UCD GEARY INSTITUTE FOR PUBLIC POLICY
DISCUSSION PAPER SERIES

Private Practice in Public Hospitals: Should Senior Consultants Be Prioritized?

Xidong GUO

Dr. Sarah PARLANE

School of Economics, University College Dublin.

Geary WP2023/01
Jan 05, 2023

Private Practice in Public Hospitals:

Should Senior Consultants Be Prioritized?

Xidong GUO

Dr. Sarah PARLANE

School of Economics, University College Dublin.

Abstract: This paper proposes a normative analysis which investigates the optimal management of private practices within a public hospital. Private income supplementation induces consultants to attend to more patients which reduces waiting lists and the public cost of healthcare. It is however optimal to cap the consultants' private income, regardless of seniority. When first-degree discrimination is possible, the more productive (senior) consultants receive a higher private income than their junior counterparts when priority is given to shortening waiting lists. However, they must charge a lower fee when priority is given to protecting the private patient's consumer surplus. When discrimination is not possible, the design of envy-free contracts enables senior consultants to extract rents and these rents increase with the private fee charged by their junior colleagues. As a result, and in this situation, junior consultants systematically get a lower private supplemental income when working alongside senior consultants.

JEL Codes: D86, I11, I18, L32

1. Introduction

In many countries, the consultants hired in public hospitals are authorized to provide private, fee-paying outpatient consultations *within* the public hospital. This generates windfall gains that can reduce contracting costs and, thus, public health expenses. A growing literature shows that such an income supplementation plays a critical role in developing countries where public funding is limited.[1] However, this specific form of dual practice is also quite instrumental in wealthier countries where patients face long waiting times to access public care.

The provision of private practices within public hospitals is a particularly contentious matter in developed countries where health authorities are expected to provide a truly universal healthcare and allow all patients to access public care within a limited amount of time. With public hospitals operating at capacity, the decision to ban private practices within such hospitals may give a sense that priority is given to the provision of public care. And yet, the welfare implications associated with this specific form of dual practice is unclear.

On the one hand, there is evidence that private practices set up in public hospitals tend to be a privilege geared towards *senior*, more experienced, consultants to attract and retain these high-profile professionals. The ability to treat patients privately makes public employment more attractive or discourages such consultants from setting up independent private practices (García-Prado and González, 2007). On the other hand, allowing consultants to treat private patients alongside public patients has been shown to trigger perverse incentives. Brekke and Sørgard (2007) and Morris et al. (2008) show that physicians can curtail the supply of *public* care to stimulate the demand for private care. Along the same lines, González (2005) and Barros and Olivella (2005) show that it can lead to some form of cream-skimming as specific public patients are being persuaded to opt for private care. González (2004) provides a rationale supporting a more optimistic outcome showing that consultants may provide high quality care to promote their reputation and boost their demand for private care. The argument made in the latter analysis is more salient for *junior* consultants who have yet to build their reputation. Thus, and arguably, giving experienced consultants a higher fixed salary and better resources, could address retention and attractiveness of a public employment without resorting to dual practice which could trigger misguided incentives. Along these lines, Barigozzi and Burani (2016) address efficient sorting of medical doctors between public hospital or a private hospital.

[1] See, for instance, Macq et al. (2001) for a report and Bir and Eggleston (2003) for a theoretical analysis.

To the best of our knowledge, very few papers have assessed the welfare implications associated with the provision of private care *within* public hospitals. And the privileged access given to the more experienced *senior* consultants to private care lacks any theoretical or empirical support. Empirically, a proper assessment might require access to potentially confidential data. In this paper we identify the conditions that would rationalise such a decision using a theoretical analysis. From a broader perspective, the issue we analyse is whether more productive consultants should be the ones incentivised when addressing a shortage of care supply. Incentivising such consultants to see more patients via a greater private income supplementation means that a larger number of patients will be treated given their expertise. However, this argument leaves aside the potential losses in "consumer surplus" that private patients could incur.

This paper proposes a normative analysis which investigates the optimal management of private practices within a public hospital. In other words, we focus on the strategic allocation of private care. We consider a situation where decisions are taken by a representative of the national Health Authority who takes into consideration the patients' and consultants' welfare as well as contracting costs. We consider that senior consultants are more experienced and able to treat more patients. Their reputation also allows them to attract more private patients than their junior counterparts for any given fee. We focus on the provision of outpatient treatments as we are interested in the rationale supporting private income supplementation for consultants.[2] We endogenize the impact of the private income supplementation on the consultants' incentives to treat a greater number of patients. The benefits of such decisions are accounted for considering that waiting lists for consultations can be lengthy and lead to some welfare losses for the patients.

Following the seminal works of Iversen (1997) and Olivella (2003), we consider that the provision of public and private health care by the same consultant are vertically differentiated. Basically, when a patient requests a private appointment the waiting time is shorter generating an increment in well-being. Moreover, since the provision of care, private or public, is done within the same hospital, the consultants do not dissociate the private care from public care.

[2] Public hospitals can also allow patients to access private rooms following inpatient treatments. This aspect is not considered here as the revenue that emerges from this possibility goes towards the hospital's funds.

We assume that patients differ in their willingness to pay for private consultations. The demand for private care is then endogenously determined based on the fee. Under this approach, the proportion of private patients decreases with the private fee. Hence, as the fee increases, a greater share of a consultant's income comes from her private practice, but a lower proportion of her patients are private patients. This means that the privilege given to some of the consultants must be clearly defined. Indeed, privilege could be understood as the authorization to treat a larger number of private patients. This is achieved when the private fee, and the consultant's private income supplement, are reduced. By opposition, privilege could be understood as granting some consultants a larger private income in returns for providing care to fewer private patients willing to pay a high fee for their services.

Our analysis focuses initially on a setting where the Health Authority has the ability to rely on perfect discrimination and offer, to each consultant, a contract based on their specific characteristics. This section enables us to conduct some informative comparative statics with regards to several exogenous variables, including the consultant's level of expertise. We then extend the analysis to a scenario where the Health Authority relies on second degree discrimination. More specifically, it considers a context where senior and junior consultants are hired simultaneously and offered the possibility to select the contract that most suits them. In this context, the objective is to design envy-free contracts maximising total welfare.[3]

We capture some of the established benefits of dual practice. Firstly, private practices lower public health expenditures as the private fee contributes to a consultant's overall earnings. Secondly, the private fee entices consultants to attend to more patients, which enables more patients to access care. Therefore, as we assume that public funds are costly to raise, it would be optimal to set the fee so as to maximize the consultant's private revenue, regardless of seniority, if we left aside concerns about the patient's economic consumer surplus. As this surplus is taken into consideration, the private fee, and consequently the consultant's private income, must be capped and the question becomes: should senior or junior consultants get any form of preferential treatment?

---

[3] See Varian (1974) for a definition of envy-free contracts. This concept is similar to incentive compatibility. We use this terminology as we capture situations where seniority is possibly verifiable but not necessarily contractible. Perfect discrimination can be subject to legal challenges impeding the offering of distinct contracts to specific individuals.

When the utility losses experienced by those remaining on waiting lists are large, the Health Authority needs to incentivise consultants to attend to more patients. This is done by rising the private fee, and thus the private income supplement. In this case the regulated fee is set higher for senior consultants. When the waiting time generates smaller utility losses, a different outcome emerges as the Health Authority focuses on protecting the patients' consumer surplus. Indeed, since senior consultants attract a larger proportion of private patients for any given fee, a large number of patients lose part of their consumer surplus when the fee increases. Thus, when utility losses associated with waiting lists are low, the regulated fee is set lower for senior consultants.
As we do not account for capacity constraints, which would restrict the number of new patients being attended to, we can conjecture that, in an economy where hospitals curtail the number of patients due to a lack of capacity, a cap of the fee is all the more justified. Furthermore, since the goal would primarily be to protect the patients' consumer surplus, senior consultants should, more often than less, charge a lower private fee.

As we expand the analysis to cater for a setting where both, senior and junior consultants are offered the same set of contracts, two distinct equilibria emerge: a separating and a pooling equilibrium. The pooling equilibrium emerges when the fee charged by senior consultants under perfect discrimination is substantially lower than the fee charged by junior consultants. In either equilibrium, junior consultants lose out as their private fee is reduced compared to a situation with perfect discrimination. When distinct contracts are made available, senior consultants are in a position to extract rents and these rents increase with the private fee charged by their junior counterparts. Hence, to maximize welfare, the Health Authority must reduce the regulated fee for junior consultants. In the separating equilibrium junior consultants charge a lower fee than senior consultants. Senior consultants may nonetheless attend to a larger number of private patients as patients are willing to pay more to see these physicians. In the pooling equilibrium all consultants receive the same private fee and senior consultants benefit from the presence of junior colleagues as they see their private fee rise but just so to match the one gathered by junior consultants.

The next section provides a review of related papers as a well as documentation supporting the context that we analyse. Section 3 describes the model. Section 4 describes the patients' and consultants' behaviour. Section 5 focuses on a situation where perfect discrimination is allowed and characterises the optimal contract in such a context. Section 6 extends the results to a situation where all consultants are offered the same set of contracts and select the ones that suits them most. Finally, we conclude in Section 7.

2. Supporting Literature

A large literature is devoted to the analysis of dual practice. Most of the papers focusing on wealthier, developed countries consider dual practice as a possibility for consultants to split their time between a public hospital and a separate, private clinic or hospital.[4] Garattini and Padula (2018) provide a historical background for such a practice and summarise its costs and benefits. Eggleston and Bir (2006) provide a summary of seminal papers addressing dual practice. Very recently Kuhn and Nuscheler (2020) analyse dual practice as a mean to offer specific treatments. This paper considers explicitly the provision of private care *within* public hospitals. This practice has been adopted in 16 OECD member states and is managed differently in different countries.[5] Paris et al. (2010) explains that, in some countries (such as Belgium) this privilege is geared towards consultants who are self-employed and paid via a fee-for-service. In other countries such as France, Ireland, and the UK, where consultants are employees of the public hospitals, dual practice is potentially part of the contractual agreement made with the public hospital.

Where allowed, the provision of private care within public hospitals is typically regulated. In the UK, consultants with a full-time contract from English National Health Service (NHS), have a private income limited to 10% of their NHS salary (Raffel, 2007). In France, consultants who engage in private practice cannot earn a private income that is above 30% of their overall income (Kiwanuka et al., 2011). In Ireland, consultants can see at most 20 private patients out of 100 (Health Service Executive, 2019). In general, issues surrounding the optimal regulation of dual practices have received much attention in the literature. García-Prado and González (2007) provide an extensive review of the different regulatory policies that are used in different countries and highlight their associated benefits and risks. González and Macho-Stadler (2013) provide a theoretical comparison of distinct regulatory measures. What emerges from this literature is the need to adapt to the economic environment present in each and every country.

The necessity to allow dual practices in order to reduce waiting lists for public care is an argument that is often put forward. Whether this argument is

[4] See for instance Barros and Siciliani (2011), García-Prado and González (2011), González et al. (2017), and Mueller and Socha-Dietrich (2020).

[5] Please see Table 14 of the OCED report by Paris et al. (2010), as well as Garattini and Padula (2018) and Ofer et al. (2006) and finally, Appendix B.

correct is debatable. The seminal works by Iversen (1997) and Olivella (2002) show that waiting lists can be adjusted to reduce health care costs by inducing an optimal allocation of patients between *private hospitals* and *public hospitals*. The latter are typically capacity constrained. Their analyses substantiate the fact that waiting times tend to increase in places where private care becomes available and more accessible because it reduces the demand for care within public hospitals thereby reducing their costs. Using data from Eurostat, Figure 1 below illustrates that the percentage of the population with unmet medical needs on average is much more prominent in the countries which forbid private practice within public hospitals.[6]

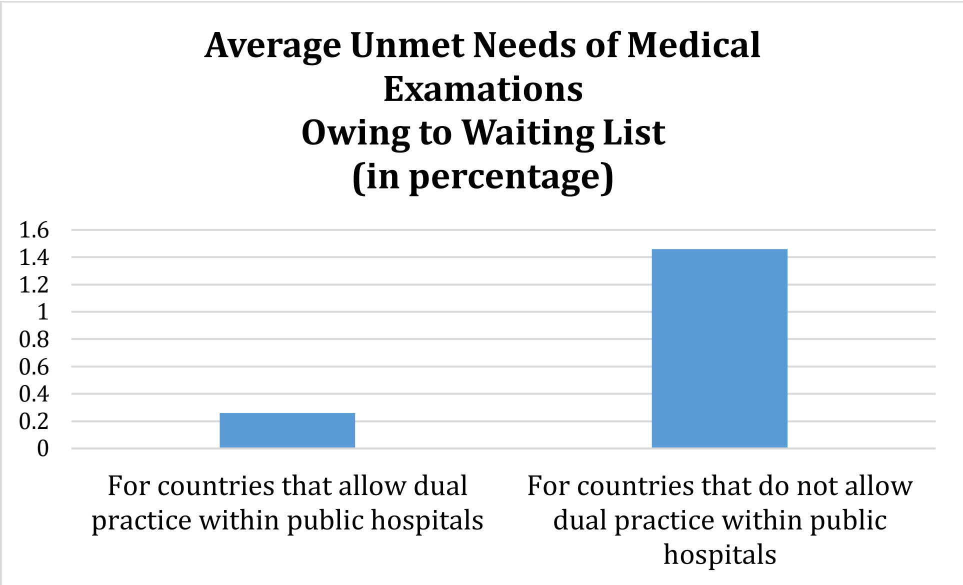

Figure 1: Unmet Needs for Medical Examination Owing to Waiting List

This chart shows the average percentage of self-reported unmet needs for medical examinations because of waiting lists, between the countries that allow private practice within the public hospitals and the countries that do not. It is based on data from EuroStat. The countries included here are Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland. For the information of the allowance of private practice within public hospitals, please see Appendix A.

As argued in the introduction, the ability to attend to private patients within the public hospital is a privilege that is, in some instances, geared exclusively towards specific consultants. This observation is more difficult to substantiate. Mueller and Socha (2018) mention in their OECD report that public

[6] Note that we are not claiming any causality from Figure 1.

hospitals use private practice privileges to attract and retain specific consultants (especially senior ones). In France, the private income supplementation that emanates from this particular form of dual practice enables public hospitals to match the salaries that senior, experienced consultants would get in private hospitals (Paris et al., 2010). According to a report produced by the British Medical Association (2021), in the UK this privilege is based upon the consultant's ability, experience and references - and it must generally be approved by the hospital's Medical Advisory Committee.[7]

Finally, our analysis considers that consultants are knowledge workers who do not only respond to monetary incentives. Following Besley and Ghatak (2005), Biglaiser and Ma (2007) and Delfgaauw (2007) (among others) we consider that consultants are mission-oriented workers who are also intrinsically motivated. We capture this assuming that each consultant receives some form of personal gratification when successfully diagnosing a patient. We then capture the difference in seniority as a difference in expertise and assume that senior doctors issue a correct diagnosis with a greater probability.

3. The Model

We consider the problem of a Health Authority (hereafter HA) that manages a public hospital attended by N patients in need of *outpatient* consultations from K consultants. The variable N is very large. We consider one period during which some patients will be seen (either as public patients or as private patients) while others will remain on waiting lists. When n_i patients are attended to by consultant i, the welfare loss incurred by those who are not seen is captured by $\left(N - \sum_{i=1,\ldots,K} n_i\right)C$ where $C > 0$.

Each consultant is characterized by their *productivity* which measures their ability to diagnose patients accurately and provide an effective treatment. Let $\gamma_i \in [0,1]$ be the probability with which consultant i reaches a correct diagnosis. For the purpose of the analysis, we will refer to *senior* consultants as these consultants with a higher productivity. We assume that the consultants' level of seniority is verifiable to all patients and to the HA.[8] Consultants are also altruistic. A consultant's altruism is captured assuming that, for each patient who is successfully treated, the consultant receives some gratification incrementing their utility by $\upsilon > 0$, regardless of patient's private or public status. Finally, the patients

[7] See https://www.bma.org.uk/advice-and-support/private-practice/working-in-private-practice/sas-doctors-and-private-practice

[8] Typically, a consultant's working experience is publicly accessible on the hospital's website or on their personal websites. In Section 5, we consider that while this variable may be verifiable, it may not be possible for the HA to rely on perfect discrimination.

experience an increment in their well-being when cured given by $\mu = 1$ while $\mu = 0$ if they are misdiagnosed. It follows that γ_i measures the *expected* utility of accessing care experienced by patients attended to by consultant i.

Public and private health services are vertically differentiated and the welfare that these generate are described as follows. Public care is free of charge but provided with some delay. Therefore, the expected utility a patient receives from public care is given by $\beta\gamma_i$ where the discount factor $\beta \in [0,1]$ depends on the quality of the provision of public care (e.g. the length of waiting lists). Private care is available within a shorter time, with the same consultant, but it is subject to an out-of-pocket fee $s_i \geq 0$. We consider that this fee is paid to the consultant directly.[9] When patient j is privately attended to by consultant i, she gets an expected utility given by $(\theta_j + \gamma_i - s_i)$. The variable θ_j captures the patient's willingness to pay for a private consultation. The value for θ_j depends on several intrinsic characteristics ranging from the patient's income and insurance plan to the possible perceived anxiety associated with waiting. We therefore consider that the value it takes is the realization of a random variable $\tilde{\theta}$ distributed over $[0, \overline{\theta}]$ according to a probability distribution function $f(.)$ with an associated cumulative distribution function $F(.)$. We assume that $\overline{\theta}$ is finite but large enough.

Finally, the contracts that the HA issues specify the private fee that consultants can ask for (s) and a fixed income (t). The HA recognizes that consultants are knowledge workers who are experts in their disciplines. As such, each consultant decides on the number of outpatients they can attend to, and we let n_i denote this variable. We assume that consultant i takes decisions considering that they bear a cost $\frac{1}{2}(n_i)^2$ when attending to n_i patients.

4. The patients' and consultants' decisions.

- **The patients**

Considering the difference between public and private services, for any fee $s_i > 0$, patient j will request to see the consultant i privately provided the willingness to pay of patient j for private care is high enough:

[9] In some countries, such as Australia, specialists transfer part of the private fee to public hospitals (Mueller and Socha, 2018). In Ireland, public hospitals get zero share from the private insurance when patients use diagnosis machines (Independent.ie, 2019); in some circumstances where private insurance provides a bundle payment to the hospital, then the hospital has power to decide the allocation of funds (The Competition Authority, 2005). From the perspective of total surplus, this transfer between a consultant and the public hospital would not impact our results. For more on implementation of private care see García-Prado and González (2007).

$$\gamma_i + \theta_j - s_i \geq \beta\gamma_i \Leftrightarrow \theta_j \geq s_i - (1-\beta)\gamma_i. \quad (1)$$

The threshold value $(s_i - (1-\beta)\gamma_i)$ decreases with γ_i and increases with β. Therefore, for any given fee, the demand for private consultations is higher when the consultant has a high ability and lower when the quality of public care increases.[10] Let $\hat{\theta} \equiv \max\{0, s_i - (1-\beta)\gamma_i\}$, for any given $s_i > 0$ a proportion $\left(1 - F(\hat{\theta})\right)$ of patients are willing to see the consultant privately. We denote the consultant's expected private revenue per patient by

$$r_i(s_i) \equiv s_i\left(1 - F(\hat{\theta})\right). \quad (2)$$

We make the following assumptions.

Assumption 1: The density function $f(.)$ is such that, for all i, $f(\hat{\theta}) + f'(\hat{\theta})s_i > 0$ and such that $f(0) = \varepsilon$ where ε is arbitrarily small so that $1 - f(0)(1-\beta)\gamma_i > 0$.

Under this assumption, the expected private revenue $r_i(s_i)$ is concave in s_i and such that there exists $\overline{s}_i > (1-\beta)\gamma_i$ that maximizes $r_i(s_i)$.[11] This assumption captures situations where the demand for private care is sufficiently inelastic. As such, neither the demand nor the private revenue falls sharply when the private fee increases slightly above $(1-\beta)\gamma_i$.[12]

- **The consultants**

The consultants' decision, in terms of the number of patients that they can attend to, is based on their productivity (captured by γ), their level of altruism (captured by ν) and on the compensation that they get. The overall utility gathered by consultant i is given by

$$U_i(n_i) = t_i + \gamma_i n_i \nu + r_i(s_i)n_i - \frac{1}{2}(n_i)^2. \quad (3)$$

Let $u_i(n_i) = \gamma_i n_i \nu + r_i(s_i)n_i - \frac{1}{2}(n_i)^2$, so that $U_i(n_i) \equiv t_i + u_i(n_i)$. Simple calculations show that a consultant characterized by a level of productivity and altruism (γ, ν) who can charge s for private consultations, treats n^* patients where

$$n^* = \left(\gamma\nu + r(s)\right). \quad (4)$$

[10] This is consistent with the literature (see Besley et al., (1999)).
[11] In Section 5 where we consider consultants are heterogeneous, we assume $\tilde{\theta}$ is uniformly distributed. The assumption holds provided $\overline{\theta}$ is very large.
[12] Figures 3 and 4 in Appendix 1 give a visual representation of the proportion of patients seeking private care, as a function of the private fee, which would satisfy the above assumption.

This number increases with the productivity parameter γ, but also with the private income supplementation. This means that the consultants' incentives respond to the value of the private fee as well as the quality of public care captured by β. For a higher value of β, the proportion of private patient is lower meaning that private practices are less lucrative. Therefore, for a given private fee, the consultants see fewer patients. At the solution, we have

$$U(n^*) = t + \frac{1}{2}(n^*)^2. \tag{6}$$

Letting U_R denote a consultant's reservation utility. A consultant with a level of productivity and altruism (γ, ν) accepts the contract from the HA provided $U(n^*) \geq U_R$.

5. Optimal Contracting under Perfect Discrimination.

In this section we assume that the HA can rely on first-degree discrimination and offers each consultant a specific contract which is tailored to their intrinsic characteristics. To solve for the optimal contract in this case, and without any loss in generalities, we assume that all consultants have the same ability γ and let K be the overall number of consultants.

The HA's representative issues contracts that maximize the total surplus. The objective function captures the following elements:

(i) The patients' welfare which accounts for the benefits gathered when being attended to, and the losses incurred when remaining on the waiting list.

(ii) The consumer surplus gathered by private patients measured as the difference between their willingness to pay and the private fee.

(iii) The consultants' well being captured via their utility function (6).

(iv) The transaction costs associated with raising public funds. The HA is only accountable for the fixed wage t_i, because the private income supplementation is paid by the patients who are seen privately (or by their private health insurance).

Therefore, the overall total surplus that is maximized by the HA is given by

$$\begin{aligned} TS = \sum_{i=1,\dots,K} [n_i PW(\gamma, s) + U_i(n_i) - (1+\lambda)t_i] \\ - \left(N - \sum_{i=1,\dots,K} n_i\right) C, \end{aligned} \tag{7}$$

where $\lambda \in]0,1[$ is the shadow cost of raising public funds, and $PW(\gamma, s)$ is the per-patient expected welfare that arises when receiving a medical treatment:

$$PW(\gamma, s) = \beta\gamma F(\hat{\theta}) + \gamma\left(1 - F(\hat{\theta})\right) + \int_{x=\hat{\theta}}^{\overline{\theta}} (x - s) f(x) dx. \tag{8}$$

The first term is the expected welfare received by a patient when accessing *public* health care. The second term is the expected welfare gathered by a patient who receives a *private* treatment. And, finally, the last term is the consumer surplus gathered by patients attended to privately. Expression (8) can be rewritten as

$$PW(\gamma, s) = \begin{cases} \gamma + \theta^e - s, & \text{if } s \leq (1-\beta)\gamma, \\ \beta\gamma + \displaystyle\int_{s-(1-\beta)\gamma}^{\overline{\theta}} \left(1 - F(x)\right) dx, & \text{otherwise,} \end{cases}$$

where $\theta^e = E(\tilde{\theta})$ denotes the expected willingness to pay for private care.

The expressions above deserve careful consideration. The product $(1-\beta)\gamma$ captures the increment in the quality of care that a patient receives when opting for a private consultation. When $s \leq (1-\beta)\gamma$ all patients request private appointments and the surplus generated is the sum of the quality of private care and the expected consumer surplus given by $(\theta^e - s)$. When the fee increases beyond $(1-\beta)\gamma$, the patients who request private consultations pay for the increment in the quality of care that they receive thereby annihilating the benefit of receiving faster treatments. Hence, in that case, the provision of care generates a welfare that accounts for the provision of *public* care only and the consumer surplus extracted by private patients.

The HA must design a contract (t and s) that maximizes the total surplus perfectly anticipating the choice of consultants described in Section 4. Furthermore, as the HA representative is able to rely on first-degree discrimination, the only constraint it faces is a participation constraint. Given that there is a cost of raising public funds, the utility of consultants is socially costly. It is therefore optimal to set t such that $U(n^*) = U_R$. Lemma 1 enables us to refine the set of optimal values for the private fee.

Lemma 1*: It is not optimal to set the private fee below or equal to* $(1-\beta)\gamma$.

Proof: Consider all $s \leq (1-\beta)\gamma$. For such values of the private fee we have $r(s) = s$ and $n^* = (\gamma\nu + s)$. Moreover, given that the fixed wage is optimally set so that $U(n^*) = U_R$, the total surplus can be written as

$$TS = K\left[(\gamma v + s)(\gamma + \theta^e - s + C) + \frac{1}{2}(1+\lambda)(\gamma v + s)^2 - \lambda U_R\right] - NC.$$

The function TS is concave and increasing over the range $[0, (1-\beta)\gamma]$ since

$$\frac{d^2 TS}{ds^2} < 0 \text{ and } \left.\frac{dTS}{ds}\right|_{s=(1-\beta)\gamma} = K[\gamma(1+\lambda v) + \theta^e + C - (1-\lambda)(1-\beta)\gamma] > 0.$$ [13]

Therefore, for any fee $s \leq (1-\beta)\gamma$, all patients want to access private care and a marginal increase in the fee increases total welfare because the marginal gains from having to pay a lower wage (t) surpass the marginal losses in surplus incurred by patients.■

Lemma 1 establishes that the optimal fee is greater than $(1-\beta)\gamma$. For any such fees we have $\hat{\theta} = s - (1-\beta)\gamma$, $r(s) = s\left(1 - F(\hat{\theta})\right)$ and $n^* = (\gamma v + r(s))$. The total surplus is given by

$$TS = K\left[n^*\left(C + \beta\gamma + \int_{x=\hat{\theta}}^{\overline{\theta}} \left(1 - F(x)\right)dx\right) + \frac{1}{2}(1+\lambda)(n^*)^2 - \lambda U_R\right] - NC.$$

Proposition 1: *The optimal contract (t^*, s^*) is such that each consultant receives their reservation utility $(U(n^*) = U_R)$ and $s^* \in](1-\beta)\gamma, \overline{s}[$. The optimal fee is such that the number of patients is given by*

$$n^*(s^*) = \frac{r'(s^*)}{\left[\left(1 - F(\hat{\theta})\right) - (1+\lambda)r'(s^*)\right]}\left[C + \beta\gamma + \int_{x=\hat{\theta}}^{\overline{\theta}} \left(1 - F(x)\right)dx\right].$$

Proof: See Appendix 2.

At the solution we have

$$\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \left(1 - F(x)\right)dx\right] r'(s^*) + (1+\lambda)n^* r'(s^*) - n^*\left(1 - F(\hat{\theta})\right) \equiv 0. \tag{9}$$

A marginal increase in the private fee incentivises the consultant to see more patients provided the fee is lower than $\overline{s} = \arg\max r(s)$. When this is so, the consultant sees $r'(s)$ more patients and the welfare implications are similar

[13] Concavity is ensured by assumption as we consider that $\lambda < 1$.

to those arising when considering the optimal level of monopolistic output. Raising the fee marginally incentivises consultants to attend to more patients thereby allowing more patients to access care. But it penalises the cohort that was accessing care prior to increasing the fee and who now face a higher fee.

The expression $\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}}\big(1 - F(x)\big)dx\right] r'(s)$ in (9) captures the surplus extracted by the $r'(s)$ patients who are taken off the waiting lists when the private fee increases slightly. These no longer incur losses C. These receive an expected utility of $\beta\gamma$ when seen publicly or else they get an expected consumer surplus given by $\int_{x=\hat{\theta}}^{\overline{\theta}}\big(1 - F(x)\big)dx$ when seen privately. The second term, $(1 + \lambda)n^* r'(s)$, captures the savings in public health expenditures that arise when the private fee is marginally increased. As private income supplementation increases, less public funds must be raised to remunerate the consultants. Finally, the last term, $n^*\left(1 - F(\hat{\theta})\right)$ comes in negatively as it measures the loss in consumer surplus incurred by the "*original*" private patients who must now pay a slightly higher fee.

There are two important features that emerge from Proposition 1.

Corollary 1: *At the solution, each consultant's expected private income is capped* $(r'(s^*) > 0)$*. Furthermore, a further increase in the private fee would lower the private patients' consumer surplus by an amount larger than the savings in public funds that such an increase would generate.*

The first point comes from the fact that equation (9) holds provided $r'(s^*) > 0$. This means that the optimal fee is set at a level where the private income is capped. Moreover, this private income results from a low fee and high proportion of private patients rather than a high fee and a low proportion of private patients. It also means that a marginal increase in the fee increases the consultant's private income and, therefore, their incentive to attend to more patient. The only reason why the maximization of the private revenue is not optimal is due to the consideration of the private patients' consumer surplus. Absent of any such concern, it is optimal to set the private fee such that $r'(s) = 0 \Leftrightarrow s = \overline{s}$, so that the private revenue is maximized. This would, in turn, maximize the number of patients being attended to, thereby reducing the waiting lists. Moreover, the monetary transfer t that consultants receive from the hospital would also be lower, which saves the cost of raising public funds.

The second point in Corollary 1 captures the fact that, at the solution we must have $(1 + \lambda)n^* r'(s^*) - n^*\left(1 - F(\hat{\theta})\right) < 0$. This means that, a further rise in the private fee would lead to savings, in terms of lower public funds, that can only be made at the expense of the private patients' consumer surplus.

We are now in a position to assess whether senior consultants should be prioritised when allocated a private income supplementation.

Lemma 2: *The optimal fee s^* is (i) non-increasing with intrinsic motivation ν and (ii) non-decreasing with the cost of waiting C. However, the sign of the derivative of s^* with respect to the quality of public care β and consultant's ability γ is not trivial.*

Proof: See Appendix 3.

Point (i) of Lemma 2 states that consultants who possess a higher intrinsic motivation should receive a lower private fee leading to a lower private income. The intrinsic motivation ν and the private revenue are substitutes when it comes to incentives. Consultants with a greater intrinsic motivation are inclined to attend to more patients and would settle for a lower fixed wage. Allowing such consultants to charge a higher fee would exacerbate their natural inclination to see more patients and lower the surplus received by private patients. The losses the private patients incur is not compensated by the lower wage that the consultants settle for. Thus, all in all, it is optimal to decrease the fee received by more intrinsically motivated physicians.

The logic behind point (ii) is straightforward. When C is high, the HA wants to reduce the number of patients on waiting lists by incentivizing consultants to see more patients. This is done by increasing the private fee which, in turn, increases the private revenue. Even though the private patients lose some surplus, the benefits resulting from a reduction of waiting lists are large enough to compensate their losses.

When it comes to the variables β and γ, the evaluation of the comparative statics is complex. To understand why this is so, it is useful to rewrite the equality satisfied by the optimal fee as follows:

$$H(s^*) = \left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1 - F(x)\big)dx\right] r'\big(s^*, \hat{\theta}\big) \\ + \left(\gamma\nu + s^*\left(1 - F\big(\hat{\theta}\big)\right)\right)\left[(1+\lambda)r'\big(s^*, \hat{\theta}\big) - \left(1 - F\big(\hat{\theta}\big)\right)\right] \equiv 0. \tag{10}$$

A marginal increase in β or in γ has multiple implications. We separate the direct impact that the variables β and γ have on $H(s^*)$ from the more subtle ones.

- **Direct impacts**

When β or γ increase, patients get a greater welfare when accessing public care. This is captured by $\beta\gamma$ in the first bracket. Given that $r'(s^*) > 0$, this would call for a rise in the fee. In other words, as the quality of public care increases, consultants should be incentivised to attend to more patients.

The variable γ also has a direct impact on the number of patients that are attended to which is captured by the term γv. This impact would call for a downward adjustment of the fee since $\left[(1+\lambda)r'\left(s^*,\hat{\theta}\right) - \left(1 - F\left(\hat{\theta}\right)\right)\right] < 0$. This suggests that the senior consultant's inclination to attend to more patients must be curtailed. A lower fee requires that a greater share of their income be paid via a fixed transfer t. Nevertheless, the losses that this generates are recouped by a benefit as fewer patients are subjected to a marginal loss in their consumer surplus.

- **Impacts via the share of private patients.**

Now we consider the impact of β and that of γ on the share of private patients captured by the variable $\hat{\theta}$. Note that the threshold value $\hat{\theta}$ increases with β but that it decreases with γ. Patients are less likely to request a private consultation when the quality of public care increases and more likely to do so when the consultant's ability is high.

There are two reasons why patients should be encouraged to seek private care. The first is that it allows them to extract a consumer surplus. The second is that it reduces the public health expenditures as more of the consultant's income originates from the private practice. Therefore, these arguments call for the private fee to be reduced when β increases. By opposition, the same arguments call for a rise in the fee for senior consultants since an increase in the consultant's ability generates a greater proportional demand for private care.

A rise in $\hat{\theta}$ is however welcome for one reason: it reduces the proportion of those we called the "*original*" patients who incur a loss when the fee is marginally increased. This would call for a lower fee geared towards senior consultants and a higher one when the quality of public care increases.

- **Impacts via monetary incentives.**

Finally, we capture the impacts that β and γ have on $r'\left(s^*, \hat{\theta}\right)$, which measures the provision of monetary incentives. Here again, β and γ have opposite impacts. There are negative synergies between s^* and β when it comes to incentives. The marginal increase in the number of patients attended to, given by $r'(s^*)$, is decreasing in β. This is because a higher proportion of patients will seek public care when β increases, which weakens the consultant's incentive to see more patients. By opposition, positive synergies exist between s^* and γ when it comes to incentives as $r'(s^*)$, is increasing in γ.[14] This is because a higher proportion of patients will request private care as the consultant's ability increases. As the function $r(.)$ is concave in s, and, at the solution we have $r'(s^*) > 0$, a reduction in the private fee helps mitigate the impact generated by an increase in β, while an increase in the fee mitigates the impact generated by an increase in γ.

- **Overall impact**

It is not possible to deduct which force is the dominating one. However, the above tells us that senior consultants should be receiving a lower private income supplementation when priority is given to those referred to as "*original*" patients, who would suffer a welfare loss when the fee increases. This impact would be stronger when hospitals are capacity constrained so that consultants are limited in the number of additional patients that they can attend to.

Using simulations, we show that the negative effects can be the dominating ones and that a reduction of the private fee is optimal when the quality of public care increases and when the ability of the consultant increases. Figure 2 and 3 in Appendix represent the optimal fee as a function of the parameter β (figure 2) and as a function of the parameter γ (figure 3). Both depict situations where the optimal fee is decreasing.

- **Summarising the outcomes that emerge under perfect discrimination**

Let us assume that there are two types of consultants, senior and junior consultants. These are characterized by their ability and let us assume that $\gamma = \gamma_H$ for senior consultants and $\gamma = \gamma_L$ for junior consultants where $\gamma_H > \gamma_L$. Let s_H^* and s_L^* denote the optimal fees that prevail under first-degree discrimination. These values are the solutions to equation (9) above with $\gamma \in \{\gamma_H, \gamma_L\}$. The table below summarises which conclusions emerge based on the different possible

[14] We have $\frac{dr'(s^*)}{d\beta} = \gamma\left(-f(\hat{\theta}) - s^* f'(\hat{\theta})\right)$ and $\frac{dr'(s^*)}{d\gamma} = (1-\beta)\left(f(\hat{\theta}) + s^* f'(\hat{\theta})\right)$. Under Assumption 1, the former is negative and the latter positive.

outcomes. From Lemma 2, we know that there is no clear ranking between s_H^* and s_L^* and we may have $s_H^* > s_L^*$ or $s_H^* = s_L^*$ or $s_H^* < s_L^*$. Each of these might translate into a different form or privilege.

	$s_H^* > s_L^*$	$s_H^* = s_L^*$	$s_H^* < s_L^*$
Proportion of private patients	Inconclusive	Greater for senior consultants	Greater for senior consultants
Private revenue $r(s_i), i = H, L$	Greater for senior consultants	Greater for senior consultants	Inconclusive
Total number of patients $n_i, i = H, L$	Greater for senior consultants	Greater for senior consultants	Inconclusive
Number of private patients	Inconclusive	Greater for senior consultants	Inconclusive

Table 1: comparing outcomes for all possible private fees.

When the patients' willingness to pay for private care is uniformly distributed over $[0, \overline{\theta}]$, one can establish that senior consultants see a higher proportion of private patients provided $(1-\beta)(\gamma_H - \gamma_L) > (s_H^* - s_L^*)$. They receive a greater private revenue provided $(s_H^* - s_L^*)\left(\overline{\theta} - (s_H^* + s_L^*)\right) + (1 - \beta)(\gamma_H s_H^* - \gamma_L s_L^*) > 0.$[15] Therefore, all of the inconclusive cases in Table 1 would become "Greater for senior consultants" when $(\gamma_H - \gamma_L)$ is very large.

Generally speaking, senior consultants are prioritized in terms of income if they receive a higher fee than junior consultants. Alternatively, if they receive a lower fee, they are prioritized in terms of the proportion of private patients that they attend to.

6. Optimal Envy-Free Contracts

In this section, we consider that perfect discrimination is not possible for legal reasons or to reduce transaction costs. Instead, all consultants are offered the same *set of contracts* from which they can pick the contract that suits them

[15] For $\overline{\theta}$ sufficiently large we always have $\left(\overline{\theta} - (s_H^* + s_L^*)\right) > 0$ as the private fees are capped.

the most.[16] We assume that senior and junior consultants are characterized by their ability and set $\gamma = \gamma_H$ for senior consultants (or type H consultants), and $\gamma = \gamma_L$ for junior consultants (or type L consultants), where $\gamma_H > \gamma_L$.

The question addressed here can be rephrased as assessing whether senior consultants should have a greater access or revenue from private care when the HA needs to design envy-free (or incentive compatible) contracts? The analysis becomes more complex. Therefore, to make it tractable, we restrict our attention to a setting where the following assumption applies.

Assumption 2: The patients' willingness to pay for private care is uniformly distributed over $[0, \overline{\theta}]$. The variable $\overline{\theta}$ is assumed to sufficiently large so that Assumption 1 holds.

Nature determines the proportion $q \in [0,1]$ of type H consultants. As in the previous section, the HA perfectly anticipates the number of patients that the consultant attends to. In this setting however, this number depends on the consultant's type and on the contract that they choose. Specifically, simple calculations maximizing equation (3) show that a consultant with productivity γ_i who selects contract (t_j, s_j) will attend to

$$n_{ij}^* = [\gamma_i v + r_i(s_j)] \tag{11}$$

patients where

$$r_i(s_j) = s_j \frac{\overline{\theta} - \hat{\theta}_{ij}}{\overline{\theta}}, \text{and } \hat{\theta}_{ij} \equiv \max\{0, s_j - (1-\beta)\gamma_i\}.$$ [17]

At the solution such a consultant gathers a utility $U_{ij} = t_j + \frac{1}{2}[\gamma_i v + r_i(s_j)]^2$.

As it accounts for the presence of junior and senior consultants, the HA offers two contracts. Contract (t_H, s_H) is geared towards type H consultants, and contract (t_L, s_L) is designed for type L consultants. In equilibrium each contract must be voluntarily accepted by the type of consultants that it is meant for. This means that the HA faces two participation and two incentive constraints. The participation constraint guarantees that a consultant with productivity γ_i gets a utility that is least as great as their reservation utility when taking contract (t_i, s_i):

[16] In 2008 a contract reform took place in Ireland whereby public consultants were given the possibility to select their contract from a set of contracts differing in their private practice allowance (HSE, 2009).

[17] For any given private fee s_i, senior consultants enjoy a greater private revenue and higher marginal revenue $r_H(s_i) \geq r_L(s_i)$ and $r_H{}'(s_i) \geq r_L{}'(s_i)$. Therefore, for any given fee, senior consultants attend to more patients as we have $n_{Hi}^* > n_{Li}^*$.

$$U_{ii} = t_i + \frac{1}{2}[\gamma_i \nu + r_i(s_i)]^2 \geq U_R.$$

Since consultants have a choice, the contracts must be written so that each consultant selects the contract that is meant for them. In other words, contracts must be envy-free. In equilibrium, a consultant with productivity γ_i takes the contract (t_i, s_i) as opposed to (t_j, s_j) provided $U_{ii} \geq U_{ij}$:

$$t_i + \frac{1}{2}[\gamma_i \nu + r_i(s_i)]^2 \geq t_j + \frac{1}{2}\left[\gamma_i \nu + r_i(s_j)\right]^2$$

$$\Leftrightarrow t_i - t_j \geq \frac{1}{2}\left[r_i(s_j) - r_i(s_i)\right]\left[2\gamma_i \nu + r_i(s_j) + r_i(s_i)\right]$$

where $i \in \{H, L\}, j \in \{H, L\}, i \neq j$.

We reach a first conclusion which points to the necessity to privilege senior consultants when discrimination is not possible.[18]

Lemma 3: *When the Health Authority cannot rely on perfect discrimination, the envy-free constraints hold if and only if the private fee charged by senior consultants is at least as great as the private fee charged by senior consultants, that is* $s_H \geq s_L$.
Proof: See Appendix.

In this context the HA must solve the following optimisation problem

$$\max_{t_L, t_H, s_L, s_H} K[qTS_{HH} + (1-q)TS_{LL}] - NC,$$

where

$$TS_{ii} = n_{ii}^*\big(C + PW(\gamma_i, s_i)\big) + \frac{1}{2}(1+\lambda)(n_{ii}^*)^2 - \lambda U(n_{ii}^*),$$

and where $PW(\gamma_i, s_i)$ is given by (8).

The constraints that the HA is subjected to are the participation constraints $(PC_i, i = H, L)$ and the envy-free constraints $(EF_i, i = H, L)$

$$PC_i\colon t_i + \frac{1}{2}[\gamma_i \nu + r_i(s_i)]^2 \geq U_R, i = H, L$$

$$EF\colon s_H \geq s_L.$$

[18] This result holds in general settings, that is it also holds when the willingness to pay for private care is not uniformly distributed.

As is commonly the case in contract theory, the more productive consultants have access to some *rents*. In other words, any contract that is acceptable to a junior consultant is acceptable to a senior consultant who would accept a lower monetary compensation than their junior partner. Therefore, if contracts are designed in such a way that a senior consultant prefers contract (t_H, s_H) it must be the case that this contract gives this consultant a utility that equals, at least, the one they would get by taking (t_L, s_L).

Corollary 2: Senior consultants gather some rents so that $U_{HH} > U_R$.
Proof: Relying on the envy-free constraint for type H and the participation constraint for type L, we have

$$t_H + \frac{1}{2}[\gamma_H v + r_H(s_H)]^2 \geq t_L + \frac{1}{2}[\gamma_H v + r_H(s_L)]^2 > t_L + \frac{1}{2}[\gamma_L v + r_L(s_L)]^2 \geq U_R.$$

The first inequality holds as contracts must be envy-free for type H consultants. The second inequality holds because of $r_H(s_i) \geq r_L(s_i)$. The last one holds as the participation constraint for type L consultants. ■

The main implication from Corollary 2 is that the participation constraint for senior consultant is redundant. Moreover, since public funds are associated with a shadow cost, it is in the interest of the HA to minimise the fixed wages t_H and t_L. Clearly, it is therefore optimal to set t_L such that and $U_{LL} = U_R$. Moreover, it is optimal to set t_H such that senior consultants are indifferent between the two contracts so that their envy-free constraint binds:

$$t_H = t_L + \frac{1}{2}[r_H(s_L) - r_H(s_H)][2\gamma_H v + r_H(s_L) + r_H(s_H)]. \tag{12}$$

Given the above, the rents $R(s_L)$ that are available to senior consultants are solely dependent on s_L and given by

$$R(s_L) = \frac{1}{2}[\gamma_H v + r_H(s_L)]^2 - \frac{1}{2}[\gamma_L v + r_L(s_L)]^2. \tag{13}$$

Furthermore, these rents are non-decreasing with s_L since

$$\frac{dR(s_L)}{ds_L} = r_H{}'(s_L)[\gamma_H v + r_H(s_L)] - r_L{}'(s_L)[\gamma_L v + r_L(s_L)] \geq 0. \tag{14}$$

Corollary 3: The HA can reduce the rents gathered by senior consultants by lowering the junior consultants' private fee.

The statement above provides an argument suggesting that preferential treatment be given to senior consultants in terms of supplemental private income. The proposition below characterizes the first type of equilibrium that can arise when consultants can select their contracts.

Recall that s_H^* and s_L^* denote the optimal fees that prevail under first-degree discrimination. These are the ones characterised in the Section 5. In the same section, we establish that we could either have $s_H^* \geq s_L^*$ or $s_H^* < s_L^*$. Let s_H^{**} and s_L^{**} denote the optimal fees in this new situation. At $s_L = s_L^{**}$, we have

$$(1-q)[C + PW(\gamma_L, s_L) + (1+\lambda)n_{LL}]\frac{dr_L}{ds_L} + (1-q)n_{LL}\frac{dPW}{ds_L} - \lambda q \frac{dR}{ds_L} \equiv 0. \tag{15}$$

We obviously have $s_L^{**} < s_L^*$ (see Appendix 4). However, it is not clear how s_L^{**} compares to s_H^* and this is key to determining which type of equilibrium we reach.

Proposition 2: *Let* s_H^* and s_L^* *denote the optimal fees under first-degree discrimination and* s_H^{**} *and* s_L^{**} *the optimal fees when discrimination is not possible. Two possible outcomes can emerge.*

- *A separating equilibrium can emerge in which the private fee charged by senior consultants is the same as the one they would charge under perfect discrimination, that is* $s_H^{**} = s_H^*$. *In equilibrium, private fee charged by junior consultants is lowered and we have* $s_L^{**} < s_H^*$ *and* $s_L^{**} < s_L^*$.
- *A pooling equilibrium can emerge wherein all consultants get the same private fee* $s^{**} \in [s_H^*, s_L^*]$. *In equilibrium, senior consultants see more private patients and get a higher private income than junior consultants.*

Proof: See Appendix.

The separating equilibrium systematically emerges as the unique outcome when $s_L^* \leq s_H^*$ meaning that the fee charged by junior consultants is lower under first-degree discrimination. In the separating equilibrium, based on Table 1, we know that senior consultants receive a higher private income and that they attend to more patients. However, the number of private patients that they attend to is not necessarily larger than that of junior consultants as a lower proportion of their patients may be private patients. In the pooling equilibrium, which only emerges when $s_L^* > s_H^*$, the senior consultants treat a higher proportion of private patients owing to their higher ability. As they charge the same fee as their junior counterpart, senior consultants also get a higher private income supplementation and see more patients.

The table below summarises our results.

Objective	**Perfect discrimination**	**Envy-free contracts**
Reduction of the waiting lists (C is large)	Senior consultants get a higher fee $s_H^* > s_L^*$.	Senior consultants get s_H^* and the fee of junior consultants is

		reduced to limit the rents extracted by senior consultants
Protecting the private patients' consumer surplus (C is low)	Senior consultants get a higher fee $s_H^* < s_L^*$.	If s_L^* is not much larger than s_H^*, senior consultants get s_H^* and the fee of junior consultants is reduced. Otherwise, all consultants charge the same private fee $s^{**} \in [s_H^*, s_L^*]$.

Regardless of the objective, junior consultants lose out when working alongside senior colleagues.

7. Conclusion

In many developed countries access to public health care in hospitals is constrained and patients have to wait to access care. How can the reliance on private practice be used to best address this issue? Should more productive consultants be given any privilege in their access to private practices within the public hospital? These are the questions that we aimed to address in this paper.

The private income supplementation induces consultants to attend to more patients, which reduces waiting times, and reduces the public cost of healthcare. Those features support the maintenance of private practices within public hospitals. However, the provision of private care generates a market where patients accessing private care get a consumer surplus. To protect these consumers, it is optimal to regulate the fee and cap the consultants' private income.

When first-degree discrimination is possible, it is not clear whether the more productive (senior) consultants should set a higher private fee so as to receive a higher private income. In nutshell, we show that they should set a higher private fee (and get a higher private income) when priority is given to shortening waiting lists. The provision of incentives is achieved at the expenses of private patients who will see their consumer surplus reduced.

When discrimination is not possible the fee charged by the less experienced, junior consultants must be set lower than the fee these consultants would get under first-degree discrimination. Hence junior consultants systematically get a lower private supplemental income when working alongside senior consultants. This is because senior consultants can extract rents and these rents increase with the private fee charged by their junior counterparts.

Two equilibriums emerge. When, under perfect discrimination, the fee charged by senior consultants is higher or not much lower that the fee charge by junior consultants, the HA must simply reduce the fee charged by junior consultants to achieve a separating equilibrium. In equilibrium senior consultants receive a higher private income and attend to more patients. However, their number of private patients is not necessarily larger than that of junior consultants. When, under perfect discrimination, the fee charged by senior consultants is much lower than the fee charged by junior consultants the HA reaches a pooling equilibrium wherein both types of consultants charge the same fee. Basically, it raises the private fee that senior consultants can charge and lowers the private fee that junior consultants charge. In equilibrium, senior consultants benefit from the presence of junior colleagues as they see their private fee rise when compared to what they would charge under first-degree discrimination.

We hope that these research outcomes will inform the debate surrounding the provision of private practice within public hospital. While the achievement of a truly universal healthcare is most desirable, we would not be supportive of the removal of private practice when the provision of public care is subject to capacity issues. Private practices within public hospitals have positive economic implications based on the incentives that they provide and on the savings that they generate.

In general, we believe that our analysis emphasises the fact that the decision to remove private practices from public hospitals is one that deserves a particularly cautious approach. It is naïve to believe that the removal of private patients from public hospitals will shorten waiting lists. In particular, one could extend this analysis considering the impact that a reduction of the supply of private care would have in an economy where private hospitals are not always operating under perfect competition. Indeed, the removal of private practices within public hospitals could reduce the affordability of private care and lead to an increase in the number of patients seeking public care.

References

Aran, M. and Rokx, C., 2014. Turkey on the way of universal health coverage through the health transformation program. World Bank Group [online].

Available from:
https://documents1.worldbank.org/curated/en/243161468308038702/pdf/913260WP0UHC0C00Box385331B00PUBLIC0.pdf [accessed 11 May 2022]

Barigozzi, F. and Burani, N., 2016, Competition and screening with motivated health professionals. Journal of health economics, 50, pp.358-371.

Barros, P. P. and Olivella, P., 2005, Waiting lists and patient selection. Journal of Economics and Management Strategy, 14, pp. 623–646

Barros, P. P., & Siciliani, L. (2011). Public and private sector interface. In Handbook of health economics (Vol. 2, pp. 927-1001). Elsevier.

Besley, T. and Ghatak, M., 2005. Competition and incentives with motivated agents. American economic review, 95(3), pp.616-636.

Besley, T., Hall, J. and Preston, I., 1999. The demand for private health insurance: do waiting lists matter?. Journal of public economics, 72(2), pp.155-181.

Biglaiser, G. and Albert Ma, C.T., 2007, Moonlighting: public service and private practice. The RAND Journal of Economics, 38(4), pp.1113-1133.

Bir, A. and Eggleston, K., 2003. Physician dual practice: access enhancement or demand inducement. *Tufts University Department of Economics Working Paper*, *11*.

Brekke, K.R. and Sørgard, L., 2007. Public versus private health care in a national health service. Health economics, 16(6), pp.579-601.

British Medical Association, 2021. SAS doctors and private practice[online]. Available from: https://www.bma.org.uk/advice-and-support/private-practice/working-in-private-practice/sas-doctors-and-private-practice [accessed 11 May 2022]

Campbell, A., 2013. The abolition of user fees in the Jamaican public health system: impact on access, care provided and the work of the professional nurse.

Delfgaauw, J., 2007. Dedicated doctors: Public and private provision of health care with altruistic physicians. Tinbergen Institute Discussion Paper No. 2007-010/1

Eggleston, K. and Bir, A., 2006. Physician dual practice. Health policy, 78(2-3), pp.157-166.

García-Prado, A. and González, P., 2007. Policy and regulatory responses to dual practice in the health sector. Health Policy, 84(2-3), pp.142-152.

García-Prado, A. and González, P., 2011. Whom do physicians work for? An analysis of dual practice in the health sector. Journal of health politics, policy and law, 36(2), pp.265-294.

Garattini, L. and Padula, A., 2018. Dual practice of hospital staff doctors: hippocratic or hypocritic? Journal of the Royal Society of Medicine, 111(8), pp. 265–269.

González, P., 2004, Should physicians' dual practice be limited? An incentive approach. Health economics, 13(6), pp.505-524.

González, P., 2005, On a policy of transferring public patients to private practice. Health Economics, 14, pp. 513–527.

González, P. and Macho-Stadler, I., 2013. A theoretical approach to dual practice regulations in the health sector. Journal of health economics, 32(1), pp.66-87.

González, P., Montes-Rojas, G. and Pal, S., 2017, Dual Practice of Health Workers: Theory and Evidence from Indonesia. IZA Discussion Paper Series, 11038, Bonn, Germany.

Health Service Executive (2019), *Consultants' Contract* [online]. Available from: https://www.hse.ie/eng/staff/resources/terms-conditions-of-employment/ccontract/consultant-contract-2008-25th-june-2019.pdf [accessed 27 February 2020]

HSE, 2009. *Revised salary scales for medical consultants* [online]. Available from: https://circulars.gov.ie/pdf/circular/hse/2009/13.pdf [accessed 12th June 2022]

Hsiao, W., 1998. A Preliminary Assessment of Sri Lanka's Health Sector and Steps Forward. Devolution: Inter-Govemment Fiscal Transfer Mechanism. GOSL-World Bank study in conjunction with IPS Health Policy Programme, World Bank Resident Mission, Colombo.

Immergut, E.M. and Comisso, E., 1992. Health politics: interests and institutions in Western Europe. CUP Archive.

Independent.ie. (2019). Public hospitals get no money for doing private patient scans [online]. Available from: https://www.independent.ie/irish-news/health/public-hospitals-get-no-money-for-doing-private-patient-scans-38626666.html [accessed 08 Feb 2022]

Iversen, T. (1997). The effect of a private sector on the waiting time in a national health service. Journal of Health Economics, 16(4), 381-396.

Kiwanuka, S.N., Kinengyere, A.A., Rutebemberwa, E., Nalwadda, C., Ssengooba, F., Olico-Okui, Pariyo, G.W. (2010). Dual practice regulatory mechanisms in the health sector: a systematic review of approaches and implementation. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Klein, R., 2018. The public-private mix in the UK. In The Public Private Mix for Health (pp. 53-72). CRC Press.

Kuhn, M. and Nuscheler, R., Saving the public from the private? Incentives and outcomes in dual practice, Journal of Public Economic Theory, pp.1120-1150

Macq, J., Ferrinho, P., De Brouwere, V. and Van Lerberghe, W., 2001. Managing health services in developing countries: between ethics of the civil servant and the need for moonlighting: managing and moonlighting. *Human Resources for Health Development Journal*, *5*(1-3), pp.17-24.

Morris S, Elliott B, Ma A, McConnachie A, Rice N, Skatun D, et al. Analysis of consultants NHS and private incomes in England in 2003/4. J R Soc Med 2008; 101: 372–380

Mueller, M., & Socha, K. (2018). OECD report: Assessing private practice in public hospitals [online]. Available from: https://assets.gov.ie/26530/88ebd7ddd9e74b51ac5227a38927d5f9.pdf [accessed 08 Feb 2022]

Mueller, M., & Socha-Dietrich, K. (2020). Reassessing private practice in public hospitals in Ireland: An overview of OECD experiences.

Ofer, G., Greenstein, M. and Rosen, B., 2006. Selecting your Surgeon: the Private–Public Mix in Public Hospitals in Jerusalem; Considerations of Efficiency and Equity. CESifo Economic Studies, 52(3), pp.513-547.

Olivella, P. (2003). Shifting public-health-sector waiting lists to the private sector. European Journal of Political Economy, 19(1), 103-132.

Paris, V., Devaux, M., & Wei, L. (2010). Health systems institutional characteristics: a survey of 29 OECD countries.

Raffel, M.W. ed., 2007, Health care and reform in industrialized countries. Penn State Press.

Siciliani, L., & Hurst, J. (2005). Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries. Health policy, 72(2), 201-215.

Sommersguter-Reichmann, M. and Stepan, A., 2017. Hospital physician payment mechanisms in Austria: do they provide gateways to institutional corruption?. Health Economics Review, 7(1), pp.1-13.

Sutton, M. and Long, J., 2014. International models of career structure for medical specialists, other tenured medical posts and general practitioners in public (funded) health systems.

The Competition Authority, 2005. Guidance in respect of Collective Negotiations relating to the Setting of Medical fees [online]. Available from: https://www.ccpc.ie/business/wp-content/uploads/sites/3/2017/04/Medical-Fees-Guidance-Note.pdf [accessed 09 June 2022]

Topeli, A., 2010. Intensive Care Medicine in Turkey. HealthManagement [online]. Available from: https://healthmanagement.org/c/it/issuearticle/intensive-care-medicine-in-turkey [accessed 11 May 2022]

Varian, Hal R. (1974) "Equity, envy, and efficiency." Journal of Economic Theory, 9(1): 63-91.

Weil, L.G., Nun, G.B. and McKee, M., 2013. Recent physician strike in Israel: a health system under stress?. Israel journal of health policy research, 2(1), pp.1-11.

APPENDIX

Appendix A: List of variables

General	**Exogenous variables**
$N > 0$	Number of patients seeking care
$K > 0$	Number of consultants
$q \in [0,1]$	Proportion of senior consultants
$\beta \in [0,1]$	Quality of public care.
$\mu \in \{0,1\}$	Patient's value to being treated. Accurate diagnosis leads to $\mu = 1$.
$C > 0$	Patient's cost when remaining on waiting lists.
$\theta \in [0, \overline{\theta}]$	Patient's willingness to pay for a private consultation.
$\gamma \in [0,1]$ $\gamma_i \in [0,1]$	Quality of consultant/Probability of issuing a correct diagnosis/ expected utility of care.
$i \in \{H, L\}$	Variable indexed by i in Section 5.
$v > 0$	Consultant's gratification from issuing a correct diagnosis.
$\lambda \in]0,1[$	Shadow cost of public funds.
Contractual	**Endogenous variables**
$s > 0$	Private fee.
$t > 0$	Fixed monetary transfer.
$\hat{\theta} = s - (1-\beta)\gamma$	Threshold above which patients request private care.
$n > 0$ $n_i > 0, i = H, L$	Number of patients treated by a consultant, possibly indexed by $i = H, L$.
$r(s) = s\left(1 - F(\hat{\theta})\right)$	Expected private income per patient.

Appendix B: Private Practice in OECD Countries

Countries	Private Practice is allowed.	Private Practice within public hospitals is allowed.
Australia	√	√
Austria	√	√
Belgium	√	√
Canada	×	×
Chile	√	√
Costa Rica	○	
Czech Republic	√	×
Denmark	√	×
Finland	√	×

France	√	√
Germany	√	√
Greece	○	√
Hungary		×
Iceland	√	×
Ireland	√	√
Israel	√	√
Italy	○	√
Japan	○	√
Korea		×
Latvia	√	
Luxembourg		√
Mexico		×
Netherlands	√	√
New Zealand		×
Norway	√	×
Poland	√	×
Portugal	○	√
Slovenia	○	
Spain	√	×
Sweden	○	Was √, then ×
Switzerland	√	√
Turkey		Was √, then ×
United Kingdom	√	√

Note: The tick √ stands for "Yes, always". The circle ○ stands for "Yes, in some circumstances only". The cross × stands for "No". The question mark ? stands for "Unclear, please refer to footnote". The blank cell refers to "No information found". We exclude American, Colombia, Estonia, Lithuania, and Slovak Republic, because there is no information about these countries from OECD survey.

List of sources of information in Appendix B

Main sources: OECD Survey on health system characteristics 2008-2009 and 2016.
Chile: OECD Health System Characteristics Survey 2016, Question 31 (Comments)
Finland: Please see Sutton and Long (2014) and Garattini and Padula (2018).
Germany: Please see Garattini and Padula (2018).

Ireland: Please see Irish consultant's contract 2008 from Health Service Executive (2019).
Israel: Please see Ofer et al. (2006).
Italy: Please see Garattini and Padula (2018).
Norway: Please see Garattini and Padula (2018).
Sweden: According to Immergut and Comisso (1992), senior consultants could treat private patients even within public hospitals. According to OECD Survey on health system characteristics 2008-2009, Sweden no longer allows this private provision of care.
Turkey: Please see Topeli (2010) and World Bank Group report by Aran and Rokx (2014), new arrangements were introduced in 2010, which required public doctors to practice exclusively in the public sector.

Appendix 1: The demand for private care and the private revenue.

Under Assumption 1 the function $r_i(s_i)$ is continuous and differentiable almost everywhere since we have

$$r_i'(s_i) = \begin{cases} 1 \text{ for } s_i \leq (1-\beta)\gamma_i, \\ \left(1 - F(\hat{\theta})\right) - s_i f(\hat{\theta}) \text{ for } s > (1-\beta)\gamma_i. \end{cases}$$

As $f(0) = \varepsilon$ where ε is arbitrarily small, we have $\lim_{s_i \to (1-\beta)\gamma_i} \left(1 - F(\hat{\theta})\right) - s_i f(\hat{\theta}) = 1$.

Furthermore, there exists $\bar{s}_i > (1-\beta)\gamma_i$ that maximizes consultant i's revenue from private care and it is such that

$$\frac{d}{ds_i} r_i(s_i) = 0 \text{ at } s_i = \bar{s}_i.$$

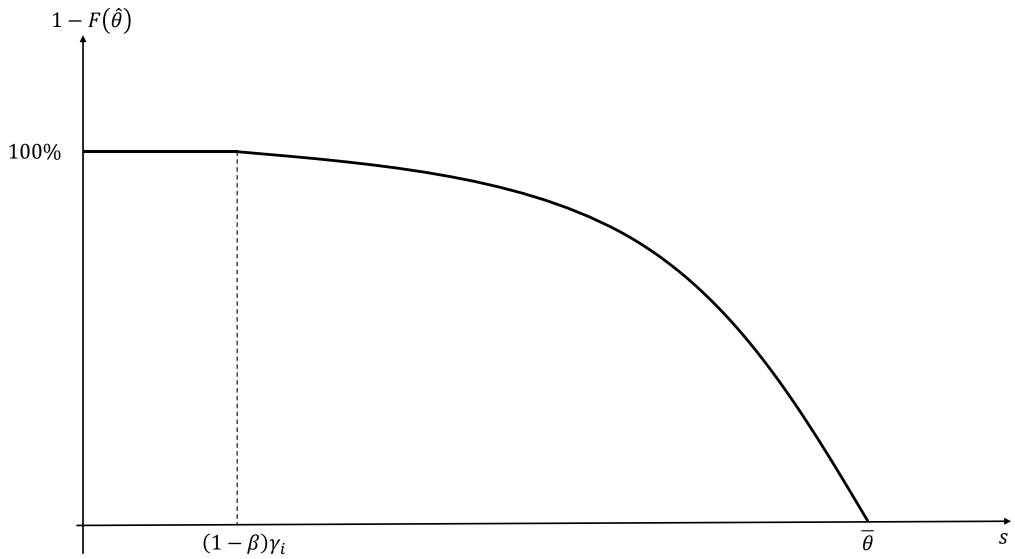

Figure 4: Private Fee as a Function of the Distribution of Patients

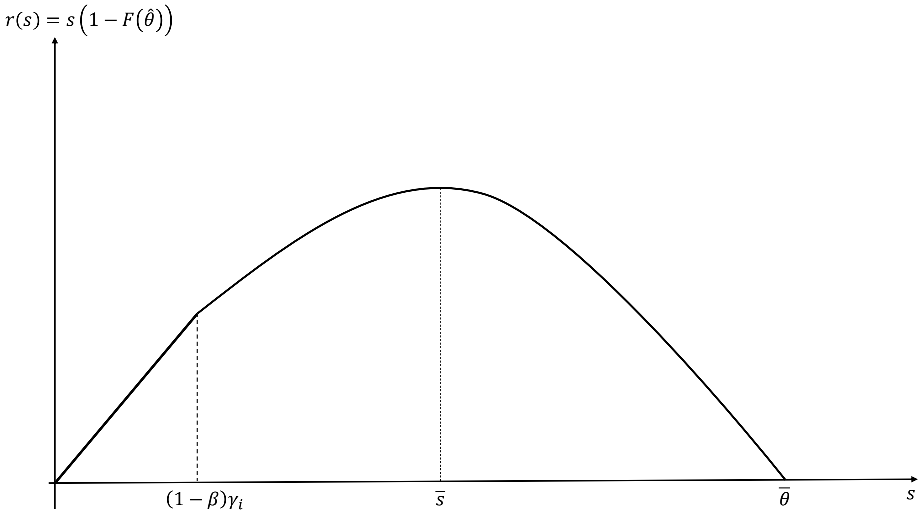

Figure 5: Existence of Revenue Maximised Private Fee

Appendix 2: Proof of Proposition 1.

The first derivative of the total surplus, when $\hat{\theta} = s - (1-\beta)\gamma > 0$, is given by

$$\frac{dTS}{ds} = \left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1 - F(x)\big)dx\right] r'(s) + (1+\lambda)n^* r'(s) - n^*\left(1 - F(\hat{\theta})\right),$$

where n^* is given by (4). Notice that we have

$$\lim_{s\to(1-\beta)\gamma} \frac{dTS}{ds} = \beta\gamma + C + \int_{0}^{\overline{\theta}} \big(1 - F(x)\big)dx + \lambda n^* > 0.$$ [19]

We can re-write the first order condition as

$$r'(s^*)\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1 - F(x)\big)dx\right] - n^*\left[\left(1 - F(\hat{\theta})\right) - (1+\lambda)r'(s^*)\right] = 0,$$

At the solution we must have

$$r'(s^*)\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1 - F(x)\big)dx\right] > 0 \to r'(s^*) > 0.$$

Thus, the private revenue is capped.

The second derivative of the total surplus can be written as sum of negative terms:

$$\frac{d^2TS}{ds^2} = r''(s)\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1 - F(x)\big)dx\right] - r'(s)\left(1 - F(\hat{\theta})\right)$$

$$-r'(s)\left[\left(1 - F(\hat{\theta})\right) - (1+\lambda)r'(s)\right] - n^*\left[\lambda f(\hat{\theta}) + (1+\lambda)\frac{d}{ds} s f(\hat{\theta})\right] < 0.$$

Given that $r'(s^*) > 0$, and $r''(s^*) < 0$, the first two terms are non-positive. Note from the first order condition that we must have

$$n^*\left[\left(1 - F(\hat{\theta})\right) - (1+\lambda)r'(s^*)\right] \geq 0 \to r'(s^*) \leq \frac{\left(1 - F(\hat{\theta})\right)}{(1+\lambda)}.$$

Therefore, the third term is non-positive. Finally, since $\frac{d}{ds} s f(\hat{\theta}) > 0$ the last term is also non-positive.

Appendix 3: Proof of Lemma 2.

[19] One can easily show that the total surplus is continuous at $s = (1-\beta)\gamma$ and it is differentiable a.e..

From the proof of Proposition 1 we know that the optimal private fee solves

$$H(s^*) = \left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} (1 - F(x))dx\right] r'(s^*) + n^*(1+\lambda)r'(s^*) \\ - n^*\left(1 - F(\hat{\theta})\right) = 0,$$

where n^* is given by (4).

Given any $x \in \{\nu, C, \beta, \gamma\}$ we have

$$\left.\frac{\partial H}{\partial s}\right|_{s^*} \frac{ds^*}{dx} + \left.\frac{\partial H}{\partial x}\right|_{s^*} = 0.$$

The second order condition holds at s^* so that $\left.\frac{\partial H}{\partial s}\right|_{s^*} < 0$. Therefore the sign of $\frac{ds^*}{dx}$ is the same as the sign of $\left.\frac{\partial H}{\partial x}\right|_{s^*}$.

- Parameter of interest: ν.

We have

$$\left.\frac{\partial H}{\partial \nu}\right|_{s^*} = -\gamma\left[\left(1 - F(\hat{\theta})\right) - (1+\lambda)r'(s^*)\right] < 0.$$ [20]

This inequality indicates that when the intrinsic motivation of consultants is higher, the HA should propose a lower private fee to respond.

- Parameter of interest: C.

We have

$$\left.\frac{\partial H}{\partial C}\right|_{s^*} = r'(s^*) > 0.$$

This inequality indicates that when the cost associated with waiting list is higher, the HA should propose a higher private fee to respond.

- Parameter of interest: β.

Simple calculations show that, at the solution we have

[20] In Appendix 2 we show that at the solution we must have $\left[\left(1 - F(\hat{\theta})\right) - (1+\lambda)r'(s^*)\right] > 0$.

$$\left.\frac{\partial H}{\partial \beta}\right|_{s^*} = \gamma r'(s^*)F(\hat{\theta}) + n^*\gamma f(\hat{\theta}) - \gamma \frac{dr(s^*)}{d\hat{\theta}}\Big(\big(1-F(\hat{\theta})\big) - (1+\lambda)r'(s^*)\Big)$$
$$+\gamma \frac{dr'(s^*)}{d\hat{\theta}}\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1-F(x)\big)dx\right].$$

We have

$$\gamma r'(s^*)F(\hat{\theta}) + n^*\gamma f(\hat{\theta}) - \gamma \frac{dr(s^*)}{d\hat{\theta}}\Big(\big(1-F(\hat{\theta})\big) - (1+\lambda)r'(s^*)\Big) > 0.$$

However, we also have

$$\frac{dr'(s^*)}{d\hat{\theta}}\left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1-F(x)\big)dx\right] < 0.$$

Therefore, the overall sign of $\left.\frac{\partial H}{\partial \beta}\right|_{s^*}$ is not obvious.

- Parameter of interest: γ.

Simple calculations show that

$$\begin{aligned}\left.\frac{\partial H}{\partial \gamma}\right|_{s^*} &= \big[1-F(\hat{\theta}) + F(\hat{\theta})\beta\big]r'(s^*) \\ &\quad - \left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1-F(x)\big)dx\right](1-\beta)\ \frac{dr'(s^*)}{d\hat{\theta}} \\ &\quad - (1+\lambda)n^*(1-\beta)\ \frac{dr'(s^*)}{d\hat{\theta}} \\ &\quad - \left[\nu - \frac{dr(s^*)}{d\hat{\theta}}(1-\beta)\right]\Big(1-F(\hat{\theta}) - (1+\lambda)r'(s^*)\Big) \\ &\quad - n^* f(\hat{\theta})(1-\beta).\end{aligned}$$

We have

$$\big[1-F(\hat{\theta}) + F(\hat{\theta})\beta\big]r'(s^*) - \left[\beta\gamma + C + \int_{x=\hat{\theta}}^{\overline{\theta}} \big(1-F(x)\big)dx\right](1-\beta)\ \frac{dr'(s^*)}{d\hat{\theta}}$$
$$- (1+\lambda)n^*(1-\beta)\ \frac{dr'(s^*)}{d\hat{\theta}} > 0.$$

However, we also have

$$-\left[v - \frac{dr(s^*)}{d\hat{\theta}}(1-\beta)\right]\left(1 - F(\hat{\theta}) - (1+\lambda)r'(s^*)\right) - n^* f(\hat{\theta})(1-\beta) < 0.$$

Therefore, the overall sign of $\left.\frac{\partial H}{\partial \gamma}\right|_{s^*}$ is not obvious. ■

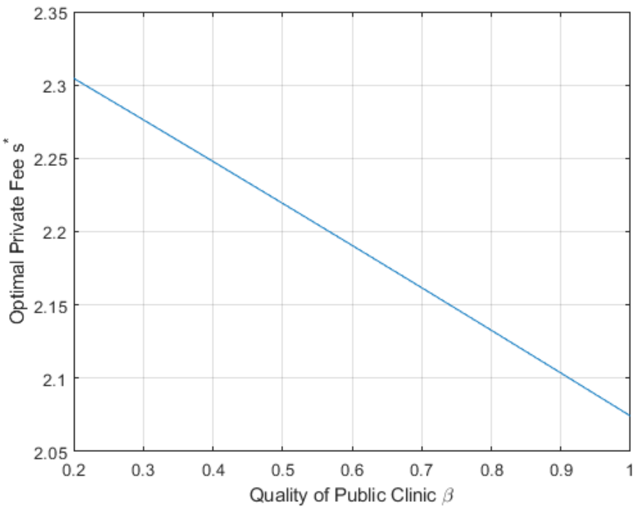

Figure 2: Optimal private fee as a function of the Quality of Public Clinic

(We let $v = 5$, $\lambda = 0.2$, $\overline{\theta} = 5$, $C = 10$, and $\gamma = 0.8$.)

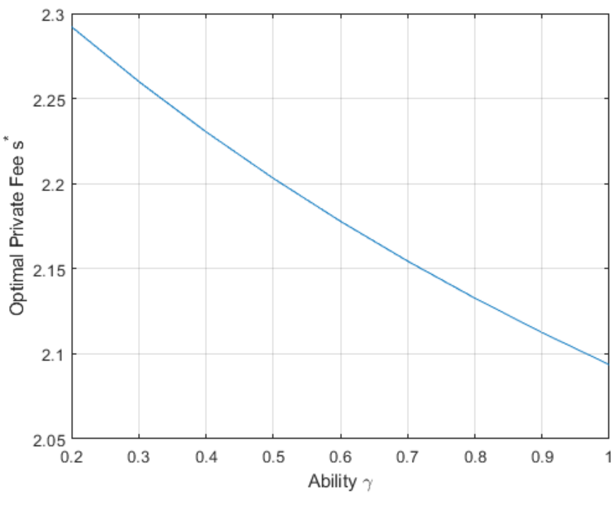

Figure 3: Optimal private fee as a function of the Ability of Consultants

(We let $v = 5$, $\lambda = 0.2$, $\overline{\theta} = 5$, $C = 10$, and $\beta = 0.8$.)

Appendix 4: proof of Lemma 3.

The envy-free constraints can be written as

$$EF_H: t_H - t_L \geq \frac{1}{2}[\gamma_H \nu + r_H(s_L)]^2 - \frac{1}{2}[\gamma_H \nu + r_H(s_H)]^2,$$

$$EF_L: t_H - t_L \leq \frac{1}{2}[\gamma_L \nu + r_L(s_L)]^2 - \frac{1}{2}[\gamma_L \nu + r_L(s_H)]^2.$$

In words, the constraint EF_H determines a lower bound for $(t_H - t_L)$, while EF_L specifies an upper bound. Thus, *for both constraints to hold*, we must have

$$[\gamma_H \nu + r_H(s_L)]^2 - [\gamma_H \nu + r_H(s_H)]^2 \leq [\gamma_L \nu + r_L(s_L)]^2 - [\gamma_L \nu + r_L(s_H)]^2$$

$$\begin{aligned} \Leftrightarrow [r_H(s_L) - r_H(s_H)][2\gamma_H \nu + r_H(s_H) + r_H(s_L)] \\ \leq [r_L(s_L) - r_L(s_H)][2\gamma_L \nu + r_L(s_L) + r_L(s_H)] \end{aligned}$$

$$\begin{aligned} \Leftrightarrow [r_H(s_H) - r_H(s_L)][2\gamma_H \nu + r_H(s_H) + r_H(s_L)] \\ \geq [r_L(s_H) - r_L(s_L)][2\gamma_L \nu + r_L(s_L) + r_L(s_H)]. \end{aligned}$$

Let $EF(s_L, s_H)$ be given by

$$\begin{aligned} EF(s_L, s_H) = [r_L(s_H) - r_L(s_L)][2\gamma_L \nu + r_L(s_L) + r_L(s_H)] \\ - [r_H(s_H) - r_H(s_L)][2\gamma_H \nu + r_H(s_H) + r_H(s_L)]. \end{aligned}$$

The variables (s_L, s_H) must be set such that $EF(s_L, s_H) \leq 0$ for both envy-free constraints hold (strictly or not).

- Note that $EF(s, s) = 0$ for all $s_L = s_H = s$.
- Note that the partial derivatives of the function $EF(s_L, s_H)$ are:

$$\frac{\partial EF}{\partial s_L} = 2r_H'(s_L)[\gamma_H \nu + r_H(s_L)] - 2r_L'(s_L)[\gamma_L \nu + r_L(s_L)] \geq 0,$$

and

$$\frac{\partial EF}{\partial s_H} = 2r_L'(s_H)\big(\gamma_L \nu + r_L(s_H)\big) - 2r_H'(s_H)\big(\gamma_H \nu + r_H(s_H)\big) \leq 0.$$

Therefore, it follows that $EF(s_L, s_H) \leq 0 \Leftrightarrow s_L \leq s_H$.

- ($\Rightarrow$) Consider any (s_L, s_H) such that $EF(s_L, s_H) \leq 0 \Rightarrow EF(s_L, s_H) \leq EF(s, s), s \in \{s_L, s_H\}$. Given the sign of the partial derivatives, $EF(s_L, s_H) \leq EF(s_L, s_L) \Rightarrow s_H \leq s_L$. Similarly, $EF(s_L, s_H) \leq EF(s_H, s_H) \Rightarrow s_H \leq s_L$.
- ($\Leftarrow$) Consider any (s_L, s_H) such that $s_L \leq s_H$. Given the signs of the partial derivatives we have $s_L \leq s_H \Rightarrow EF(s_L, s_H) \leq EF(s, s) = 0, s \in \{s_L, s_H\}$.

Appendix 5: Proof of proposition 2

As discussed in the text, to reduce the weight of public expenditures, the fixed transfer t_L is set such that the participation constraint holds for the L-type and t_H is such that the envy-free constraint holds for the H-type.

The HA must solve

$$\max_{t_L, t_H, s_L, s_H} K[qTS_{HH} + (1-q)TS_{LL}] - NC$$

subject to $(s_L - s_H) \leq 0$, where

$$TS_{ii} = n_{ii}^*\big(C + PW(\gamma_i, s_i)\big) + \frac{1}{2}(1+\lambda)(n_{ii}^*)^2 - \lambda U(n_{ii}^*),$$

$$U(n_{LL}^*) = U_R,\ U(n_{HH}^*) = U_R + R(s_L),$$

n_{ij}^* is given by (11) in the text.

Using Lagrange's method, we have

$$\mathcal{L} = K[qTS_{HH} + (1-q)TS_{LL}] - NC - \delta(s_L - s_H).$$

The first order conditions (FOC) are such that we must have

$$\frac{\partial \mathcal{L}}{\partial s_H} = qKH(s_H; \gamma_H) + \delta = 0, \tag{C1}$$

$$\frac{\partial \mathcal{L}}{\partial s_L} = (1-q)KH(s_L; \gamma_L) - \lambda qK\frac{dR}{ds_L} - \delta = 0, \tag{C2}$$

$$\delta(s_L - s_H) = 0, \tag{C3}$$
$$(s_L - s_H) \leq 0 \text{ and } \delta \geq 0,$$

where $H(s_i; \gamma_i) = H(s_i)|_{\gamma=\gamma_i}$, and $H(s_i)$ is given by (10) in the text.

Let s_H^* and s_L^* denote the optimal fees under first-degree discrimination. They are defined such that $H(s_H^*; \gamma_H) = H(s_L^*; \gamma_L) = 0$.

- **Unconstrained solution ($\boldsymbol{\delta = 0}$, $\boldsymbol{(s_L - s_H) < 0}$)**

Let s_H^{**} and s_L^{**} denote the optimal fees in this situation. When $\delta = 0$ the optimisation problem is separable in s_H and s_L as (C1) depends only on s_H and (C2) only on s_L. Therefore:

(i) The unconstrained candidate for s_H is such that $H(s_H; \gamma_H) = 0 \Leftrightarrow s_H^{**} = s_H^* > (1-\beta)\gamma_H$. We know that the second order condition holds at that solution.

(ii) The unconstrained candidate for s_L, is such that we have $s_L^{**} < s_L^*$ because

$$\left.\frac{\partial \mathcal{L}}{\partial s_L}\right|_{s_L^*} = -\lambda q K \left.\frac{dR}{ds_L}\right|_{s_L^*} < 0.$$

Let $TSL(s_L)$ capture the function that the variable s_L must maximize, we have

$$TSL(s_L) = (1-q)\left[\big(\gamma_L \nu + r_L(s_L)\big)\big(C + PW(\gamma_L, s_L)\big) + \frac{1}{2}(1+\lambda)\big(\gamma_L \nu + r_L(s_L)\big)^2\right] - \lambda q R(s_L).$$

The variable s_L^{**} is the solution to

$$\frac{dTSL}{ds_L} = (1-q)KH(s_L^{**};\gamma_L) - \lambda q K \left.\frac{dR}{ds_L}\right|_{s_L^{**}} = 0.$$

Depending on the exogenous variables two possibilities arise.

<u>Possibility 1:</u> $s_L^{**} \le (1-\beta)\gamma_L$ in which case all patients request to see type-L consultants privately. This is a solution provided

$$(1-q)\left(C + \frac{1}{2}\overline{\theta} + \gamma_L\big(1 + \lambda\nu - (1-\beta)(1-\lambda)\big)\right) \le \lambda q \nu(\gamma_H - \gamma_L),$$

meaning, among other things, that q is large enough or that $(\gamma_H - \gamma_L)$ is large enough. In this case, it is easy to show that for all $s_L \in [0, (1-\beta)\gamma_L\,]$ the function $TSL(s_L)$ is strictly concave and the solution forms a maximum.

<u>Possibility 2:</u> $(1-\beta)\gamma_L < s_L^{**}$. In this case we have

$$PW(\gamma_L, s_L) = \beta\gamma_L + \frac{1}{\overline{\theta}} \int_{s_L-(1-\beta)\gamma_L}^{\overline{\theta}} \big(\overline{\theta} - x\big)dx.$$

and

$$r_L(s_L) = \frac{s_L\big(\overline{\theta} - s_L + (1-\beta)\gamma_L\big)}{\overline{\theta}} \text{ and } r_H(s_L) = \frac{1}{\overline{\theta}}\big(\overline{\theta} - \hat{\theta}_{HL}\big).$$

The first order condition is such that

$$(1-q)[C + PW(\gamma_L, s_L) + (1+\lambda)n_{LL}]\frac{dr_L}{ds_L} + (1-q)n_{LL}\frac{dPW}{ds_L} - \lambda q \frac{dR}{ds_L} = 0.$$

To establish that the second order condition holds at the solution, we take the derivative of the expression on the left-hand side, multiply it by $\frac{dr_L}{ds_L}$ and use the first order condition to replace $(1-q)\left[C+PW(\gamma_L,s_L)+(1+\lambda)\left(\gamma_L\nu+r_L(s_L)\right)\right]\frac{dr_L}{ds_L}$. We must then show that expression E below is non-positive:

$$E=-\frac{2}{\overline{\theta}}\left[\lambda q\frac{dR}{ds_L}-(1-q)n_{LL}\frac{dPW}{ds_L}\right]+(1-q)\left(\frac{dr_L}{ds_L}\right)^2\frac{dPW}{ds_L}+(1-q)\frac{1}{\overline{\theta}}n_{LL}\frac{dr_L}{ds_L}-\lambda q\frac{dr_L}{ds_L}\frac{d^2R}{ds_L^2},$$

Note that since $\frac{dr_H(s_L)}{ds_L}\geq\frac{dr_L(s_L)}{ds_L}$ and since $\frac{d^2r_H(s_L)}{ds_L^2}\geq\frac{d^2r_L(s_L)}{ds_L^2}$ we have

$$\frac{dR}{ds_L}\geq\frac{dr_L(s_L)}{ds_L}\left((\gamma_H-\gamma_L)\nu+r_H(s_L)-r_L(s_L)\right)$$

and

$$\frac{d^2R}{ds_L^2}\geq\frac{d^2r_L(s_L)}{ds_L^2}\left((\gamma_H-\gamma_L)\nu+r_H(s_L)-r_L(s_L)\right).$$

Using the two inequalities above, we have

$$E\leq(1-q)\left(\frac{dr_L}{ds_L}\right)^2\frac{dPW}{ds_L}-(1-q)\frac{n_{LL}}{\overline{\theta}^2}\left[\overline{\theta}+(1-\beta)\gamma_L\right]<0.$$

Finally, we must identify whether the unconstrained candidate (s_H^*,s_L^{**}) that we have identified is always such that the envy-free constraint $(s_L^{**}-s_H^*)<0$ holds.

- If the optimal fees that would prevail under first-degree discrimination are such that $s_L^*\leq s_H^*$ then the solution depicted above is an optimal solution since $s_L^{**}<s_L^*$.
- If, however, $s_H^*<s_L^*$ we could have a situation where $s_L^{**}\leq s_H^*<s_L^*$ and that would mean that (s_H^*,s_L^{**}) would still be optimal. However, if we had $s_H^*<s_L^{**}<s_L^*$ the candidate (s_H^*,s_L^{**}) does not lead to envy-free contracts and we must look for an constrained solution.

(iii) Constrained solution $(\boldsymbol{\delta>0},(\boldsymbol{s_L-s_H})=\mathbf{0})$

Assume that the fees are such that where $s_H^*<s_L^{**}<s_L^*$ so that the previous solution fails to satisfy the envy-free constraint. We must look for a constrained solution that is characterized such that $s_L=s_H=s^{**}$.

When the variable θ is uniformly distributed, the total surplus for any given type can be shown to be a strictly concave function in s as we have

$$\frac{dH}{ds} = -\frac{2}{\overline{\theta}}\big(C + PW(s,\gamma)\big) - \frac{\big(\gamma\nu + r(s)\big)}{\overline{\theta}}(1+2\lambda) + (1+\lambda)\left(\frac{dr}{ds}\right)^2 + 2\frac{dr}{ds}\frac{dPW}{ds} < 0.$$ [21]

Since we must have $\delta > 0$, at the solution s^{**} characterised by (C1) and (C2) we have

$$H(s^{**};\gamma_H) < 0 \Longrightarrow s^{**} > s_H^*, \tag{C4}$$
$$H(s^{**};\gamma_L) > 0 \Longrightarrow s^{**} < s_L^*. \tag{C5}$$

Conditions (C1) and (C2) are satisfied at $s_L = s_H = s^{**}$ provided

$$M(s^{**}) = qH(s^{**};\gamma_H) + (1-q)H(s^{**};\gamma_L) - \lambda q \frac{dR}{ds_L}\bigg|_{s^{**}} = 0.$$

Notice that $M(s_H^*) > 0$ since we are considering situations where $s_H^* < s_L^{**}$ and, for the same reason, we have $M(s_L^{**}) < 0$. Therefore, there exists at least one value $s^{**} \in\,]s_H^*, s_L^{**}[$ which satisfies $M(s^{**}) = 0$.

Let $s^{**} \in [s_H^*, s_L^{**}]$ and the associated $\delta > 0$ be the solution to (C1) and (C2). Let us prove that the second order condition holds at such values. Recall that, for the contracts to satisfy the envy-free constraints we must have $EF(s_L, s_H) \leq 0 \Leftrightarrow (s_L - s_H) \leq 0$. We will show that when we set $s_i = s^{**}$ we prove that setting $s_j = s^{**}$ is optimal within the range of envy -free private rates. Recall that the function $H(.)$ is the derivative of the total surplus.

- Assume that $s_L = s^{**}$. Since $\delta > 0$, (C1) holds where the total surplus is decreasing in s_H at s^{**}. Furthermore, since $s^{**} > s_H^*$, and due to concavity, the total surplus is decreasing over the relevant envy-free range $s_H \geq s^{**}$. Hence, among the constrained contracts setting $s_H = s^{**}$ reaches a maximum.
- Assume that $s_H = s^{**}$. Since $\delta > 0$, (C2) indicates that the total surplus is increasing in s_L at s^{**}. Furthermore, since $s^{**} < s_L^*$, the total surplus is increasing over the relevant envy-free range $s_L \leq s^{**}$. Hence, among the constrained contracts setting $s_L = s^{**}$ reaches a maximum. ■

[21] To prove the above inequality, one should evaluate the expression at $(1+\lambda)\left(\frac{dr}{ds}\right)^2 = 2\left(\frac{dr}{ds}\right)^2$ which gives an upper bound as $\lambda \in (0,1)$. This upper bound is clearly negative.