

# POLICY IMPLICATIONS OF TRANSFERRING PATIENTS TO PRIVATE PRACTICE

Paula González

WP-AD 2002-12

Correspondence: Universidad de Alicante. Departamento de Fundamentos del Análisis Económico. Apdo. Correos 99. 03080 Alicante (Spain). Phone: (34) 965 90 36 14, Ext. 2627. Fax. (34) 965 90 38 98. E-mail: paula@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

First Edition: July 2002.

Depósito Legal: V-2752-2002

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

---

\* I am truly grateful to Carmen Herrero and Javier López-Cuñat for their invaluable help and support. I also would like to thank Izabela Jelovac, Inés Macho-Stadler, Pau Olivella, Nicolás Porteiro and the participants at the I Encuentro de Economía Industrial (Barcelona) and at the 2<sup>nd</sup> Health Economics Workshop (Lisbon) for their helpful comments. Financial support from the Generalitat Valenciana Project GV01-371 and from the IVIE is gratefully acknowledged. The usual disclaimers apply.

\*\* University of Alicante.

# **POLICY IMPLICATIONS OF TRANSFERRING PATIENTS TO PRIVATE PRACTICE**

**Paula González**

## **A B S T R A C T**

We construct a model to analyze the willingness of Health Authorities to reach agreements with private hospitals to have some of their public sector patients treated there. When physicians are dual suppliers, we show that a problem of cream-skimming arises and reduces the incentives of the government to undertake such a policy. We argue that the more disperse the severities of the patients are, the greater the reduction in the incentives will be. Moreover, we characterize the distortion that the cream-skimming phenomenon imposes on the characteristics of the policy, when this is implemented.

**KEYWORDS:** Regulation; Physician's Incentives; Public-Private Health Services; Patient Selection.

# 1 INTRODUCTION

Public health services, worldwide, are plagued by over-crowding 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. The general population is particularly sensitive to the congestion within the system as they suffer the direct effects of long waiting-lists for urgently needed operations.

The general discomfort caused by the back-log has been forcing several national health authorities, the Spanish Ministry of Health included, to turn to private hospitals and clinics for assistance in reducing their ever-increasing waiting-lists.<sup>1</sup> The Spanish Health Authority, moreover, in an effort to optimize its health system, not only allows certain patients on its Social Security waiting-lists to be treated at private hospitals, but also uses its own operating-theaters outside regular working hours.

At the Spanish regional level, the Catalanian government approved a budget of almost seven million Euros to shorten waiting-lists during 2001.<sup>2</sup> The Valencian Region has been undertaking the 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. Moreover, in this region 4.26 million Euros were spent last year to defray the debt to private clinics that have participated in the "Impact Plan" for reducing surgical waiting-lists.<sup>3</sup>

Such temporary programs, however, cannot solve the problem and can turn out to be extremely costly. Finding the correct balance between cost-containment and improvements in the provision of health care services has, therefore, become a major endeavor in most European economies.

This makes the study of the adequacy and optimality of the policies of distributing patients between the public and private sectors crucial. There is quite a lot of controversy over whether hospital specialists are able to influence and manage these waiting-lists for elective surgery to their own private benefit, which would generate an important negative impact on the public sector budget.

---

<sup>1</sup>The British government, in addition to this, has recently decided to allow a significant percentage of its patients awaiting surgery to be operated in France.

<sup>2</sup>See, for instance, the journal "La Vanguardia", 19th April 2001.

<sup>3</sup>See the journal "Información", 4th January 2001.

It is common in countries with public health services and waiting-lists, that the doctors who work for government hospitals also have their private practice. In the UK, for instance, most private medical services are provided by physicians whose main commitment is to their public sector duties. A report by the Competition Commission (1994), estimated that 61% of NHS physicians in the UK have significant private practices. In addition to this, and according to Yates (1995), an NHS specialist undertakes, on average, two private operations a week. In the Southern European countries, this phenomenon seems to be even more common.

Furthermore, there is a significant difference between the forms of payment to the doctors in the public and private sectors. In the private practice the physician charges a fee for his services, while in the public sector he has a fixed salary.

These two features, doctors acting in both private and public sector and different remuneration schemes in both sectors, raise a basic matter. Patient-selection (cream-skimming) by the physicians may appear, i.e., the physicians can have incentives to strategically divert the easiest cases to their private practice.<sup>4</sup>

This behavior, moreover, can hardly be avoided, as the evaluation of the diagnostic information required to assess the severity of a patient can only be performed by a trained physician. The control over the severities of the patients who receive treatment in each sector is, therefore, likely to be out of the monitoring capacity of the Health Authority.<sup>5</sup>

The aim of this paper is to analyze the consequences of transferring patients to private practices, and the circumstances under which the Health Authority should implement it.

In our analysis, it is implicitly assumed that the private sector is operating under capacity. This is also an empirical observation. For instance, Bosanquet (1999) states that "at present, there is under-occupation in private hospitals (in the UK), with occupancy rates at 50% or less".

Our starting point is a simple model in which the policy-maker (the Health Authority) contracts a hospital specialist for treating patients with different severities, and reaches

---

<sup>4</sup>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 costly cases.

<sup>5</sup>Arrow (1963) was the first to analyze the health care market, taking the differences of information held by the different agents involved into account. Gaynor (1994) and Propper (1995) provide interesting discussions about this topic.

agreements with private hospitals to have the remaining patients treated there. The Health Authority agrees to pay a fixed fee per operation performed by the private sector.

Our analysis is appropriate for treatments when the patient's condition is not life-threatening in the absence of treatment. These medical disciplines usually require facilities that both the public and the private sectors possess. Hence, it is reasonable to consider, as we do in this model, that there are no differences in quality between the two sectors. Moreover, these non-urgent treatments are precisely the ones included in most plans for diverting patients.

The objective of our work is two-fold. On the one hand, we characterize the physician's behavior when the government undertakes a policy of transferring some of the public health patients to private practice. We show that when the government is not able to monitor the physician's behavior with regard to which severities to treat, a problem of cream-skimming arises. The physician will transfer the least severe cases to the private sector.

On the other hand, we also study how this feature affects the decision of the Health Authority concerning when to carry out the policy and how to distribute the patients between the two sectors. We show that the presence of cream-skimming reduces the incentives of the Health Authority to undertake the policy. The reason for this is the increase in the costs borne by the public sector due to the existence of patient-selection. The fact that the physician only transfers the mildest cases to private practice, increases the average severity of those patients who remain in the public sector and the expenditure the Health Authority faces also increases.

We find, moreover, that the relevant measure for evaluating the importance of the problem of patient-selection is the relative dispersion of the severities of the patients. The higher the dispersion is, 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 is.

We also characterize the distortion that the cream-skimming phenomenon imposes on the characteristics of the policy of transferring patients (when it is eventually implemented). This helps us to establish comparisons with the actual performance of these kinds of measures. In this respect, there is empirical evidence supporting the idea that, when the policy is undertaken, the number of operations that are performed in the public sector decrease slightly. Our model provides rationality to this phenomenon, based on

the strategic behavior of the physicians. Since they keep the most severe patients in the public sector, the amount of patients that can be treated there, for a given level of effort, is reduced.

The physician's response to the form of the compensation contract has been widely covered by the literature, which generally focuses on retrospective versus capitation reimbursement methods. The main concern of these works is the effect of the reimbursement rule on either the intensity of health services (see, for instance, Ellis and McGuire (1986, 1990), Selden (1990), Blomqvist (1991), and Rickman and McGuire (1999)), or on the provider selection of who will be treated (see, for instance, Dranove (1997)), or both on the intensity and extent of the treatment (see, for instance, Ma (1994), Ellis (1999)). Only Rickman and McGuire (1999) consider, as we do, the fact that the physician can supply either private health-care to a patient or public.

In our model, intensity of treatment is not considered, as we focus exclusively on the physician's selection of patients. Even if this were the aim of some of the papers mentioned above, our approach is quite different from theirs. We take the remuneration system as given (and fixed) by the institutional framework and we study the reaction of the policy-maker to the potential strategic behavior of the physician. Hence, in our work, cream-skimming does not appear as a consequence of the remuneration system chosen by the Public Authority, but rather due to the different structure of payments in the public and private sectors.

This paper should also be included in the literature that considers a mix of public and private sector services provided by physicians. Iversen (1997) has modelled the impact of public sector waiting-lists on the demand for private care. He concentrates on the patient's decision. He assumes that all patients who are willing to pay for private treatment are served in the private sector. As such, he clearly rules out the possibility of cream-skimming on the part of the doctors.

Patient-selection by the physician is also ruled out in Olivella (2001). He analyzes the incentives of the public health administration (fixing long waiting times for public treatments) to divert costs from the public to the private sector and studies the conditions under which this deviation enhances welfare. In our model, such a behavior does not appear since it is the Health Authority who pays the cost of treating all the social security patients (independently of where they are treated).

Finally, the doctor's strategic behavior plays an important role in Barros and Olivella (1999). 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 private treatment (paying a flat fee). In our model, patients are completely passive agents 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>6</sup>

The rest of the paper is organized as follows: In the following section we present the model. Section 3 computes the optimal policy in the benchmark scenario. In section 4 we study the behavior of the physician concerning the selection of patients and the response of the Health Authority. In Section 5, we show how our results can be extended to several variants of our model. Finally, Section 6 provides some concluding remarks and the policy implications of our analysis. All of the proofs are in the Appendix.

## 2 THE MODEL

There is a continuum of individuals requiring health care, all of whom demand elective treatment. The size of this population of potential patients is normalized to  $N$ . 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 density function  $f(s)$  defined on  $(\underline{s}, \bar{s})$  which we assume to be uniform. Let  $\Delta s = \bar{s} - \underline{s}$  be the difference between the extremes of the domain for  $s$ : A patient with severity  $s$  is assumed to obtain a benefit from a treatment defined by  $Q(s)$ ; ( $Q$  can be, for instance, a monetary value associated to the QALYs).

We consider a situation in which the social pressure on the Health Authority to reduce the excessive congestion in the public health service is severe. To do so, the Health Authority undertakes a policy in which the potential population of patients receives treatment within a given period of time. Note that this construction is equivalent to considering

---

<sup>6</sup>According to Barros and Olivella (1999), "full cream-skimming" is a situation in which all the mildest patients end up being treated in the private sector.

that only a fraction of patients is treated, and that this amount is exogenously given in our model and reflected by  $N$ .

For this purpose, the Health Authority contracts an agent, represented by a specialist, to treat a certain number of patients and reaches agreements with private hospitals to have the remaining patients treated there.<sup>7</sup> We denote by  $x$  the number of operations performed in the public sector; hence,  $N - x$  patients will be transferred to the private sector.

When treating patients, the Health Authority incurs in two different kinds of costs in the public sector: a transfer  $T$  to the physician it contracts and a constant cost of treatment  $k^{pb}$  per patient. We take the cost of the public treatment to be linear for the sake of expositional clarity. In Section 5, we provide some insights into the robustness of our results to other more general cost structures.

Moreover, in order to be consistent with real-life observations, we assume that the Health Authority makes a constant payment for each operation sent to the private sector. This payment covers both, the fee agreed with the private specialist ( $w$ ) and the cost of providing treatment to each patient in the private sector ( $k^{pv}$ ). The private cost of treatment is likely to be linear if the private sector is operating well under capacity.

With this construction, we allow for differences in the cost of providing treatment in both systems. For the sake of clarity, we define  $\Delta k = k^{pb} - k^{pv}$ , as the difference between the costs of treatment in either sector. Although, in principle, we impose no restrictions on the sign of  $\Delta k$ , it may be reasonable to consider the case of  $\Delta k > 0$  more likely to occur. This can be sustained on the grounds of dis-economies of scale or congestion problems in the public sector, as well as bureaucratic or administrative inefficiencies.<sup>8</sup>

We model the physicians' behavior 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,

---

<sup>7</sup>The private and public practice may even be done in the same hospital, under different types of contracts.

<sup>8</sup>In favor of this argument, data from Norway indicates that for some types of treatment the price charged by private hospitals is considerably lower than the costs in public hospitals. A further discussion of these cost differences is given in Hoel and Sæther (2000).



therefore, we must not only take his revenues and costs in the public sector into account, but those in his private practice as well.

In order to perform his tasks, the physician has to exert some effort ( $e^{pb}$  for the patients he treats in the public sector and  $e^{pv}$  for those treated in private). These levels of effort depend on the number of patients he treats in each sector ( $x$  and  $N - x$ , respectively) and on the average severity of the patients ( $b^{pb}$  and  $b^{pv}$ ). We define the effort as the product of these two components. The cost borne by the physician is also affected by a parameter  $\mu$ ; which measures the physician's skills or knowledge to be able to perform his tasks. We consider, however, that all the physicians share the same level of ability, which is common-knowledge among all the agents in the model. As such, the costs of the effort exerted by the physician in each sector are given by:

$$\begin{aligned} c^{pb} &= c(e^{pb}; \mu) = c(xb^{pb}; \mu) \\ c^{pv} &= c(e^{pv}; \mu) = c((N - x)b^{pv}; \mu): \end{aligned}$$

The function  $c(\cdot)$  is increasing and convex in the level of effort exerted and decreasing in the physician's ability. Moreover, we assume that  $c(0; \mu) = c_e(0; \mu) = 0$ ;  $\forall i = pb; pv$ :

This construction with separable efforts can yield situations in which the physician may have an incentive to distribute patients and severities between the two sectors, as this decreases his dis-utility by the total effort exerted. As we will show later, this possibility of cost-induced patient-selection is ruled out in our model by the effort function we have chosen. The fact that  $e^{pb} = xb^{pb}$  and  $e^{pv} = (N - x)b^{pv}$  ensures that when changes in the number of patients treated in a given sector and in their average severity, leave the cost to this sector unaltered, then this change does not alter the costs borne in the other sector either.

We can define the utility function of the physician, as follows:

$$U^s = T + w(N - x) - c(xb^{pb}; \mu) - c((N - x)b^{pv}; \mu). \quad (1)$$

The aim of this work is to study the potential strategic behavior that a physician may have, in his performance as a dual supplier. We, therefore, specifically ignore the possibility of the physician's strategically behaving within either sector, in the sense of exerting little effort (shirking). This is why we consider that the physician will exert the maximum level of effort that he considers compatible with his earnings, either by ethical commitment

or because he is fully monitored. Hence, he will treat patients in the public sector as long as his net revenues do not fall below his reservation value (normalized to zero in this model).<sup>9</sup>

The Health Authority's surplus derived from the care provided is given by the difference between the social net benefit of the treatment and the social cost that the production of the services generates. Hence, denoting by  $\mathbf{b}$  the average severity of the potential population of patients, i.e.,  $\mathbf{b} = \int_{\underline{s}}^{\overline{s}} s \frac{1}{\Phi_s} ds$ , the government's objective function is as follows:

$$H = Q\mathbf{b}N_i \left[ T + k^p x + (k^{pv} + w)(N_i - x) \right]$$

Or, by re-arranging terms we get:

$$H = (Q\mathbf{b} - k^{pv}) N_i [T + \Phi k x + w(N_i - x)] \quad (2)$$

Since all the patients eventually receive the treatment they need, maximizing this objective function is equivalent to minimizing the costs derived from undertaking the policy.

Note that we assume that the Health Authority does not take the utility function of the physician into consideration. In other words, the government is not maximizing a social welfare function.

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 (private physician) bargain over the value of the private fee  $w$  that the private physician will receive per operation.<sup>10</sup> At the first stage, the Health Authority contracts a specialist, specifying the salary he will receive ( $T$ ). At the second stage, the physician takes two simultaneous decisions: On the one hand, he selects the severities that he wants to treat in each sector. On the other hand, he decides on the number of operations he will perform in his public duty, and 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 specialist can neither control the demand for health care nor select the severities to be treated in

---

<sup>9</sup>In this model the physician treats all the patients he receives in the private sector. As such, there is no room for strategic behavior in the effort he exerts in his private practice.

<sup>10</sup>We show later that, in equilibrium, the bargaining set is not empty, i.e., the maximum wage the Health Authority is willing to pay exceeds the minimum the physician will accept for attending to public patients in his private practice.

either system. Stage 2 is therefore only partially 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 specialized physicians, the Health Authority cannot monitor the physician in his selection of the patients who are to receive treatment in either sector.

We start by analyzing the optimal policy in the ...rst setting.

### 3 BENCHMARK SCENARIO

In this section, we assume that the Health Authority can preclude the physician's selecting the patients he wants to treat in either sector. Hence, patients will be uniformly distributed between the public and the private sectors. We can thus ensure that the average severity of the operations is the same in both sectors, i.e.,  $\mathbf{b}^{pb} = \mathbf{b}^{pv} = \mathbf{b}$ .

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

**Assumption 1**  $0 < w_j \leq k < \alpha_x(N\mathbf{b}, \mu)$ :

Under Assumption 1, the difference between the private fee and the treatment-cost differential has to be positive and bounded above by a certain value. With this assumption we are only requiring that: On the one hand, if the private sector is less costly in terms of the treatment provided, we do not want this difference to be so high that it compensates the fee paid to the private physician. If this was the case, the public sector could purchase all the health services from the private sector instead of providing them, 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 not be so high that the policy of transferring patients is not undertaken, even in this framework where manipulation is not possible.

In order to characterize the solution in this framework, we solve the game by backwards induction. At the second stage, the physician chooses the amount of operations he will perform in the public sector. He will treat patients up to the point at which performing an additional operation would force him to make a loss. Therefore, the number of operations performed in the public sector,  $x$ , is such that  $\alpha_x(N\mathbf{b}, \mu) = T$ . We denote it by  $x(T)$ :

In the ...rst stage, the Health Authority maximizes its objective function. The opti-

mization program that the government faces is as follows:

$$\begin{aligned} \max_T H &= (Qb_i - k^{pv}) N_i [T + \Phi kx + w(N_i - x)] \\ \text{s.t.} \quad x &= x(T) : \end{aligned}$$

The following lemma characterizes the optimal sharing of patients between public and private practice and the salary that induces it.<sup>11</sup>

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

$$^a_x(x^a; \mathbf{b}, \mu) + k^{pb} = w + k^{pv};$$

with  $^a_x(x^a; \mathbf{b}, \mu) = T^a$ :

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.

The salary the Health Authority pays to the physician induces him to perform in his public practice the number of operations that equalizes the marginal costs of treating patients in both sectors.

It is straightforward to verify that the optimal level of patients treated in the public sector is increasing in the physician's ability and in the private fee  $w$ . We can also easily see the effect of the treatment-cost differential. If  $\Phi k > 0$ , the Health Authority incurs a smaller cost per operation in the private sector and, hence, is willing to transfer more patients for a given value of  $w$ . Moreover, since the fee per operation paid to the physician in the private sector is fixed (independent of the patient's severity), the optimal number of patients treated in the public sector is decreasing in the average severity of the population ( $\mathbf{b}$ ).

We proceed now to study the effects of dealing with a physician who can strategically choose the kind of patients to be treated in either sector. This will allow us to analyze the consequences of this potential strategic behavior on the willingness of the Health Authority to undertake the policy.

---

<sup>11</sup>The proof of this lemma is straightforward and therefore we omit it.

## 4 PATIENT SELECTION

Our concern in this section is to analyze whether the results differ when the physician has the ability to select patients and decide which cases to treat in his public practice and which go to the private sector. In other words, the Health Authority is not able to monitor the physician's choice of the severities of the patients treated in either sector, and this variable cannot be included in the terms of the contract.

As we have argued in the Introduction, this is an issue of great controversy in mixed health-care systems, as the diagnosis process that leads to the assessment of the severity of a given patient can only be performed by a trained physician. This means that the control over the severities of the patients who receive treatment in either sector is probably not possible for the Health Authority. Therefore, 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 again the game by backwards induction. At the second stage, the physician decides on the amount of operations he will perform in the public sector. His optimal number of operations does not depend merely on the salary he receives, since the average severity of the patients he treats  $\bar{b}^{pb}$  is also a variable of choice. Therefore,  $x$  is such that  $\int_0^1 x \bar{b}^{pb} \mu = T$  and we denote it by  $x^{i_T; \bar{b}^{pb}}$ .

The physician also decides which severities he wants to treat in either sector, subject to the restriction that  $x^{i_T; \bar{b}^{pb}}$  operations have to be performed in his public practice. Therefore, he does not have complete freedom in the choice of  $\bar{b}^{pb}$ ; as there may be values that are not compatible with the sharing of patients set. The physician will choose the value of  $\bar{b}^{pb}$  (and therefore also of  $\bar{b}^{pv}$ ) in order to maximize his total revenue. Since he is a dual supplier, he will consider the effects of his strategic behavior concerning his two sources of income. Therefore, the program he faces is:

$$\begin{aligned} \max_{\bar{b}^{pb}} U^S &= T + w(N - x) \int_0^1 (x \bar{b}^{pb}; \mu) \int_0^1 ((N - x) \bar{b}^{pv}; \mu) \\ \text{s.t. } x &= x^{i_T; \bar{b}^{pb}}; \end{aligned} \quad (3)$$

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

**Proposition 1** For a given sharing of patients between the two sectors ( $x$  and  $N - x$ ), the specialist will transfer the least severe cases to the private practice. Formally:

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

$$s \geq \frac{\mu}{\beta} + \frac{\phi s x}{N}; \quad \beta > 0$$

This proposition shows that the physician wants to treat only the mildest cases in the private practice, leaving the most difficult ones for the public sector. This behavior, 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 receives a salary whereas, in the private sector, his earnings are on a fixed fee-for-service basis. Therefore, the more operations the physician performs in his private practice, the higher the earnings he obtains. Furthermore, for a given level of effort exerted, (i.e., for a given cost), the “easier” the operations he performs the more patients he can treat. Note that we are not facing a situation of cost-induced patient-selection. Cream-skimming appears in our model not as an attempt by the physician to incur smaller costs, but rather as a way of increasing his earnings.

From Proposition 1, we obtain that the outcome of stage 2 consists of a pair  $(b^{pb}, x)$  that simultaneously fulfill the following conditions:

$$b^{pb} = \frac{\mu}{\beta} + \frac{\phi s x}{N}$$

$$x = x^T; \quad b^{pb} > 0$$

In our framework, cream-skimming cannot be avoided as the government is not able to monitor the physician choice of the severities to treat and, the contract, hence, cannot be contingent on the severities to be treated in the public sector. This set-up, however, is consistent with what one observes in most mixed health-care systems, where the physicians earn a salary that is not contingent on the number of patients they treat or their severities.

We shall now study how this problem of cream-skimming affects the decision of the Health Authority on when to undertake such a policy, and how to distribute the patients between the two sectors.

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

The maximization program of the Health Authority in this scenario is as follows:

$$\begin{aligned} \max_T & (Qb_i - k^{pv}) N_i [T + \Phi kx + w(N_i - x)] \\ \text{s.t.} & x = x^i_T; b^{pb} \\ & : b^{pb} = \beta_i \frac{\Phi s x}{2N} \end{aligned} \quad (4)$$

The following lemma provides the interior candidate for solution in this framework. Note that when the patient-selection arises, the Health Authority's objective function is much more complicated. In particular, it is no longer true that the function is always concave. The first order necessary condition for optimality is, therefore, not sufficient in general.<sup>12</sup>

**Lemma 2** With patient-selection, in the interior candidate for solution, the number of patients treated in the public sector ( $x^m$ ) and the salary the physician receives ( $T^m$ ) are such that:

$$a_x^i x^m b^{pb}; \mu^\Phi + k^{pb} i \frac{\Phi S}{2N} a_{b^{pb}}^i x^m b^{pb}; \mu^\Phi = w + k^{pv};$$

with  $a_x^i x^m b^{pb}; \mu^\Phi = T^m$  and  $b^{pb} = \beta_i \frac{\Phi s x^m}{2N}$ :

With patient-selection, new effects appear in the Health Authority's first order condition which will determine the optimal policy. The cost per operation in the public sector has now increased, as the average severity of the patients treated there is higher. This makes the treatment provided by the public sector more expensive and, hence, leads to a greater transfer of patients to private practice. An additional effect, which goes in the opposite direction, however, appears:  $\frac{\Phi S}{2N} a_{b^{pb}}^i x^m b^{pb}; \mu^\Phi$  reflects how an increase in the number of operations in the public sector has a positive impact on the public costs (through the decrease it induces in the average severity of the patients treated there).

To be able to close the model and perform comparisons with the benchmark case, we need to consider a specific cost of effort function. This will allow us to characterize the effects of the cream-skimming phenomenon on the behavior of the government and on its willingness to undertake the policy.

---

<sup>12</sup>Moreover, we need to impose a regularity condition in the cost of effort function to ensure that the f.o.c is well defined. If the total effect of a change in  $x$  on the cost of effort is positive, i.e.  $\frac{d^a_x(x b^{pb}; \mu)}{dx} = a_x^i x b^{pb}; \mu^\Phi + \frac{\Phi S}{2N} a_{b^{pb}}^i x b^{pb}; \mu^\Phi > 0$  for every  $x \in [0; N]$ ; an interior candidate for optimum exists. Under this condition, the proof of the lemma is straightforward and is therefore omitted.

Hence, hereinafter, the dis-utility of the physician's efforts in either sector is given by:

$$\begin{aligned}
 a^{pb} &= a(x, \mathbf{b}^{pb}; \mu) = \frac{1}{2} \frac{i}{\mu} x \mathbf{b}^{pb} \zeta_2 \\
 a^{pv} &= a((N - i) x, \mathbf{b}^{pv}; \mu) = \frac{1}{2} \frac{i}{\mu} (N - i) x \mathbf{b}^{pv} \zeta_2.
 \end{aligned}$$

Before proceeding to analyze the optimal response from the Health Authority, we need to study the curvature of its objective function. Since the amount of patients that receive public treatment is determined by the salary the physician receives, it is equivalent to the Health Authority's deciding directly on the salary or on the number of patients to be treated. For the sake of notational clarity, let  $d = \frac{\zeta_2 s}{\mathbf{b}}$  denote the relative dispersion of the patients' severities. The following lemma characterizes the curvature as a function of  $x$ .

**Lemma 3** The curvature of the Health Authority's objective function, under patient-selection ( $H^m$ ), is as follows:

1. If  $d \leq 4 \frac{i}{2} \frac{\rho_2}{3}$ ; then  $H^m$  is always concave.
2. If  $d > 4 \frac{i}{2} \frac{\rho_2}{3}$ ; then:
  - (a) For any  $x \in [0, \frac{1}{d} N]$ ;  $H^m$  is concave at  $x$ .
  - (b) For any  $x \in [\frac{1}{d} N, N]$ ;  $H^m$  is convex at  $x$ .

$$\text{With } \frac{1}{d} = \frac{i}{\mathbf{b}} + \frac{1}{2} \frac{\zeta_2 \rho_2}{3}, \quad \frac{1}{d} < 0.$$

Lemma 3 allows us to study the curvature of the population's health function, in terms of the relative dispersion of the patients' severities, measured as the ratio of the difference between the boundary severities ( $\mathbf{s}$  and  $\underline{\mathbf{s}}$ ) and the average severity ( $\mathbf{b}$ ). This lemma highlights the relevance of the relative dispersion, as it is a measure of how serious the problem of cream-skimming is. The strategic behavior of the physician (in transferring the mildest cases to the private practice) is fostered by the wide range of severities, since his gains in diverting patients are higher. We show that when the relative dispersion of the severities is low, the objective function is still concave in the entire domain. Thus, when the Health Authority does not suffer much from the problem of patient-selection, the program has a unique candidate to optimum. If the severities are sufficiently dispersed, however, the objective function has a convex section, that is bigger the higher the dispersion is. 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 government's maximization problem.

**Proposition 2** In the presence of patient-selection, the Health Authority decides, through the choice of the salary, to undertake the policy of transferring patients to private practice if the value of the private fee is below a certain threshold. The higher the relative dispersion of the patients' severities is, the more demanding this condition is. Formally:

1.  $T^m$  is such that  $x^m < N$  if  $w_i - c_k < \frac{b^2 N}{\mu^2} G(d)$ :

2.  $T^m = \frac{1}{2} \frac{b^2 N}{\mu^2}$  is such that  $x^m = N$ , otherwise.

Where  $G(d) = \begin{cases} 1 - \frac{d}{2} & \text{if } d \leq \frac{4}{3} \\ g(d) & \text{if } d > \frac{4}{3} \end{cases}$  is a continuous function and  $g(d)$  is such that  $g'(d) < 0$  and  $\lim_{d \rightarrow 1} g(d) = \frac{1}{2}$ :

This result 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. It is straightforward to verify that, for this particular effort cost function that we have considered here, in the absence of patient-selection, the decision of the Health Authority will be to distribute patients between the two sectors as long as  $w_i - c_k < \frac{b^2 N}{\mu^2} = \bar{w}$  (and, by Assumption 1, this condition always holds). If patient-selection by the physician cannot be avoided, this condition is more demanding. For the government to undertake this policy of distribution of patients between sectors, the upper bound of the private fee is now smaller ( $\frac{b^2 N}{\mu^2} G(d) < \frac{b^2 N}{\mu^2}$  since  $G(d)$  is always less than one).

The reason for this result is that by choosing to pay a salary  $T^m = \frac{1}{2} \frac{b^2 N}{\mu^2}$ , the Health Authority is completely eliminating the possibility of patient-selection. Inducing the physician (through the salary it pays to him) to treat all the patients in the public sector, there is no chance of avoiding the mildest cases. When the private fee is sufficiently low, however, the Health Authority decides to suffer the "cream-skimming" problem, in order to bear a lower cost for the patients transferred.

Moreover, the threshold of the private fee from which the Health Authority is willing to carry out the policy, is decreasing in the relative dispersion of the patients' severities.

When the relative dispersion of the severities is low, the Health Authority does not suffer the problem of cream-skimming very much (since all the 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  $d \neq 0$ , the condition for undertaking the policy converges to the one required in Assumption 1. However, as the relative dispersion of the severities increases, the condition necessary for the public 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 policy is not undertaken unless the private fee is sufficiently low.

We now compare the interior solution in this setting to the optimal one in the benchmark case.

**Proposition 3** 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 induces a higher transfer of patients to private practice, provided that the private fee is below a certain value. Otherwise, there is a lower transfer.

ii).- If the relative dispersion of the severities exceeds a critical value, the Health Authority always induces a higher transfer of patients to private practice.

Formally:

$$1. x^m < x^a; \text{ when } d \geq 0.7625 \text{ and when } d < 0.7625 \text{ and } w_j - k < \frac{b^2 N}{\mu^2} \odot (d) :$$

$$2. x^m > x^a; \text{ when } d < 0.7625 \text{ and } w_j - k > \frac{b^2 N}{\mu^2} \odot (d).$$

$$\text{With } \odot (d) = \frac{1}{4d} (3d + 6) \left( \frac{p}{d^2 + 4d + 36} \right)^{\alpha} :$$

This proposition shows that when the policy of transferring patients is undertaken, the consequences of the cream-skimming on the number of patients transferred differ in the relative dispersion of the severities. This result is a consequence of two contrary effects. On the one hand, the marginal cost of treating an extra patient in the public sector is higher with patient-selection, as the average severity of the patients treated is higher. On the other hand, by increasing the number of operations in the public sector we not only save the private fee ( $w$ ), but also reduce the possibility of cream-skimming and, hence, decrease the expected level of severity that the Health Authority faces. When the

relative dispersion is sufficiently high, the negative effect of treating patients in the public sector always dominates and, thus, fewer patients are kept in the public sector, even if we already know that this increases the capacity of the physicians to select patients. In contrast, when the relative dispersion is low, the final result is determined by the value of the private fee. A high level of  $w$  implied a relatively low transfer of patients in the benchmark scenario; we show that, in this case, the presence of cream-skimming leads to an even smaller transfer. Conversely, when the first best situation was to transfer a high proportion of the patients (low  $w$ ) to private practice, the distortion implies to transferring even more. Figure 1 outlines all of these possibilities.

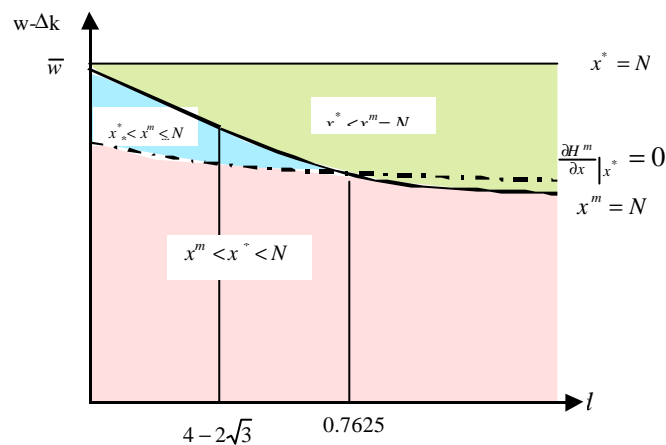


Figure 1: Comparison of the solutions with patient selection ( $x^m$ ) and without it ( $x^*$ ).

Proposition 3 can also be interpreted in terms of the salary paid to the physician. In this dimension, however, it is more difficult to obtain the sign of the distortion. The reason for this is that, when the presence of cream-skimming induces a higher transfer of patients to private practice, two effects come into conflict: First, fewer patients are treated in the public sector, which implies lower costs for the physician and a smaller salary, and secondly, the patients who are treated are relatively more severe (hence, they induce a higher salary to cover the physician's effort cost). As a result, the impact of patient-selection on the salary the physician receives is ambiguous in this region. Nevertheless, when the relative dispersion of the severities is low, and the private fee sufficiently high, there is no such ambiguity. In equilibrium, fewer patients are transferred to the private

sector. Moreover, due to patient-selection, the patients who are left in the public sector are the most severe ones. In this region, therefore, the existence of cream-skimming induces the Health Authority to pay a higher salary to the physician.

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 under patient-selection. This illustration is made for the case in which the policy is implemented and the relative dispersion of the severities is sufficiently high ( $d > 0.7625$ ).

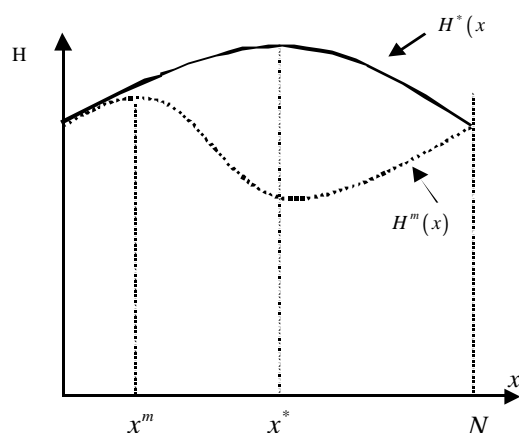


Figure 2: Health Authority's objective functions when  $d > 0.7625$ :

We have considered the negotiations between the government and private hospitals, about the value of the fee to be paid per operation performed in the private sector, 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.  $w \geq k < \frac{b^2 N}{\mu^2} G(d)$  and, at the same time, are acceptable to the physician  $w \geq \frac{1}{2} \frac{N - x}{\mu} b^{pv} c^2$ . The result is presented in the following remark:

**Remark 1** In equilibrium, and for  $k \geq 0$ ; the bargaining set is not empty (for every value of  $d$ ), i.e., the maximum wage the Health Authority is willing to pay exceeds the minimum that the physician requires to accept public patients in his private practice.

This remark shows that if the private sector is not more inefficient than the public sector is, we can ensure that there are values of  $w$  that are both physician and government

compatible, independently of the relative dispersion of the severities. If  $\Phi k < 0$ , there also exists room for negotiation, provided the relative dispersion of the severities is not too high. Figure 3 illustrates the bargaining set between the Health Authority and the physician for  $\Phi k \geq 0$ :

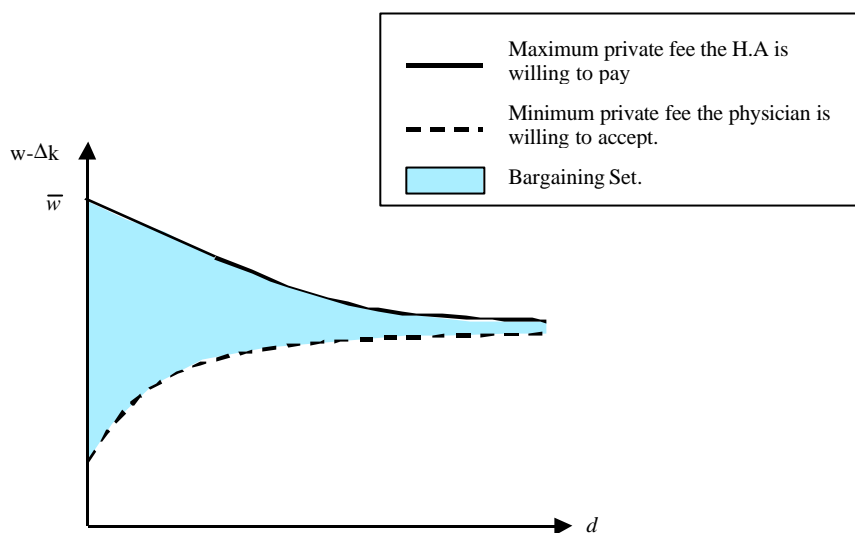


Figure 3: Bargaining Set for values of  $\Phi k \geq 0$ :

## 5 COMMENTS AND EXTENSIONS

In this paper we have shown that when physicians are dual providers, a problem of cream-skimming may arise. This strategic behavior makes the government less willing to undertake a policy of transferring some of the public sector's patients to private hospitals. When the policy is undertaken, however, in most of the domain of the variables, more patients are eventually treated in private practice than in the absence of cream-skimming.

In this section, we discuss some of the ingredients of our model by proposing alternative constructions and providing insights into their impact on the results. In particular, we introduce modifications that affect the structure of the costs of treatment in the public sector, and allow for heterogeneous physicians in the model.

### 5.1 Cost of Public Treatment

Our assumption of a constant marginal cost of treatment in the public sector, was made for the sake of clarity in the presentation. The construction we have chosen allows us to

concentrate on the physician's incentives in selecting patients and to fully characterize the Health Authority's response to such behavior.

Nevertheless, other more general structures can be considered for the costs. The alternatives that we have in mind are: (1) Marginal cost of treatment increasing with the average severity of the patients; (2) Marginal costs increasing in the number of treatments provided (dis-economies of scale); and (3) Capacity constraints in the public sector.

In all these set-ups, the incentives of the physician to select patients remain unaltered and, hence, the same problem of cream-skimming arises. Moreover, the crucial measure for assessing the seriousness of the problem continues to be the relative dispersion of the severities. Even if the government's reaction is not qualitatively altered, there are quantitative differences in the results arrived at under these different cost structures.

The first two alternatives considered (marginal costs increasing with the severity or in the number of operations) alter the curvature of the Health Authority's objective function. Even if the curvature cannot be fully characterized now, the effect of the change in the cost structure on the results is clear. When the marginal costs increase with the severity of the patients, the presence of cream-skimming fosters the non-concavity of the function, as it increases the average severity in the public sector. Consequently, the boundary solution (not undertaking the policy) is more likely to prevail. On the contrary, when the costs are convex in the number of operations, it results in an impulse of the concavity of the program and, hence, in a reduction of the region where the boundary solution is optimal.

To analyze the situation where public costs are affected by capacity constraints, we consider the simplest structure, i.e., costs that are linear in the number of patients, but with the cost parameter increasing when the number of operations exceed a certain threshold. Formally, there exists a level of operations  $\bar{x} < N$  such that if  $x < \bar{x}$ ; the costs are  $k_1^{pb}x$ ; while, if  $x > \bar{x}$ , total costs are  $k_1^{pb}\bar{x} + k_2^{pb}(x - \bar{x})$ ; with  $k_2^{pb} > k_1^{pb}$ : In this set-up, the analysis of the curvature of the objective function is not different from that of our original model. The only difference is that, under this new cost structure, the boundary solution becomes more costly, since it implies bearing the cost of a higher extra capacity. The condition for the Health Authority to be willing to carry out the policy is, therefore, less demanding here than it was in the original model.

## 5.2 Heterogeneous Physicians

Our analysis assumes that all physicians have the same level of ability and, therefore, this level is observable by the Health Authority (or, what is equivalent, to considering a single physician).

There are two reasons for choosing such a set-up: First, we wanted to focus on the potential strategic behavior that a physician may have in his performance as a dual supplier. We, therefore, specifically ignored the possibility of the physician to strategically behave within each sector, in the sense of his exerting a low level of effort or shirking off. Dealing with homogeneous physicians allows us to avoid the additional problems derived from the physician's having a double strategic behavior.

Secondly, we wanted to study physicians' incentives to select patients under a mixed contractual structure that is close to the one we observe in many mixed health care systems. Under these schemes, the earnings of the physicians in either sector are the same, independently of their type.

There is a general consensus in the literature, however, concerning the problems that the different degrees of information of the agents involved in the provision of health care generate (for an overview of principal-agent theory, see LaPortont and Tirole (1993)). To be more specific, the providers of medical services generally have more information than the government does, concerning their own skills, i.e., their capacity to treat the patients at a given cost of effort. The important question here, therefore, is whether our results are robust in the existence of heterogeneous physicians. This would introduce a problem of adverse selection in the model, in addition to the one of patient-selection.

Let us now consider two types of physicians: a "high skilled" physician ( $\mu = \mu_h$ ) and a "low skilled" one ( $\mu = \mu_l$ ). Earning a mixed salary is not incentive compatible for the "high skilled" physician, since he can pretend to be "low skilled" and reduce the dis-utility of his effort.

To solve this problem and concentrate on patient-selection, a truthful-revelation contract should be designed. A contract that ensures the honest revelation by the doctor should include two components: a mixed salary and a bonus. The Health Authority would offer a contract  $f(T_h; B_h) ; (T_l; B_l)$  to the physician, under which he receives a salary  $T_h$ , plus a bonus  $B_h$ , if he announces  $\mu_h$  as his level of ability, and analogously for  $(T_l; B_l)$ . The bonus is included to induce a high skilled physician to sign his contract. In equilibrium,

then, it is strictly positive for the high type (informational rent) and zero for the low type.

Under this new sort of contract, however, (and maintaining the same reimbursement scheme in the private sector), the incentives for the physician to select patients remain unaltered. Only the high skilled physician receives a bonus, but this does not depend on his decision regarding the selection of patients. The magnitude of his informational rents is only affected by the decision taken by the low skilled physician. The problem of cream-skimming is, therefore, the same under adverse selection.

The reaction of the Health Authority, however, differs. When the government does not observe the physician's ability, a new effect appears: its incentives to undertake the policy of distributing patients between the two sectors increase, even in the presence of cream-skimming. This effect is due to the fact that the physician's informational advantage allows him to earn informational rents, which can be reduced by transferring more patients to the private sector. Even if sending less patients to private practice alleviates the problem of cream-skimming, it raises the value of the bonus received by the high skilled physician and is, therefore, more costly. As a consequence of this, there is a range of parameter values (in particular, private fees) for which the policy of distributing patients between the private and the public sectors is only optimal when the Health Authority suffers from asymmetric information about the physician's ability. The Health Authority is more willing to undertake the policy, as it is a way of relaxing the agency problem suffered and "transferring" the informational disadvantage to the private sector.



## 6 CONCLUDING REMARKS AND POLICY IMPLICATIONS

The motivation behind this study was the enormous congestion that exists in public health services worldwide, which has forced several Health Authorities to devise especial programs to alleviate it. Temporary programs, however, may be extremely costly and, hence, the study of the regularity and adequacy of the measures undertaken becomes a major concern.

We have analyzed a situation in which the Health Authority undertakes a policy to reduce the proportion of patients that remain untreated, transferring a fraction of such cases to private hospitals. We have shown that a problem of cream-skimming arises. Due to the different structure of the physician's remuneration in either system, specialists prefer to treat only the mildest cases in their own private practices. We have, then, shown how this problem makes the Health Authority be more reluctant to implement this policy.

We have also characterized the range of parameters that makes the Health Authority find it more profitable to treat all of its patients in the public sector. Moreover, we found that the crucial variable that measures the importance of the problem the Health Authority faces is the relative dispersion of the patients' severities.

When the policy is undertaken, we study the effects of patient-selection on the amount of patients that are finally transferred to the private sector. We also find that the results are ambiguous. When the relative dispersion of the severities is high enough, the negative effect of treating patients with higher average severities in the public sector always dominates. More patients are, therefore, sent to private hospitals. When we deal with medical disciplines in which the dispersion of the severities is low, the result is determined by the value of the private fee. In this case, the behavior of the Health Authority is always more extreme than in the absence of cream-skimming.

Our analysis provides some policy recommendations concerning the optimality of this kind of measures. When designing a policy to transfer patients from the public to the private sector, the policy-maker should consider the fact that the difference that exists between the reimbursement systems of the two sectors can create perverse incentives for the physicians. To be more specific, if the reimbursement rule in the public sector is a fixed salary, while in the private practice functions on a fee-for-service basis, as is the case

in several countries where mixed health-care systems exist, a problem of cream-skimming arises. The physicians transfer the mildest cases to the private practice, thus increasing the cost per operation borne by the public sector.

Our results suggest that the decision to undertake the policy or not is also influenced by another important issue, apart from the fee per operation agreed with the private sector. The type of illness and, in particular, how dispersed the severities of the patients are 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 likely it is that the policy will be undertaken.

Moreover, empirical evidence shows that when a policy of sending some patients to private hospitals is implemented, the number of operations that are eventually performed in the public sector slightly decreases. These empirical findings are consistent with our results. In our model, this is because the patients who remain in the public sector are more costly. In equilibrium, except for a small region of the parameter values, the implementation of the policy generates a reduction in the number of operations performed in the public sector. In the region where this does not hold, the Health Authority reacts to the selection of patients by paying the physician a higher salary.

We have chosen to work with a uniform distribution of patients' severities for analytical convenience, even though a distribution in which the severities are more concentrated around the mean might seem to be more realistic. However, we guess that this would not alter the main features of the model. There would be incentives for patient-selection, and the other results should not be qualitatively altered.

Finally, it is worth noting that all of the analysis we have performed, are done under the assumption that there is no difference between the quality of the services provided by the public and private sectors. This assumption is reasonable for the kind of treatments we consider (non-urgent elective surgery). In fact, at least in Spain, the patients who demand elective surgery are those who are transferred from the public to the private sector. This assumption would be clearly inappropriate for modelling other kinds of illnesses that require very expensive or high-tech treatment. To include these medical disciplines in the analysis would require a model with heterogeneous quality among the different providers.

# Appendix

Proof of Proposition 1:

At stage 2, the reduced form of the optimization program the physician faces is as follows:

$$\begin{aligned} \max_{b^{pb}} U^s &= T + w(N - x) i^a(x; b^{pb}; \mu) i^a((N - x); b^{pv}; \mu) \\ \text{s.t. } x &= x^i(T; b^{pb}; \mu) \end{aligned}$$

The derivative is given by:

$$\frac{\partial U^s}{\partial b^{pb}} = i^w \frac{dx}{db^{pb}} i^{\frac{d^a i^a(x; b^{pb}; \mu)}{db^{pb}}} i^{\frac{d^a i^a((N - x); b^{pv}; \mu)}{db^{pb}}}; \quad (5)$$

By stage 3 we know that  $x = x^i(T; b^{pb}; \mu)$  is such that  $i^a(x; b^{pb}; \mu) = T$ , which implies that  $i^a(x; b^{pb}; \mu)$  is constant with respect to  $b^{pb}$ .

Since  $s \subseteq U(s; \mathcal{S})$ ,  $b^{pb}$  and  $b^{pv}$  are related by the following equation:

$$b^{pb} \frac{x}{N} + b^{pv} \frac{(N - x)}{N} = b$$

From the above expression we get:

$$b^{pv} (N - x) = bN - b^{pb}x$$

Since  $b^{pb}x$  is constant with changes in  $b^{pb}$ ,  $b^{pv} (N - x)$  is also constant in this variable.

Hence:

$$\frac{d^a i^a(N - x; b^{pv}; \mu)}{db^{pb}} = 0$$

Therefore, condition (5) is reduced to:

$$\frac{\partial U^s}{\partial b^{pb}} = i^w \frac{dx}{db^{pb}};$$

and from here it is straightforward that:

$$\frac{\partial U^s}{\partial b^{pb}} > 0; \quad \partial b^{pb}:$$

The physician would like to treat the patients with the highest average severity in the public sector.

We can write  $b^{pb}$ , using the formula for the conditional expectation. We let  $C$  be the subset of severities treated in the public sector (for a given number of operations  $x$ ). Formally:

$$b^{pb} = E(s|C) = \frac{1}{Pr(C)} \int_C s \phi_s ds$$

For a given value of  $Pr(C) = \frac{x}{N}$  this expectation is increasing in the location of  $C$  in  $(\underline{s}; \bar{s})$ . Thus,  $b^{pb}$  is maximum when  $C$  is maximized in the interval  $(\underline{s}; \bar{s})$ , from which we get that  $C = [\underline{s}; \bar{s} - \frac{\phi_{sx}}{N}]$ . Therefore, the physician chooses to treat the patients with severities in the range  $[\underline{s}; \bar{s} - \frac{\phi_{sx}}{N}]$  in the private sector and leaves those in the interval  $[\bar{s} - \frac{\phi_{sx}}{N}; \bar{s}]$  in the public sector: This implies that, in equilibrium,  $b^{pb}$  and  $x$  are such that:

$$\begin{aligned} b^{pb} &= \bar{s} - \frac{\phi_{sx}}{2N} \\ x &= x^i_T; b^{pb} \end{aligned}$$

This completes the proof.  $\square$

Proof of Lemma 3:

We study the curvature of the Health Authority's objective function in the restricted domain given by the behavior of the physician at stage 2. From this stage we know that  $x$  is such that  $\frac{1}{2} \frac{x}{\mu} b^{pb} = T$  and that  $b^{pb} = \bar{s} - \frac{\phi_{sx}}{2N}$ : By substituting these two constraints in the program, we will characterize the curvature of the objective function as a function of  $x$ : The optimization program at the first stage is as follows:

$$\max_x H^m = (Qb_i - k^{pv}) N - \frac{1}{2} \frac{x}{\mu} \bar{s} - \frac{\phi_{sx}}{2N} + \phi_k x + w(N - x) \quad \#$$

The f.o.c is given by:

$$\frac{\partial H^m}{\partial x} = \frac{1}{\mu} \bar{s} - \frac{\phi_{sx}}{2N} + \frac{\phi_{sx}^2}{2N\mu^2} + \phi_k - w = 0$$

Or:

$$\frac{x^2 \bar{s}}{2N\mu^2} b^{pb} + w - \phi_k = \frac{x}{\mu^2} \bar{s} b^{pb} \quad !$$

Computing the second order condition and rearranging terms yields:

$$\frac{\partial^2 H^m}{\partial x^2} = \frac{1}{\mu^2} \bar{s}^2 + 3\phi_s \bar{s} - \frac{3}{2} \phi_s^2 \bar{s}^{-2} \quad ; \quad \bar{s} = \frac{x}{N}$$

This derivative is increasing in  $\mu$ : We compute its roots and we ...nd:

$$\mu_1 = \frac{4}{\phi_s} \frac{3 + \rho_3}{3} > 1 \text{ and } \mu_2 = \frac{4}{\phi_s} \frac{3 - \rho_3}{3}: \text{ Simple manipulations allow us to re-write } \mu_2 \text{ as } \mu_2 = \frac{1}{d} + \frac{1}{2} \frac{3 - \rho_3}{3} = \mu^*(d): \text{ It is straightforward to see that } \mu^*(d) < 0.$$

From here, and taking into account that  $\mu > 1$  since  $x \in N$ ; it can be shown that:

- i) If  $\mu^*(d) > 1$ ;  $\frac{\partial^2 H^m}{\partial x^2} < 0 \forall x \in [0; N]$ , i.e., the objective function is always concave.
- ii) If  $\mu^*(d) < 1$ ;  $\frac{\partial^2 H^m}{\partial x^2} < 0$  for every  $x \in [0; \mu^*(d)N]$ . Thus, we can ensure that for  $x \in [0; \mu^*(d)N]$  the objective function is concave, and for  $x \in [\mu^*(d)N; N]$  it is convex.

In re-writing the conditions for  $\mu^*(d)$  in terms of the relative dispersion of the severities, we ...nd that:

$$\mu^*(d) > 1, \quad d \leq 4 \frac{1}{2} \frac{\rho_3}{3} \text{ and}$$

$$\mu^*(d) < 1, \quad d > 4 \frac{1}{2} \frac{\rho_3}{3}.$$

This completes the proof.  $\square$

### Proof of Proposition 2:

From the f.o.c computed in Lemma 3 it is straightforward to verify that  $x = 0$  can be never a solution, since:

$$\frac{\partial H^m}{\partial x} \Big|_{x=0} = w_i - \phi k > 0:$$

This, together with the other conditions found in Lemma 3, provides a complete characterization of the optimization problem:

1.-If  $d \leq 4 \frac{1}{2} \frac{\rho_3}{3}$ , there exists a unique candidate to optimum ( $x$ ). This solution will be interior, i.e.,  $x \in (0; N)$ ; if and only if:

$$\frac{\partial H^m}{\partial x} \Big|_{x=N} = w_i - \phi k_i \frac{N}{\mu^2} \frac{\mu}{b_i} \frac{\phi_s}{2} < 0, \quad w_i - \phi k < \frac{b^2 N}{\mu^2} \frac{\mu}{1} \frac{d}{2}:$$

2.-If  $d > 4 \frac{1}{2} \frac{\rho_3}{3}$ , there exists, at most, a single interior candidate to optimum ( $x$ ), such that  $x \in [0; \mu^*(d)N]$ , with  $\mu^*(d) = \frac{1}{d} + \frac{1}{2} \frac{3 - \rho_3}{3}$ . The boundary solution  $x = N$  is also a potential candidate to optimum, provided that:

$$\frac{\partial H^m}{\partial x} \Big|_{x=N} > 0, \quad w_i - \phi k > \frac{b^2 N}{\mu^2} \frac{\mu}{1} \frac{d}{2}$$

This is a necessary, although not sufficient condition, for  $x = N$  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 on a fraction  $\theta$  of the patients, rather than perform  $N$  operations. We find that:

$$\text{If } w_i \Phi k < \frac{b^2 N}{\mu^2} g(d); \text{ with } g(d) = \max_{\theta \in (0;1)} \frac{1}{2} (1 + \theta) \theta^{2d} \frac{\mu (1 + \theta)}{4} d + 1$$

$$g'(d) < 0 \text{ such that } H^m(x = \theta N) > H^m(x = N):$$

We can see that  $g(d)$  is such that  $g'(d) < 0$ ;  $\lim_{d \rightarrow 1} g(d) = \frac{1}{2}$ .

If  $w_i \Phi k$  exceeds the above threshold, then we can ensure that the boundary solution is optimal, since for this parameter configuration  $\frac{\partial H^m}{\partial x} |_{x=N} > 0$ : Therefore, the solution to the principal's problem is:

$$i) x^m < N \text{ if } w_i \Phi k < \frac{b^2 N}{\mu^2} G(d);$$

$$\text{where } G(d) = \begin{cases} < 1 + \frac{d}{2} \text{ if } d < 4 + \frac{2}{\mu^2} \\ g(d) \text{ if } d \geq 4 + \frac{2}{\mu^2} \end{cases}$$

is a continuous function and  $g(d)$  is such that

$$g'(d) < 0 \text{ and } \lim_{d \rightarrow 1} g(d) = \frac{1}{2};$$

$$ii) x^m = N; \text{ otherwise.}$$

And this completes the proof.  $\square$

### Proof of Proposition 3:

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

$$w_i \Phi k < \frac{b^2 N}{\mu^2} G(d):$$

The f.o.c of the Health Authority's problem in the presence of patient-selection was given by:

$$\frac{\partial H^m}{\partial x} = \frac{\tilde{A}}{\mu^2} \theta \left[ \frac{\Phi s x}{2N} + \frac{\Phi s x^2}{2N\mu^2} \right] + w_i \Phi k = 0$$

Considering that  $\theta = b + \frac{1}{2}\Phi s$ , we can rewrite it in terms of  $b$  as:

$$\frac{\tilde{A}}{\mu^2} \left( b + \frac{1}{2}\Phi s \right) \left[ \frac{x}{N} \frac{\Phi s}{2} + \frac{\Phi s x^2}{2N\mu^2} \right] + w_i \Phi k = 0:$$

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

$$\frac{\partial H}{\partial x} = \frac{\mu}{\mu^2} \frac{x b^2}{\mu^2} + w_i \Phi k = 0$$

Performing some algebraic manipulations, we find that:

$$\frac{\partial H^m}{\partial x} > \frac{\partial H}{\partial x}, \quad \frac{x}{\mu^2} \left( \frac{b^2 N}{N} \frac{d}{2} + \frac{3x}{2N} \mu \left( 1 + \frac{d}{2} \right) \mu \left( 1 + \frac{d}{4} \right) \right) > 0$$

From the inequality above we find that:

$$\frac{\partial H^m}{\partial x} \Big|_{x=x^a} > 0 \quad ( )$$

$$w_i \in \mathbb{R} > \frac{b^2 N}{\mu^2} \frac{1}{4d} \left( 3d + 6 \right) \mu \frac{1}{d^2 + 4d + 36} :$$

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 2, we find that:

1) When  $d < 7625$ :

If  $w_i \in \mathbb{R} < \frac{b^2 N}{\mu^2} \odot(d)$ ; then:

$$x^m < x^a;$$

If  $w_i \in \mathbb{R} > \frac{b^2 N}{\mu^2} \odot(d)$ ; then:

$$x^m > x^a;$$

With  $\odot(d) = \frac{b^2 N}{\mu^2} \frac{1}{4d} \left( 3d + 6 \right) \mu \frac{1}{d^2 + 4d + 36} :$

2) When  $d \geq 7625$ :

$$x^m < x^a;$$

And this completes the proof.  $\square$

## References

- [1] Arrow, K.J. (1963) "Uncertainty and the Welfare Economics of Medical Care". *American Economic Review* 53, 941-973.
- [2] Barros, P.P. and Olivella, P. (1999). "Waiting lists and Patient Selection". D.P. UAB-IAE 499-99.
- [3] Blomqvist, A. (1991) "The Doctor as a Double Agent: Information Asymmetry, Health Insurance, and Medical Care". *Journal of Health Economics* 10, 411-432.
- [4] Bosanquet, N. (1999) "A Successful National Health Service". Adam Smith Institute. London.
- [5] Competition Commission (1994) "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
- [6] Dranove, D. (1987) "Rate-setting by Diagnosis Related Groups and Hospital Specialization". *RAND Journal of Economics* 18-3, 417-427.
- [7] Ellis, R.P. (1998) "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins". *Journal of Health Economics* 17, 537-555.
- [8] Ellis, R.P. and McGuire, T.G. (1986) "Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply". *Journal of Health Economics* 5, 129-152.
- [9] Ellis, R.P. and McGuire, T.G. (1990) "Optimal Payment Systems for Health Services". *Journal of Health Economics* 9, 375-396.
- [10] Gaynor, M. (1994) "Issues in the Industrial Organization of the Market for Physician Services". *Journal of Economics and Management Strategy* 3-1, 211-255.
- [11] Hoel, M. and Sæther, E.M. (2000) "Private Health Care as a Supplement to a Public Health System with Waiting Time for Treatment". Memorandum from the Department of Economics, no 41/2000. University of Oslo.
- [12] Iversen, T. (1997) "The Effect of a Private Sector on the Waiting Time of a National Health Service". *Journal of Health Economics* 16, 381-396.



- [13] Laffont, J.J. and Tirole, J. (1993) *A Theory of Incentives in Procurement and Regulation*. Cambridge and London: MIT Press.
- [14] Ma, C.-t.A. (1994) "Health Care Payment Systems: Costs and Quality Incentives". *Journal of Economics and Management Strategy* 3-1, 93-112.
- [15] Olivella, P. (2001) "Shifting Public-Health-Sector Waiting Lists to the Private Sector". Forthcoming: *European Journal of Political Economy*.
- [16] Propper, C. (1995) "Agency and Incentives in the NHS Internal Market". *Social Science and Medicine* 40, 1683-1690.
- [17] Rickman, N. and McGuire, A. (1999) "Regulating Providers' Reimbursement in a Mixed Market for Health Care". *Scottish Journal of Political Economy* 46-1, 53-71.
- [18] Selden, T.P. (1990) "A Model of Capitation". *Journal of Health Economics* 9, 397-410.
- [19] Yates, J. (1995) "Private Eye, Heart and Hip: Surgical Consultants, the National Health Service and Private Medicine". London: Churchill Livingstone.