proposition 5.* When the health authority is willing to undertake the policy, the existence of patient selection implies that:

- (i) If the relative dispersion of the severities is sufficiently low, the health authority keeps fewer patients in the public sector, provided that the private fee is below a given value. Otherwise, more patients are kept waiting.
- (ii) If the relative dispersion of the severities exceeds a critical value, the health authority always keeps fewer patients in the public sector.

Formally, there exists a threshold  $d^0(\delta/\Delta s^2) > 0$  such that:

1.  $x^m < x^*$  if  $d \geq d^0(\frac{\delta}{\Delta s^2})$ , or if  $d < d^0(\frac{\delta}{\Delta s^2})$  and  $w < 2\delta + k^{pb} + \hat{s}^2\Phi(d, \frac{\delta}{\Delta s^2})$ .
2.  $x^m > x^*$  if  $d < d^0(\frac{\delta}{\Delta s^2})$  and  $w > 2\delta + k^{pb} + \hat{s}^2\Phi(d, \frac{\delta}{\Delta s^2})$ .

$$\text{With } \Phi(d, \frac{\delta}{\Delta s^2}) = [\frac{1}{4d}(3d + 6 - \sqrt{d^2 + 4d + 36}) - \frac{\delta}{\Delta s^2} 2d^2].$$

This proposition shows the outcome of the unresolved conflict of effects presented in Proposition 3. We see how, when the policy of transferring patients is undertaken, the consequences of cream-skimming on the number of patients transferred are driven by the relative dispersion of the severities.

When the relative dispersion is sufficiently high, the negative effect of treating more severe patients in the public sector always dominates. Although the capacity of the physician to select patients decreases with the amount of public operations he has to perform, fewer patients are kept waiting. For this range of parameters, therefore, the health authority's reaction to the problem of patient selection is a more intensive transfer of patients from the waiting-list to private hospitals.

In contrast, when the relative dispersion is low, the final result is determined by the value of the private fee. A high level of  $w$  implies a relatively low transfer of patients in the benchmark scenario;

we show that, in the present case, the presence of cream-skimming leads to an even smaller transfer. In this region, therefore, cream-skimming leads to a smaller utilization of the private sector by the health authority and, thus, longer waiting-lists. Conversely, while the first best outcome was to transfer a high proportion of the patients (low  $w$ ) to private practice, the distortion created by cream-skimming implies transferring even more (see Figure 1).

Figure 2 compares the objective functions of the health authority and the optimal sharing of patients under the alternative scenarios we have studied:  $H^*$  denotes the objective function in the benchmark case, whereas  $H^m$  stands for the one

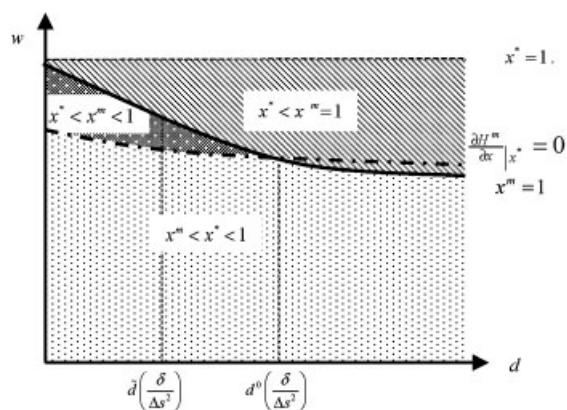


Figure 1. Sharing of patients with and without patient selection

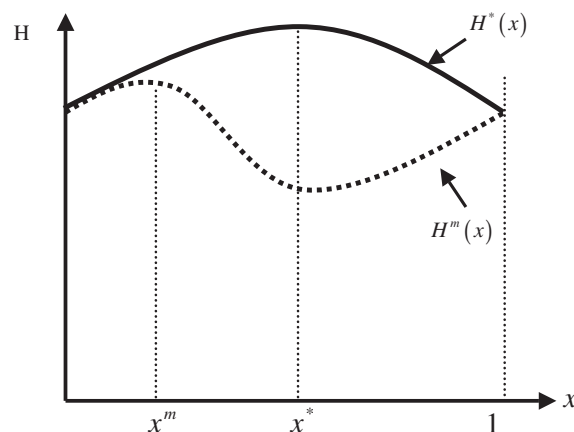


Figure 2. Health Authority's objective functions when  $d > d^0(\delta/\Delta s^2)$

under patient selection. This illustration is made for the case in which the policy is implemented, the value of  $\delta$  is not very high, and the relative dispersion of the severities is sufficiently large ( $d > d^0(\delta/\Delta s^2)$ ).

So far, we have taken the negotiations between the health authority and the private hospital, about the payment for those patients transferred, as given in our model. It is crucial, however, to know whether the equilibrium values we have computed leave room for such a bargaining process. That is, if there are values of  $w$  that make the health authority willing to undertake the policy (i.e.  $w < 2\delta + k^{pb} + \hat{s}^2 G(d, \delta/\Delta s^2)$ ) and, at the same time, are profitable to the private hospital ( $(w - k^{pv})(1 - x) - \frac{1}{2}((1 - x)s^{pv})^2 \geq 0$ ). The result is presented in the following remark:

*Remark 1.* In equilibrium, for each value of  $\delta$ , there exists a threshold  $\Delta k_{\min} < 0$  such that if  $\Delta k > \Delta k_{\min}$ , the bargaining set is not empty (for every value of  $d$ ), i.e. the maximum fee the health authority is willing to pay exceeds the minimum that the private hospital requires to accept public patients.

This remark shows that in all cases in which the private sector is more efficient than the public sector and, even in some cases in which it is not (when  $\Delta k_{\min} < \Delta k < 0$ ), there are values of  $w$  that are compatible to both the private hospital and the health authority, independently of the relative dispersion of the severities. Moreover, the region where agreement is possible is larger, the higher is the patients' loss associated with waiting for treatment ( $\delta$ ). Figure 3 illustrates the bargaining

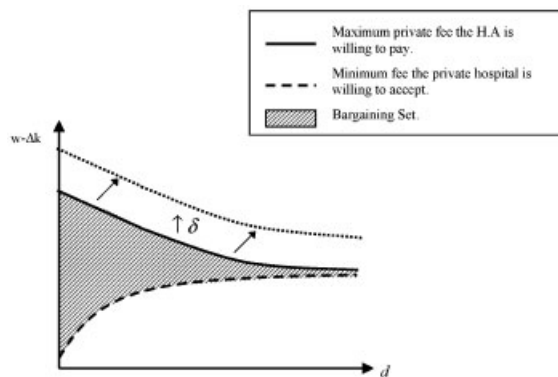


Figure 3. Bargaining set for  $\Delta k \geq \Delta k_{\min}$

set between the health authority and the private hospital for  $\Delta k \geq \Delta k_{\min}$ .

## Concluding remarks

This paper has analysed the incentives of health authorities to divert patients from the public to the private sector in a framework in which the public sector is capacity-constrained, while the private sector is not.

We have found that, due to the different structure of the physician's remuneration in each system, a problem of cream-skimming arises, i.e. physicians prefer to treat the less severe cases in their private practices. This problem has been shown to make the health authority more reluctant to implement a policy of transferring patients to private practice.

We have characterized the range of parameters that make the health authority find it more profitable to keep longer waiting-lists instead of moving patients to private practice. There are two crucial variables, apart from the private fee, that influence the trade-off the health authority faces: On the one hand, the relative dispersion of the patients' severities. The higher the dispersion, the more the physician earns from selecting patients and, at the same time, the greater the impact on the costs borne by the health authority. On the other hand, the patients' health losses associated with waiting for treatment. The larger these losses, the more willing the health authority to resort to private practice.

When the health authority decides to undertake the policy, the amount of patients that are finally transferred to the private sector cannot be determined unambiguously. We have found, however, that for a wide range of parameter values, fewer patients are kept waiting. The negative effect of cream-skimming on the marginal cost of providing treatment in the public sector dominates (as the most severe patients remain in it). Hence, more patients are transferred, despite this increasing the ability of the physician to select patients.

In terms of policy recommendations, our result suggests that when designing policies to reduce waiting-lists by transferring patients from the public to the private sector, policy-makers should consider the incentive effects of the different reimbursement systems of the two sectors. More-

over, the decision to undertake the policy or not must be influenced not only by the fee per operation agreed with the private sector, but also by the type of illness. In particular, the dispersion of illness severity is shown to be very important. The wider the range of severities, the more serious the problem of patient selection becomes and, therefore, the less beneficial the policy to the health authority.

The spirit of this work is eminently positive. Given the institutional framework and the contractual arrangements in existence, we have identified a potential problem of patient selection. A more normative analysis, in which possible solutions to the problem can be addressed, is left for further research. In spite of this, our article provides some indications on how to address the issue. One may think of two possible types of instruments. The first concerns physician reimbursement. Different contractual arrangements would alter physicians' incentives and, hence, they deserve a careful analysis.<sup>m</sup> A second option is to consider measures that limit the physician's strategic capacity. Possible measures include the introduction of exogenous rules for allocating patients between the two sectors and legal exclusions that prevent physicians from providing private treatment to patients that were on their waiting-lists.<sup>n</sup>

One potential criticism of our analysis is that we do not consider how a physician's professional ethics may alleviate cream-skimming. However, this criticism does not apply to our model since the costs associated with waiting for treatment in the public sector are equal among agents and, hence, an altruistic motive would not alter physician's incentives to perform cream-skimming. If the direct cost of waiting were increasing in the seriousness of the patients' condition, it is true that doctors' altruism may partially mitigate their incentives to select the less severe patients. In this sense, our analysis could be considered as a benchmark for the worst-case scenario.

There are other issues that have not been addressed in this work. First, we have not considered any difference between the quality of the services provided by the public and private sectors. This makes our analysis appropriate for treatments when the patient's condition is not life-threatening in the absence of treatment, as these medical conditions usually require facilities that both the public and the private sectors possess. Moreover, these non-urgent treatments are pre-

cisely the ones included in most plans for diverting patients. Secondly, the analysis would require a first step to study the negotiations between the health authority and the private hospitals concerning the payment for those operations transferred. Endogeneizing this process would be an interesting exercise that would provide the health authority with a new regulatory tool to tackle the cream-skimming problem. However, as long as the payment to the private sector is a fixed fee per operation, our results would continue to hold.

Certainly, more theoretical and empirical work in this line of research is needed. We believe, however, that this work can inform the debates on health policy and contribute to the development of a better policy making process.

## Appendix

### Proof of Proposition 1

The optimization program the physician faces is as follows:

$$\max_{\hat{s}^{pb}} U^s = T - \Psi^{pb}(x, \hat{s}^{pb}) + \pi[(w - k^{pv})(1 - x)] \\ - \Psi^{pv}((1 - x), \hat{s}^{pv}),$$

with  $\partial \hat{s}^{pv} / \partial \hat{s}^{pb} < 0$ . Since  $T = R + \Psi^{pb}(x, \hat{s}^{pb})$ , we can rewrite this problem as:

$$\max_{\hat{s}^{pb}} U^s = R + \pi[(w - k^{pv})(1 - x)] \\ - \Psi^{pv}((1 - x), \hat{s}^{pv}).$$

From this expression, it is straightforward to see that the physician's utility is higher the lower is  $\hat{s}^{pv}$  (and, therefore, the higher is  $\hat{s}^{pb}$ ). This implies that the physician chooses to transfer to the private sector those patients on the left tail of the distribution of severities. Formally, there exists a  $s' \in (\underline{s}, \bar{s})$  such that:

If  $s \in (\underline{s}, s')$  the patient is transferred to the private sector.

If  $s \in (s', \bar{s})$  the patient remains in the public sector.

In order to compute the value of  $s'$ , we have to take into account that a fraction  $x$  of patients have to be treated in the public sector. Thus:

$$\int_{s'(x)}^{\bar{s}} dF(s) ds = x \Rightarrow s'(x) = F^{-1}(1 - x).$$

Therefore, a patient with severity  $s$  will be treated in the public sector if and only if:

$$s \in (s'(x), \bar{s}), \quad \text{with } s'(x) = F^{-1}(1 - x)$$

This completes the proof. □

**Proof of Proposition 2**

In the benchmark case, Assumption 1 (rewritten as  $w - 2\delta - k^{pb} - \partial\Psi^{pb}(x, \hat{s})/\partial x|_{x=1} < 0$ ) was enough to ensure that  $x^* < 1$ . However, in the problem with patient selection, the f.o.c. evaluated at  $x = 1$  is given by:

$$w - 2\delta - k^{pb} - \frac{\partial\Psi^{pb}(x, \hat{s})}{\partial x} \Big|_{x=1} - \frac{\partial\Psi^{pb}(x, \hat{s}^{pb}(x))}{\partial \hat{s}^{pb}} \frac{\partial \hat{s}^{pb}(x)}{\partial x} \Big|_{x=1}$$

This expression is larger than  $w - 2\delta - k^{pb} - \frac{\partial\Psi^{pb}(x, \hat{s})}{\partial x}|_{x=1}$  and, therefore, Assumption 1 is not enough to ensure an interior solution for the problem with patient selection.

This completes the proof. □

**Proof of Lemma 2**

We study the curvature of the health authority's objective function in the restricted domain given by the behaviour of the physician at stage 2. From Proposition 1 we have  $\hat{s}^{pb} = \bar{s} - x\Delta s/2$ . However, it is known that  $T = R + \frac{1}{2}[x\hat{s}^{pb}]^2$ . By substituting these two constraints into the problem, we can characterize the curvature of the objective function as a function of  $x$ . The optimization program at the first stage is as follows:

$$\begin{aligned} \max_{x,R} \quad & H^m = Q - \delta x^2 \\ & - \left[ R + \frac{1}{2} \left[ x \left( \bar{s} - \frac{x\Delta s}{2} \right) \right]^2 + k^{pb}x + w(1 - x) \right] \\ \text{s.t.} \quad & R \geq \underline{U}. \end{aligned}$$

The f.o.c. is given by:

$$\begin{aligned} -x \left[ \left( \bar{s} - \frac{x\Delta s}{2} \right) \right]^2 + \frac{x^2\Delta s}{2} \left[ \left( \bar{s} - \frac{x\Delta s}{2} \right) \right] \\ + w - k^{pb} - 2\delta x = 0. \end{aligned}$$

Computing the second order condition and rearranging terms yields:

$$\frac{\partial^2 H^m}{\partial^2 x} = -\bar{s}^2 + 3x\Delta s\bar{s} - \frac{3}{2}\Delta s^2 x^2 - 2\delta.$$

This second order derivative is increasing in  $x$ . We compute its roots and find:

$$\begin{aligned} x_1 &= \frac{\bar{s}}{\Delta s} + \sqrt{\left(\frac{\bar{s}}{\Delta s}\right)^2 \frac{1}{3} - \frac{4}{3} \frac{\delta}{\Delta s^2}} > 1 \quad \text{and} \\ x_2 &= \frac{\bar{s}}{\Delta s} - \sqrt{\left(\frac{\bar{s}}{\Delta s}\right)^2 \frac{1}{3} - \frac{4}{3} \frac{\delta}{\Delta s^2}}. \end{aligned}$$

Simple manipulations allow us to re-write  $x_2$  as

$$x_2 = \left(\frac{1}{d} + \frac{1}{2}\right) - \sqrt{\left(\frac{1}{d} + \frac{1}{2}\right)^2 \frac{1}{3} - \frac{4}{3} \frac{\delta}{\Delta s^2}} \equiv \bar{x}(d, \frac{\delta}{\Delta s^2}).$$

It can be verified that  $\partial\bar{x}(d, \delta/\Delta s^2)/\partial d < 0$ .

From here, since  $x \leq 1$ , it can be shown that:

- (i) If  $\bar{x}(d, \delta/\Delta s^2) \geq 1$ ,  $\partial^2 H^m/\partial^2 x < 0 \quad \forall x \in [0, 1]$ , i.e. the objective function is always concave.
- (ii) If  $\bar{x}(d, \delta/\Delta s^2) < 1$ ,  $\partial^2 H^m/\partial^2 x < 0$  for every  $x \leq \bar{x}(d, \delta/\Delta s^2)$ . Thus, we can ensure that for  $x \in [0, \bar{x}(d, \delta/\Delta s^2)]$  the objective function is concave, and for  $x \in [\bar{x}(d, \delta/\Delta s^2), 1]$  it is convex.

In re-writing the conditions for  $\bar{x}(d, \delta/\Delta s^2)$  in terms of the relative dispersion of the severities, we find that:

$$\begin{aligned} \bar{x}(d, \delta/\Delta s^2) \geq 1 &\Leftrightarrow \delta \geq \frac{3}{8}\Delta s^2, \quad \text{or if } \delta < \frac{3}{8}\Delta s^2 \text{ and} \\ &\leq \tilde{d}\left(\frac{\delta}{\Delta s^2}\right) \equiv \frac{4 - 2\sqrt{3 - 8\delta/\Delta s^2}}{1 + 8\delta/\Delta s^2} \\ \bar{x}(d, \delta/\Delta s^2) < 1 &\text{ otherwise.} \end{aligned}$$

This completes the proof. □

**Proof of Proposition 4**

From the f.o.c. computed in Lemma 2 it is straightforward to verify that  $x = 0$  can be never a solution, since:

$$\frac{\partial H^m}{\partial x} \Big|_{x=0} = w - k^{pb} > 0.$$

This, together with the other conditions found in Lemma 2, provides a complete characterization of

the optimization problem:

1. If  $\delta \geq \frac{3}{8}\Delta s^2$ , or if  $\delta < \frac{3}{8}\Delta s^2$  and  $d \leq \tilde{d}(\delta/\Delta s^2)$ , there exists a unique candidate to optimum  $(x)$ . This solution will be interior, i.e.  $x \in (0, 1)$ , if and only if:

$$\frac{\partial H^m}{\partial x} \Big|_{x=1} = w - k^{pb} - 2\delta - \hat{s} \left( \hat{s} - \frac{\Delta s}{2} \right)$$

$$< 0 \Leftrightarrow w - k^{pb} - 2\delta < \hat{s}^2 \left( 1 - \frac{d}{2} \right).$$

2. If  $\delta < \frac{3}{8}\Delta s^2$  and  $d > \tilde{d}(\delta/\Delta s^2)$ , there exists, at most, a single interior candidate to optimum  $(x)$ , such that  $x \in (0, \bar{x}(d, \delta/\Delta s^2)]$ . The boundary solution  $x = 1$  is also a potential candidate to optimum, provided that:

$$\frac{\partial H^m}{\partial x} \Big|_{x=1} > 0 \Leftrightarrow w - k^{pb} - 2\delta > \hat{s}^2 \left( 1 - \frac{d}{2} \right).$$

This is a necessary, but not sufficient condition, for  $x = 1$  to be a solution. Hence, to choose the optimal level of  $x$  in this region we need to compare the value function for both candidates. We check the conditions under which it is better to operate a fraction  $\alpha$  of the patients, rather than perform all the operations. We find that:

$$\text{If } w - k^{pb} - 2\delta < \hat{s}^2 g \left( d, \frac{\delta}{\Delta s^2} \right),$$

$$\text{with } g \left( d, \frac{\delta}{\Delta s^2} \right)$$

$$= \max_{\alpha \in (0,1)} \frac{1}{2} \left[ (1 + \alpha) - \alpha^2 d \left( \frac{(1 - \alpha)}{4} d + 1 \right) - \frac{\delta}{\Delta s^2} 2(1 - x)d^2 \right] \exists \alpha^* \in (0, 1)$$

such that  $H^m(x = \alpha^*) > H^m(x = 1)$ .

We can see that  $g(d, \delta/\Delta s^2)$  is such that  $\partial g(d, \delta/\Delta s^2)/\partial d < 0$ ,  $\lim_{d \rightarrow \infty} g(d, \delta/\Delta s^2) = \frac{1}{2}$ .

If  $w - k^{pb} - 2\delta$  exceeds the above threshold, then we can ensure that the boundary solution is optimal, since for this parameter configuration  $\partial H^m/\partial x|_{x=1} > 0$ . Therefore, the solution to the principal's problem is:

$$(i) \ x^m < 1 \text{ if } w < 2\delta + k^{pb} + \hat{s}^2 G \left( d, \frac{\delta}{\Delta s^2} \right),$$

where

$$G \left( d, \frac{\delta}{\Delta s^2} \right) = \begin{cases} 1 - \frac{d}{2} & \text{if } \delta > \frac{3}{8}\Delta s^2, \text{ or if } \delta < \frac{3}{8}\Delta s^2 \text{ and } d \leq \tilde{d} \left( \frac{\delta}{\Delta s^2} \right) \\ g(d, \delta/\Delta s^2) & \text{otherwise} \end{cases}$$

is a continuous function,  $g(d, \delta/\Delta s^2)$  is such that  $\partial g(d, \delta/\Delta s^2)/\partial d < 0$  and  $\lim_{d \rightarrow \infty} g(d, \delta/\Delta s^2) = \frac{1}{2}$ .

(ii)  $x^m = 1$ , otherwise.

Finally, from  $\partial H^m/\partial x|_{x=1}$  it is also straightforward to see that, in both regions, the larger is  $\delta$  the more likely it is that the interior solution is optimal.

This completes the proof.  $\square$

### Proof of Proposition 5

From the previous proposition we know that  $x^m$  is interior if:

$$w < 2\delta + k^{pb} + \hat{s}^2 G \left( d, \frac{\delta}{\Delta s^2} \right).$$

The f.o.c. of the health authority's problem in the presence of patient selection was given by:

$$\frac{\partial H^m}{\partial x} = -x \left( \bar{s} - \frac{x\Delta s}{2} \right)^2 + \frac{x^2 \Delta s}{2} \left( \bar{s} - \frac{x\Delta s}{2} \right) + w - k^{pb} - 2\delta x = 0.$$

Considering that  $\bar{s} = \hat{s} + \frac{1}{2}\Delta s$ , we can rewrite this in terms of  $\hat{s}$  as:

$$-x \left( \hat{s} + (1 - x) \frac{\Delta s}{2} \right)^2 + \frac{x^2 \Delta s}{2} \left( \hat{s} + (1 - x) \frac{\Delta s}{2} \right) + w - k^{pb} - 2\delta x = 0.$$

The f.o.c. of the health authority's problem under non-manipulability was given by:

$$\frac{\partial H}{\partial x} = -x\hat{s}^2 + w - k^{pb} - 2\delta x = 0.$$

Performing some algebraic manipulations, we find that:

$$\frac{\partial H^m}{\partial x} > \frac{\partial H}{\partial x} \Leftrightarrow x\Delta s \hat{s} \left[ -(x)^2 \frac{d}{2} + \frac{3}{2}x \left( 1 + \frac{d}{2} \right) - \left( 1 + \frac{d}{4} \right) \right] > 0.$$

From the inequality above we find that:

$$\begin{aligned} \frac{\partial H^m}{\partial x} \Big|_{x=x^*} > 0 &\Leftrightarrow w - k^{pb} \\ &> \hat{s}^2 \left[ \frac{1}{4d} \left( 3d + 6 - \sqrt{d^2 + 4d + 36} \right) \right]. \end{aligned}$$

Rewriting it, we get:

$$w - k^{pb} - 2\delta > \hat{s}^2 \left[ \frac{1}{4d} \left( 3d + 6 - \sqrt{d^2 + 4d + 36} \right) \right].$$

In order to complete the characterization of the solution  $x^m$ , we need to check when the inequality above is compatible with the condition guaranteeing that  $x^m$  is interior. On combining the above condition with that of Proposition 4, we find that there exists a threshold  $d^0(\delta/\Delta s^2) > 0$  such that:

- (1) If  $d < d^0(\delta/\Delta s^2)$ , then:
    - If  $w < 2\delta + k^{pb} + \hat{s}^2\Phi(d, \delta/\Delta s^2)$ , then:  $x^m < x^*$ .
    - If  $w > 2\delta + k^{pb} + \hat{s}^2\Phi(d, \delta/\Delta s^2)$ , then:  $x^m > x^*$ .
    - With  $\Phi(d, \delta/\Delta s^2) = [(3d + 6 - \sqrt{d^2 + 4d + 36})/4d - \delta 2d^2/\Delta s^2]$ .
  - (2) When  $d \geq d^0(\delta/\Delta s^2)$  then:  $x^m < x^*$ .
- This completes the proof.  $\square$

## Acknowledgements

This is a substantially rewritten version of a paper entitled 'Policy Implications of Transferring Patients to Private Practice'. Earlier versions of this paper were written at Universidad de Alicante (Spain) and at CORE-Université catholique de Louvain (Belgium). I am truly grateful to Carmen Herrero, Javier López-Cuñat and Nicolás Porteiro for their invaluable help and support. I also would like to thank Pedro P. Barros, Tor Iversen, Izabela Jelovac, Inés Macho-Stadler, Xavier Martínez-Giralt, Pau Olivella and Andrew Street for valuable discussions and helpful suggestions. Suggestions from two anonymous referees have improved the paper considerably. Financial support from the Generalitat Valenciana GV01-371 and from the Instituto Valenciano de Investigaciones Económicas (IVIE) is gratefully acknowledged. The usual disclaimers apply.

## Notes

- a. This information has been obtained from the journal 'Información' (4th January 2001) and the Journal 'Faro de Vigo' (22nd September 2002).
- b. A report by the Competition Commission [1], estimated that 61% of NHS physicians in the UK have significant private practices. In addition to this, and according to Yates [2], an NHS specialist

- undertakes, on average, two private operations a week. In the Southern European countries, this phenomenon seems to be even more common.
- c. Cream-skimming may also appear in other frameworks. For instance, the editorial of *The Economist* (1998) addresses the criticism directed towards Health Maintenance Organizations in the US for excluding people with high expected costs from purchasing health insurance.
  - d. According to Barros and Olivella, 'full cream-skimming' is a situation in which all the mildest patients end up being treated in the private sector.
  - e. In this work it is assumed implicitly that medical guidelines have set minimum thresholds of acuteness for a patient to be admitted for treatment. The decision of the health authority is to determine the amount of patients, from those who will receive treatment, that should be transferred to private clinics.
  - f. As pointed out by an anonymous referee, if working time in the two sectors adjusted perfectly to the number of patients treated (instead of being exogenous), it would be more reasonable to consider that the total cost of the physician depends only on the total number of patients treated and on the average severity.
  - g. We show later that, in equilibrium, the bargaining set is not empty, i.e. the maximum fee the health authority is willing to pay exceeds the minimum the private hospital will accept for attending to public patients.
  - h. We are ruling out the possibility that the health authority strategically decides to select where patients are treated, as a way to decrease the costs borne. We will discuss this issue briefly at the end of the section.
  - i. The proof of this lemma is straightforward and therefore we omit it.
  - j. We thank an anonymous referee for pointing out the possibility of patient selection by the health authority.
  - k. As a consequence of this, payments to public physicians are a function of total costs and number of patients treated.
  - l. Using the fact that in the absence of patient selection the problem is always concave, it can be shown that the concavity is preserved in two cases. First, when the health loss associated with waiting for treatment ( $\delta$ ) is sufficiently large. Secondly, if the dependence of  $\hat{s}^{pb}(x)$  on  $x$  is relatively weak. This latter condition can be seen as a restriction on the distribution of the severities. In particular, we will show in the following section that, for an uniform distribution of severities, it holds if the relative dispersion of the severities ( $\Delta s/\hat{s}$ ) is below a certain threshold.
  - m. Ellis [8], Ellis and McGuire [9,10], Blomqvist [13], Dranove [14], Ma [15], Rickman and McGuire [16] and Selden [17] have written extensively on this issue.

- n. In this regard, some health authorities have imposed restrictions on the private earnings of the publicly employed physicians to limit their strategic capacity. González analyses in detail this issue [18].

## References

1. Competition Commission. Private Medical Services: A Report on the Agreements and Practices Relating to Charges for the Supply of Private Medical Services by NHS consultants. London: HMSO, 1994.
2. Yates J. Private Eye, Heart and Hip: Surgical Consultants, The National Health Service and Private Medicine. London: Churchill Livingstone, 1995.
3. Olivella P. Shifting public-health-sector waiting lists to the private sector. *Eur J Polit Econ* 2003; **19**: 103–132.
4. Bosanquet N. A Successful National Health Service. Adam Smith Institute: London, 1999.
5. Iversen T. The effect of a private sector on the waiting time of a national health service. *J Health Econ* 1997; **16**: 381–396.
6. Cullis JG, Jones PR. National health service waiting lists: a discussion of competing explanations and a policy proposal. *J Health Econ* 1985; **4**: 119–135.
7. Hoel M, Sæther EM. Public health care with waiting time: the role of supplementary private health care. *J Health Econ* 2003; **22–4**: 599–616.
8. Ellis RP. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *J Health Econ* 1998; **17**: 537–555.
9. Ellis RP, McGuire TG. Provider behavior under prospective reimbursement: cost sharing and supply. *J Health Econ* 1986; **5**: 129–152.
10. Ellis RP, McGuire TG. Optimal payment systems for health services. *J Health Econ* 1990; **9**: 375–396.
11. Boadway R, Marchand M, Sato M. An optimal contract approach to hospital financing. *J Health Econ* 2004; **23–1**: 85–110.
12. Propper C. Agency and incentives in the NHS internal market. *Soc Sci Med* 1995; **40**: 1683–1690.
13. Blomqvist A. The doctor as a double agent: information asymmetry, health insurance, and medical care. *J Health Econ* 1991; **10**: 411–432.
14. Dranove D. Rate-setting by diagnosis related groups and hospital specialization. *Rand J Econ* 1987; **18–3**: 417–427.
15. Ma C-TA. Health care payment systems: costs and quality incentives. *J Econ Manage Strat* 1994; **3–1**: 93–112.
16. Rickman N, McGuire A. Regulating providers' reimbursement in a mixed market for health care. *Scott J Polit Econ* 1999; **46–1**: 53–71.
17. Selden TP. A model of capitation. *J Health Econ* 1990; **9**: 397–410.
18. González P. Should physicians' dual practice be limited? An incentive approach. *Health Econ* 2004; **13**: 505–524.