Public and private hospitals, waiting times, and redistribution*

Chiara Canta[†] and Marie-Louise Leroux[‡]

June 1, 2012

Abstract

We study an economy in which agents all have to undertake a health treatment but differ in their ability (i.e. their income) to pay for it. We design the optimal income taxation scheme when agents can either undertake the treatment at a public hospital in which they have to bear a waiting cost, or use a private facility in which there is no waiting line but for which they have to pay a higher fee. We show that, without further intervention of the social planner, agents fail to internalize their impact on the congestion of the public hospital, and waiting times are then higher than optimal. The first-best optimum can however be implemented through lump-sum taxation and a subsidy on the private hospital fee. We also derive the optimal income taxation policy when incomes are not observable. Waiting times are used as an instrument to relax incentive constraints. If the social planner can assign patients to hospitals, the second best can be implemented by a menu of lotteries where the probability of going to the public hospital is decreasing with the level of income. Waiting times are distorted up in order to avoid mimicking by high-income agents. If agents can freely choose the hospital, the only instruments available are a lump-sum tax and a subsidy on the private hospital price. The optimal subsidy is negative whenever redistribution concerns are high enough.

JEL Codes: H21, H23, H44, I11.

Keywords: Optimal taxation; Mixed health care systems; Waiting times; Income redistribution.

*We would like to thank Pierre Pestieau and Philippe De Donder for their valuable comments on this paper.

[†]CORE, Universitæ catholique de Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. CESifo (Germany). E-mail:chiara.canta@uclouvain.be

[‡]Département des sciences économiques, ESG UQAM, C.P. 8888, Succ. Centre-ville, Montréal, QC, Canada H3C 3P8. CESifo (Germany) and CORE (Université catholique de Louvain, Belgium). E-mail: leroux.marie-louise@uqam.ca

1 Introduction

In many countries (such as the UK, Italy, and Australia) health services are provided by a tax funded universal health system. However, patients can choose to get the treatment from a private provider, paying the price of the treatment out of pocket. Such systems are often referred to as mixed systems. The reasons why some patients are willing to pay for treatment at a private hospital are related to higher perceived quality in the private sector or to longer waiting times in the public sector. The latter reason is particularly important. In the case of the UK, Besley et al. (1999) show that waiting times at public hospitals are the only significant variable explaining the demand for private health care insurance. Nevertheless, whether agents have the possibility to go to the private or the public hospital crucially depends on their ability to pay for the service and thus on their income. In this paper, we are interested in designing the optimal income taxation policy when agents, who all need to undertake a health treatment, only differ in their income. The novelty of this paper consists in considering a congested public sector, so that waiting times are endogenous and depend on the number of agents treated by the public hospital. If agents are free to choose between public and private hospitals, the taxation scheme in place affects waiting times. In turn, waiting times affect the feasible set of taxation schedules. In this framework, we show that in some cases waiting times can be a useful tool for redistribution when the social planner cannot observe individual incomes.

We use a one-period model in which agents have to undertake the same health treatment, represented by one unit of medical service, but have different incomes. To get the treatment, they can either go to a public or a private hospital. In the private hospital, there is no waiting time, but patients have to pay a fixed price. On the contrary, the public hospital is financed through taxation and delivers health care services for free. However, capacity constraints lead to waiting costs, which create some disutility to the agents. We assume a fixed public hospital capacity; this seems a reasonable assumption, at least in the short run. Thus, the waiting time is endogenously determined and depends on the number of agents seeking treatment in the public hospital. It is modeled as the difference between the demand for the public hospital and the capacity of the public hospital, normalized by the public capacity. Depending on their income, agents will have a different willingness to pay for health services and we find that, at the laissez-faire, low-income agents go to the public hospital while high-income ones go to the private one. We then characterize the optimal income taxation and allocation of patients across hospitals.

The first best policy consists in offering agents a menu of lotteries specifying the probability of going to the public hospital and the taxes to be payed if assigned to the public or the private hospital, respectively. The optimal allocation is such that the marginal utility of consumption is equalized across agents. Each agent is assigned the same positive probability to receive care in the private hospital. The optimal sorting of patients between the public and the private hospital is such that fewer agents should go to the public hospital than in the laissez-faire. This corresponds to the internalization of a simple hospital choice externality: at the laissez-faire, agents do not internalize the effect of their choice on the waiting time bared by all the other agents going to the public hospital. If the social planner cannot directly assign agents to one or the other facility, we show that the first best allocation can also be decentralized by individualized transfers and a subsidy on the private treatment.

We then analyze a setting in which the social planner cannot observe agents' income but can still assign patients to hospitals. In this case, the probability of being assigned to the public hospital depends on their reported income and is used as a mean to foster redistribution. Agents with a low income have a higher probability of being assigned to the private hospital with respect to the first best, while the opposite is true for high-income agents. The waiting time is also distorted up in order to avoid mimicking from high income agents. The optimal taxes are such that consumption are the same irrespective on the hospital the patient is assigned to ex post.

We finally turn to the case in which the social planner cannot assign patients to hospitals. In such case, the tax schedule can only be conditional on the hospital choice, and consists of a lump-sum tax and of a linear subsidy on private health services. The optimal subsidy can be positive or negative depending on the size of redistributive and efficiency concerns. If the hospital choice externality is very high with respect to redistribution concerns, the optimal subsidy is positive. If redistribution concerns dominate, the subsidy turns into a tax, since only high-income agents use the private hospital. Contrary to the previous literature, we find that redistribution through a tax on the use of the private hospital may be limited by the corresponding increase in the waiting time in the public hospital.

Our paper can be related to the vast literature on optimal income redistribution and on the public provision of private goods as a way to redistribute resources among agents. Besley and Coate (1991) first showed that, if redistribution is not perfect, universal provision of private goods, financed by a linear head tax, can enhance redistribution. Intuitively, the rich pay for the public provision but buy a service of better quality from private firms and only the poor use the public facility. Also in our model rich agents may contribute to the public facility but go the private hospital in which there is no waiting cost, so that public provision of health care is also an in-kind transfer to the poor. However, our model differs from theirs in many respects. First of all, we consider a continuum of income types. Second, we analyze an optimal tax problem in which the good used to redistribute in-kind suffers of congestion problems. Finally, we also study the optimal non linear income taxation scheme showing that, under asymmetric information, the social planner can adjust the quality of the public service (i.e. waiting costs) to relax incentive constraints. Marchand and Schroyen (2005) consider a framework in which agents differ in productivity and face a probability of illness. The public service is provided free of charge but agents face waiting times, which decrease labor supply. In their paper, waiting times are used as a rationing device that equates demand and supply in the public sector. They show that the size of the public sector is never small, and that a mixed system is optimal when the income inequality is sufficiently large. Contrary to us, they only consider a linear tax. Moreover, they assume that the social planner can directly control the size of the public sector, while in our paper, we let patients choose between the private and the public hospital. Hoel and Sæther (2003) assume that the waiting time is a quality variable that can be directly set by the social planner. In their framework, agents have the same income and they find that a positive waiting time and/or a subsidy on private services are optimal, while in our model a subsidy on the private service is not always optimal, because a tax enhances redistribution. Blomquist and Christiansen (1995) analyze the role of the public provision of private goods for redistribution under asymmetric information. Since part of the redistribution is done in-kind, mimicking from high-ability to low-ability agents is less attractive. In their model, the social planner optimally sets the quantity/quality of the good that is publicly provided. Cremer and Gahvari (1997) consider a similar setting with no restriction on the number of consumption goods and allowing for linear commodity taxes. They characterize the conditions under which in-kind transfers help redistribution, by affecting individual consumption levels and by relaxing the incentive compatibility constraints. Differently from these papers, we consider a framework where the quality of public health care depends on the number of public patients. Redistribution concerns may lead the social planner to increase or decrease congestion in the public service in order to avoid mimicking. Furthermore, we allow for a continuum of types. Finally, Boadway and Marchand (1995) study the optimality of uniform public provision on redistributive grounds but allow agents to supplement public expenditures with private provision. We focus on the case in which agents can consume at most one unit of the good and have to choose between the public and the private hospital.

All in all, the originality of our paper is to consider a waiting time which depends on the demand for the public service. Since agents do not internalize the impact of their decisions on the waiting time, the demand ends up being higher than optimal. At the same time, the social planner also has a rationale for distorting up the waiting time: this discourages mimicking of high-income agents and fosters income redistribution. This is in line with the literature: distorting the quality of the public hospital relaxes incentive compatibility constraints. However, in our model, distorting quality is possible only through distortion of the number of individuals benefiting from the redistribution in-kind.

The paper is organized as follows. In the following section, we give the basic set up and derive the laissez-faire. In Section 3, we characterize the first-best allocation and show how it can be decentralized. In the fourth section, we solve the second-best problem under asymmetric information on agents' income. Section 5 concludes.

2 The model

2.1 Basic assumptions

In our economy, there is a mass 1 of agents, who all need to undertake a health treatment and who differ only with respect to income, y. Income is a continuous variable with support $[\underline{y}, \bar{y}]$. In the following, we denote $f(y)$ and $F(y)$ as, respectively, the density and the cumulative functions of y.

Health care is provided through two hospitals: a public one and a private one. These hospitals are different with respect to two characteristics: the price agents pay and the time they have to wait for getting the treatment. If agents go to the public hospital, they pay a regulated price and they have to wait for the treatment. In the following, we denote w, the waiting time these agents have to incur and it takes the form

$$w = \begin{cases} \frac{\rho - \hat{x}}{\hat{x}} & if\ \rho \geq \hat{x} \\ 0 & if\ \rho < \hat{x} \end{cases} \tag{1}$$

where $\rho \in [0, 1]$ is the number of patients seeking care in the public hospital and $\hat{x} \leq 1$ is the capacity of the public hospital. We assume that the public hospital capacity is fixed. This assumption is reasonable since adjusting the capacity of a public hospital might be difficult, at least in the short run. In our model, we assume that the unit cost of treatment in the public sector hospital is constant and equal to k. Agents can also choose to avoid the waiting time and go to the private hospital. In that case, to get the same treatment, they have to pay a fixed price to the private hospital equal to p. In the following we assume that $k \leq p$. The justifications for this assumption are that either the private sector is not perfectly competitive or that it has some idle capacity in order to ensure no waiting times.[1]

This is a one-period model. Agents obtain utility from the consumption of a normal good, c and from the benefit of the treatment, h. The waiting time w reduces this benefit. For instance, agents may suffer during that period or end up being in a worse condition because of the time delay. The *net* benefit of the treatment in the public hospital is thus equal to $h - w$, while in the private hospital it is equal to h. For simplicity, we assume that agents have identical preferences that are quasi linear in the net benefit from the treatment:

$$u(c) + h - w,$$

where $u'(.) > 0$ and $u''(.) < 0$. For agents going to the public hospital, $w \geq 0$ while for those going in the private, $w = 0$. In our model, waiting times can be seen as the quality variable and thus, the public good is of a lower quality than the private one. We assume that h is positive and that $h > (1 - \hat{x})/\hat{x}$. This guarantees that all patients strictly prefer to get treatment than to stay untreated, even if the whole population gets the treatment at the public hospital. Note also that, for simplicity, we assume that h is the same for every agent and is independent of the choice of going to the public and the private hospital.[2]

[1] For a full discussion, see Hoel and Sæther (2003), p. 604.

[2] Alternative assumptions would certainly complicate the model. If h was agent-specific, we would have to deal with two dimensions of individuals heterogeneity; under asymmetric information, this can significantly complicate the model. If h was different in the private and

2.2 Laissez-faire

As a benchmark situation, let us first study the equilibrium allocation of agents when the social planner does not intervene. Patients of the public hospital pay the marginal cost k and incur the waiting cost, while patients in the private hospital pay p. We will refer to this situation as the laissez-faire one. To be precise, this is not exactly the laissez-faire since the presence of a public hospital is, by its own nature, a government's intervention. However, as we mentioned previously, we assume that the size of the public hospital, $\hat{x}$ is given and cannot be affected by the social planner.

The agent's problem consists in choosing whether to get the treatment in the public or in the private hospital. Since the decision to go to the private or to the public hospital is discrete, we have to compare the agent's indirect utility in one or the other situation. If he goes to the public hospital, his indirect utility is

$$u(y-k)+h-w,$$

where w is defined by (1).

On the contrary, if the patient seeks care in the private hospital, he has to pay the fee p of the private hospital but he avoids the waiting time so that his indirect utility is

$$u(y-p)+h.$$

Using the above indirect utility functions, and under the assumption that $k \leq p$, a patient with income y seeks care from the public hospital if and only if

$$\varphi(y)=u(y-k)-u(y-p)-w \geq 0.$$

Since $k \leq p$, $\varphi(y)$ is monotonically decreasing in y. Hence, at the laissez-faire, all patients having an income below the threshold $\tilde{y}^{LF}$, such that $\varphi(\tilde{y}^{LF})=0$ choose the public hospital, and the demand for the public hospital is $F(\tilde{y}^{LF})$. The income threshold $\tilde{y}^{LF}$ is implicitly defined by the following equation:

$$u(\tilde{y}^{LF}-k)-u(\tilde{y}^{LF}-p)=\frac{F(\tilde{y}^{LF})}{\hat{x}}-1. \tag{2}$$

Using the expression above, the demand for the public hospital in the laissez-

in the public hospital, the two hospitals would differ also in the efficiency of the treatment. These extensions are left for future work.

faire is

$$\rho^{LF} = \begin{cases} F(\tilde{y}^{LF}) = \hat{x}\left[u(\tilde{y}^{LF} - k) - u(\tilde{y}^{LF} - p) + 1\right] & if \;\; k \leq p \\ 0 & if \;\; k > p \end{cases}$$

It is straightforward to see that the demand for the public hospital when $k \leq p$ is always greater than the public capacity, and is increasing in the public capacity and in the price charged by the private hospital p but decreasing in k.

3 First best allocation

In this first section, we derive the optimal allocation when the social planner has perfect information on agents' income. We first derive the centralized solution. Second, we show how this solution can be decentralized when the social planner has no direct control on the hospital assignment.

3.1 Centralized solution

We assume that the social planner seeks to maximize the sum of a concave transformation Φ of agents' utility function. In our setting, the first best policy consists in the allocation of consumptions between agents with different incomes and in the repartition of the population between public and private hospitals.

The problem of the social planner consists then in choosing consumption levels for agents using the public and the private hospital, denoted respectively by c_y^{PU} and c_y^{PR}. Furthermore, the social planner sets the probability θ_y that an agent with income y is treated in the public hospital.[3] The program is the following:[4]

$$\max_{\theta_y, c_y^{PU}, c_y^{PR}} \int_{\underline{y}}^{\bar{y}} \Phi\left(\theta_y u(c_y^{PU}) + (1-\theta_y)u(c_y^{PR}) - \theta_y w + h\right) f(y)dy$$

$$\text{s.to} \quad \int_{\underline{y}}^{\bar{y}} \left[y - \theta_y(c_y^{PU} + k) - (1-\theta_y)(c_y^{PR} + p)\right] f(y)dy$$

$$w = \begin{cases} \frac{\int_{\underline{y}}^{\bar{y}} \theta_y f(y)dy}{\hat{x}} - 1 & if \;\; \int_{\underline{y}}^{\bar{y}} \theta_y f(y)dy \geq \hat{x} \\ 0 & otherwise \end{cases}$$

[3] Without redistribution concerns, one could rank the agents with respect to their willingness to pay for the public sector and decide how many of them should use the public facility. However, in the first best, it is not possible a priori to find who should go to the public hospital. This is why we set a probability to go to the public sector.

[4] We assume that the second order conditions are satisfied.

The first-order conditions with respect to c_y^{PU}, c_y^{PR} are, respectively

$$\frac{\partial \pounds}{\partial c_y^{PU}} = \Phi'\left(EU_y\right) u'\left(c_y^{PU}\right) - \lambda = 0 \ \forall y \tag{3}$$

$$\frac{\partial \pounds}{\partial c_y^{PR}} = \Phi'\left(EU_y\right) u'\left(c_y^{PR}\right) - \lambda = 0 \ \forall y, \tag{4}$$

where $EU_y = \theta_y u(c_y^{PU}) + (1-\theta_y)u(c_y^{PR}) - \theta_y w + h$ is the expected utility of an agent with income y and it depends on its probability to be assigned in the public hospital . The social marginal utilities of consumption are thus equalized for all individuals, irrespective of the hospital they seek care from so that these conditions lead to $c_y^{PR} = c_y^{PU} \ \ \forall y$.

The first order condition with respect to θ_y is

$$\begin{aligned}\frac{\partial \pounds}{\partial \theta_y} &= \Phi'\left(EU_y\right)\left[u(c_y^{PU}) - u(c_y^{PR}) - w\right] f\left(y\right) - \frac{dw}{d\theta_y}\int_{\underline{y}}^{\bar{y}} \theta_y \Phi'\left(EU_y\right) f(y)dy \\ &\quad - \lambda\left[c_y^{PU} + k - c_y^{PR} - p\right] f(y).\end{aligned} \tag{5}$$

Making use of the first order conditions on consumptions, we can rearrange the above first-order condition as

$$\frac{\partial \pounds}{\partial \theta_y} = -w\Phi'\left(EU_y\right) f\left(y\right) - \frac{dw}{d\theta_y}\int_{\underline{y}}^{\bar{y}} \theta_y \Phi'\left(EU_y\right) f(y)dy - \lambda\left[k-p\right] f(y). \tag{6}$$

It is optimal to set the probability to go to the public hospital such that the marginal cost of increasing the waiting time, i.e a decrease in utility for the agents going to the public hospital, equals the gain in efficiency due to the fact that more agents use the less expensive facility. Note that it is always optimal to set the waiting time greater than zero, since

$$\left.\frac{\partial \pounds}{\partial \theta_y}\right|_{w=0} = -\lambda\left[k-p\right] f(y) > 0,$$

under our assumption that $p - k \geq 0$. Setting (6) to zero, one obtains

$$w\Phi'\left(EU_y\right) = \lambda\left[p-k\right] - \frac{1}{\hat{x}}\int_{\underline{y}}^{\bar{y}} \theta_y \Phi'\left(EU_y\right) f(y)dy. \tag{7}$$

Since the right hand side of this equality is independent from y, $EU_y = EU$ $\forall y$. Replacing into the first order conditions with respect to consumptions, we conclude that

$$c_y^{PU} = c_y^{PR} = \bar{c}.$$

This, in turn implies that θ^* is independent of y. From (3), $\Phi'(EU)\,u'(\bar{c}) = \lambda$. and replacing into (7) we get

$$w\frac{\lambda}{u'(\bar{c})} = \lambda\,[p-k] - \frac{1}{\hat{x}}\int_{\underline{y}}^{\bar{y}} \theta^* \frac{\lambda}{u'(\bar{c})} f(y)dy \tag{8}$$

$$\iff \theta^* = \hat{x}\frac{u'(\bar{c})\,[p-k]+1}{2}. \tag{9}$$

This condition can be interpreted as follows. Since the public hospital is more efficient than the private one ($k < p$), it is always optimal that some patients get the treatment in the public hospital. The optimal probability $\theta^* > \hat{x}$ is such that the marginal increase in waiting times due to a higher public demand exactly equals the gains in efficiency. Whenever the hospital capacity is very large, or the public hospital is much more efficient than its private counterpart, it is optimal to have all patients treated in the public hospital. Corner solutions such that $\theta^* = 1$ are thus possible.

More generally, the optimal demand for the public hospital only depends on the relative efficiency of the public and private hospitals, as well as on the size of the public capacity. In fact, if the regulator can observe individual incomes and can use lump sum transfers, there is no scope for in-kind redistribution.

Note also that ex post, agents going to the public sector always have a lower utility than those who end up in the private: $u(\bar{c}) + h - w^* \leq u(\bar{c}) + h$ whenever $w^* > 0$. Hence, in our first best, agents visiting the public hospital are not compensated for the waiting cost. This is a direct consequence of the separability between consumptions and waiting times in the individual utility function.[5]

We summarize our results in the following proposition:

Proposition 1 *Assume that the social planner observes the agents' income, y, and that he can assign agents to the public or the private hospital. The first-best optimum is such that:*

i) Consumptions are equalized whether the agent goes to the private or the public hospital and across income levels: $c_y^{PR} = c_y^{PU}\ \ \forall y$.

ii) The probability to go to the public sector is the same across income levels,

[5] Relaxing this assumption would be at the cost of increased complexity without bringing more results.

$\theta^* = \theta \ \ \forall y$.

iii) Expected utilities are the same across agents.

iv) The ex post utility of an agent going to the public sector is lower than that of an agent going to the private, but it is independent of their incomes.

Let us finally compare the demands for the public hospital under the laissez-faire and the first best. In general, the demand for the public hospital is likely to be different between the first best and the laissez-faire for two reasons. First, it depends on the distribution of incomes between the laissez-faire and the first best, and in particular between $\tilde{y}^{LF} - k$ and c^{PU} and c^{PR}. The second reason is that, at the laissez-faire, agents take ρ as given so that they do not anticipate that by choosing the public hospital they increase the waiting cost suffered by all other agents getting the treatment at the public hospital. To make explicit this second reason, consider the case where the social planner does not make any income redistribution but only sets θ_y. His problem is

$$\max_{\theta_y} \int_{\underline{y}}^{\bar{y}} \theta_y \Phi \left(\theta_y u(y-k) + (1-\theta_y) u(y-p) - \theta_y w + h \right) f(y) dy.$$

The derivative with respect to θ_y is

$$\begin{aligned} &\Phi' \left(EU_y \right) \left[u(y-k) - u(y-p) - w \right] \\ &- \frac{1}{f(y)} \frac{\partial w}{\partial \theta_y} \int_{\underline{y}}^{\bar{y}} \theta_y \Phi' \left(EU_y \right) f(y) dy = 0. \end{aligned} \tag{10}$$

Again, one can show that $w > 0$, so that $\partial w / \partial \theta_y = f(y)/\hat{x}$. The second term of the derivative above is thus constant across y. The first term is negative if $u(y-k) - w < u(y-p)$, i.e. if y is greater than the laissez-faire cutoff income $\tilde{y}^{LF}$. In this case, the derivative with respect to theta is always negative, and $\theta_y = 0$. Whenever $u(y-k) - w \geq u(y-p)$, i.e. whenever y is smaller than $\tilde{y}^{LF}$, the first term is positive and decreasing in y.

Consequently, there exists a threshold $y^* \leq \tilde{y}^{LF}$ such that all individuals with income below y^* should visit the public hospital ($\theta_y = 1$), and all individuals with income higher than y^*should go to the private hospital ($\theta_y = 0$). Indeed, the marginal willingness to pay to avoid the waiting time is increasing in income so that agents with a lower income prefer to suffer the waiting time while high income agents prefer to avoid it. This threshold is implicitly defined by the following equality.

$$\Phi\left(u(y^{*}-k)-w+h\right)\left[u(y^{*}-k)-u(y^{*}-p)-w\right]$$
$$-\frac{1}{\hat{x}}\int_{\underline{y}}^{\bar{y}}\Phi'\left(u(y^{*}-k)-w+h\right)f(y)dy=0\quad\forall y \tag{11}$$

Comparing this expression with (2), one can check that $\tilde{y}^{LF}>y^{*}$ and waiting times are higher at the laissez faire than at the first best, since the demand for the public sector in higher in the former case.[6] Agents only compare the indirect utility in the public and in the private hospital, and do not internalize their own contribution to the congestion of the public hospital (second term in (11)).

Let us now discuss how the social planner could decentralize the first best optimum.

3.2 Decentralized first-best solution

If the social planner observes incomes and can directly assign agents to the public and the private hospitals (or even set a probability to end up in the public), the first-best optimum can easily be implemented through individualized lump-sum transfers from high-income to low-income agents.

However, assuming that the social planner has no direct control on the agents' hospital choice seems more natural and in that case, more instruments than just individualized taxes are needed to implement the first best solution. We show in this section that it is possible to decentralize it, by using both individualized lump-sum taxes denoted T_y and T'_y, and a uniform subsidy on the price of the private hospital, τ.[7] Without loss of generality, we assume that the price payed by patients in the public hospital is equal to zero. The timing is the following: the social planner first sets individual taxes and the level of the subsidy.[8] Given the lump sum taxes, agents choose whether to seek care in the

[6]To see this, evaluate (10) at $\tilde{y}^{LF}$. It is always negative.

[7]Note that instead of setting τ, the social planner could as well set the price p of the private hospital. Here, we assume that the social planner has no direct control on p, that is why we focus on a subsidy on the price of the private hospital.

[8]It is easy to show that a tax on the public hospital use and a subsidy on the use of the private hospital are equivalent. Since the total demand for health services is inelastic, subsidies and fees only affect the allocation of agents across hospitals (see Hoel and Sæther, 2003).

public or in the private hospital. If they choose to go to the private hospital they receive the subsidy.

In Appendix A we prove the following proposition:

Proposition 2 *The first best optimum can be decentralized through the following individualized tax and transfer scheme and a subsidy on the private health care price:*

1. Agents pay $T_y = y - \bar{c}$ with probability θ^*.

2. Agents pay $T_y^{'} = y - (\bar{c} + (1-\tau)p)$, with probability $1-\theta^*$.

3. Agents choosing the private hospital receive a subsidy on the private hospital price $\tau \in [\underline{\tau}, \bar{\tau}]$, where $\underline{\tau}$ and $\bar{\tau}$ are both smaller than one and are implicitly defined by

$$u(\bar{c}) - u\left(\bar{c} - (1-\bar{\tau})p\right) = \frac{\theta^*}{\hat{x}} - 1,$$

and

$$u\left(\bar{c} + (1-\underline{\tau})p\right) - u(\bar{c}) = \frac{\theta^*}{\hat{x}} - 1.$$

This ensures the optimal sorting of patients between hospitals: θ^* agents choose the public hospital while a $(1-\theta^*)$ agents choose to the private one.

Two aspects of this result should be highlighted. First, the optimal subsidy might be smaller than zero. Consequently, in some cases both a tax and a subsidy work in decentralizing the first best. These two options are indifferent for the final allocation, since taxes and transfers are adjusted such that, in the end, every agents choosing the private hospital have a net consumption $\bar{c}$ and those choosing the public one have $\bar{c}$. Receiving a high disposable income and having to pay a tax is equivalent to receiving a smaller disposable income and receive a subsidy on private services.

Secondly, we assume here that the transfers only depend on the agents' income, and not on their hospital choice. Once they receive the transfers, agents are free to choose their most preferred provider given the relative prices set by the social planner. To say it differently, the subsidy τ is set such that, in the end, we obtain the optimal allocation of agents between the public and the private hospitals.[9] If the lump sum transfers were also dependent on the choice of the hospital, it is clear that the subsidy would be a redundant instrument.

[9]We assumed here that the social planner cannot directly intervene in the private sector and fix the price p. If this was the case, the price paid by agents for the private sector would be higher or smaller than p depending on whether τ is positive or negative.

4 Asymmetric information

In the previous section, we showed that, under perfect information, the optimal allocation can be implemented even if the social planner cannot directly assign patients to the public or the private hospital. We now assume that the social planner cannot observe the agents' income, y (and thus, their consumption) and consider successively two cases.

In the first case, the social planner can assign patients to the public or the private hospital. She thus offers agents a menu of lotteries $(\theta_y, T_y^{PU}, T_y^{PR})$, specifying the probability to be assigned to the public hospital and the taxes to be payed when assigned to the public and to the private hospital. Each agent selects a lottery, and subsequently the social planner assigns him to one of the two hospitals according to the selected $\theta_{y'}$.

In the second case, the social planner cannot assign agents to any hospital but she can nonetheless observe the hospital choice made by each agent. Since the social planner cannot observe incomes, transfers can only depend on the hospital choice. The tax scheme consists of a tax T^{PU} if the agent chooses the public hospital, and a tax T^{PR} if the agents chooses the private one. As we will see, this tax scheme is equivalent to a tax scheme comprising a constant lump sum tax and a (positive or negative) subsidy on the private hospital price.

4.1 Unobservable incomes and possibility to assign patients to hospitals

Let us first study the case where the social planner cannot observe incomes, but can nonetheless assign agents to different hospitals. The second-best optimal scheme consists then in offering a menu of lotteries $(\theta_y, T_y^{PU}, T_y^{PR})$ which depends on the individual's declared income. The problem of the social planner is:

$$\max_{\theta_y, T_y^{PU}, T_y^{PR}} \int_{\underline{y}}^{\bar{y}} \Phi\left(\theta_y u(y - T_y^{PU}) + (1-\theta_y)u(y - T_y^{PR} - p) - \theta_y w\right) f(y)dy$$

$$\text{s.to} \int_{\underline{y}}^{\bar{y}} \left[\theta_y (T_y^{PU} - k) + (1-\theta_y) T_y^{PR}\right] f(y)dy \geq 0$$

$$\begin{aligned} \text{s.to} \quad & \theta_y u(y - T_y^{PU}) + (1-\theta_y)u(y - T_y^{PR} - p) - \theta_y w \geq \\ & \theta_{\tilde{y}} u(y - T_{\tilde{y}}^{PU}) + (1-\theta_{\tilde{y}})u(y - T_{\tilde{y}}^{PR} - p) - \theta_{\tilde{y}} w \quad \forall y, \tilde{y} \end{aligned}$$

The first constraint is the resource constraint. The last constraints represent the set of all incentive constraints. Denoting

$$EU_y = \theta_y u(y - T_y^{PU}) + (1-\theta_y)u(y - T_y^{PR} - p) - \theta_y w,$$

this set of incentive constraints can be replaced by a unique local incentive constraint of the form[10]

$$E\dot{U}_y = \theta_y u'(y - T_y^{PU}) + (1-\theta_y)u'(y - T_y^{PR} - p).$$

In the following, we assume that the second order local conditions hold, so that a first order approach is valid (see also footnote 11).

Using the local incentive constraint, the problem is to maximize the following Lagrangian:

$$\begin{aligned}
\pounds \quad = \quad & \int_{\underline{y}}^{\bar{y}} \{\Phi\left(EU_y\right) f(y) \\
& +\lambda \left[\theta_y (T_y^{PU} - k) + (1-\theta_y) T_y^{PR}\right] f(y) \\
& +\mu_y \left[\theta_y u'(y - T_y^{PU}) + (1-\theta_y)u'(y - T_y^{PR} - p)\right] \\
& +\dot{\mu}_y EU_y \\
& +\alpha_y \left[\theta_y u(y - T_y^{PU}) + (1-\theta_y)u(y - T_y^{PR} - p) - \theta_y w - EU_y\right]\} \, dy,
\end{aligned}$$

where λ is the Lagrange multiplier associated to the resource constraint, μ_y is the co-state variable and α_y is the shadow value of the constraint $EU_y = \theta_y u(y - T_y^{PU}) + (1-\theta_y)u(y - T_y^{PR} - p) - \theta_y w$.

[10] Replacing by the local incentive constraint is possible if the Spence Mirlees condition holds.

The first order conditions are

$$
\begin{aligned}
\frac{\partial \pounds}{\partial EU_y} &= \Phi'\left(EU_y\right) f\left(y\right) - \alpha_y + \dot{\mu}_y = 0 \\
\frac{\partial \pounds}{\partial T_y^{PU}} &= \lambda f\left(y\right) - \mu_y u''(y - T_y^{PU}) - \alpha_y u'(y - T_y^{PU}) = 0 \\
\frac{\partial \pounds}{\partial T_y^{PR}} &= \lambda f\left(y\right) - \mu_y u''(y - T_y^{PR} - p) - \alpha_y u'(y - T_y^{PR} - p) = 0 \\
\frac{\partial \pounds}{\partial \theta_y} &= \lambda\left(T_y^{PU} - k - T_y^{PR}\right) f\left(y\right) \\
&\quad + \mu_y\left[u'(y - T_y^{PU}) - u'(y - T_y^{PR} - p)\right] \\
&\quad + \alpha_y\left[u(y - T_y^{PU}) - u(y - T_y^{PR} - p) - w\right] \\
&\quad - \frac{\partial w}{\partial \theta_y}\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy = 0.
\end{aligned}
\tag{12}
$$

The first condition yields:

$$
\dot{\mu}_y = \alpha_y - \Phi'\left(EU_y\right) f(y). \tag{13}
$$

From the second and third conditions, we obtain that

$$
\begin{aligned}
y - T_y^{PU} &= y - T_y^{PR} - p = c_y \\
\Longleftrightarrow T_y^{PU} &= T_y^{PR} - p,
\end{aligned}
$$

so that, like in the first best, consumptions are independent of whether the agent goes to the public or the private hospital.[11]

Using the findings above, the condition on θ_y can then be simplified as follows:[12]

$$
\frac{\partial \pounds}{\partial \theta_y} = -\alpha_y w - \frac{\partial w}{\partial \theta_y}\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy + \lambda\left(p - k\right) f\left(y\right) = 0, \tag{14}
$$

[11] Since $T_y^{PU} = T_y^{PR} + p = T_y$, the second order local condition takes the form

$$
u(y - T_y) - u(y' - T_y) \geq u(y - T_{y'}) - u(y' - T_{y'}),
$$

implying T_y non decreasing in y. We assume that this condition is satisfied.

[12] As in the first best, it is always optimal to have a positive waiting time since the first order condition with respect to θ_y evaluated in $w = 0$ is equal to $\lambda\left(p - k\right) f\left(y\right) > 0$.

and using (13), this condition can be rewritten as

$$w\Phi'\left(EU_y\right) = \lambda\left[p-k\right] - \frac{1}{\hat{x}}\int_{\underline{y}}^{\bar{y}} \theta_y \Phi'\left(EU_y\right) f(y)dy$$
$$- \left[\frac{\dot{\mu}_y}{f(y)}w + \frac{1}{\hat{x}}\int_{\underline{y}}^{\bar{y}} \dot{\mu}_y \theta_y dy\right]. \quad (15)$$

From this condition, we prove in the appendix that the probability to be assigned to the public hospital in decreasing in y. Thus, contrary to the first best, agents now face a probability to go in the public sector which depends on their income. This probability is decreasing in income so as to relax incentive constraints and ensure that high-income agents do not wish to mimic low-income ones. Since high-income agents obtain a higher marginal utility from patronizing the private hospital,setting higher waiting times for low-income agents than for high-income ones permits to impose on high income agents a higher tax burden.[13] This enables some income redistribution.

Let us then compare this condition with the first best one. Condition (15) differs from the first best one, (7) by the last term in parenthesis. This is related to the introduction of the local incentive constraint, and can be decomposed into two parts. The first part, $\frac{\dot{\mu}_y}{f(y)}w$ is the effect of a marginal increase of θ_y on the incentive constraint of individual y. The second part, $\frac{1}{\hat{x}}\int_{\underline{y}}^{\bar{y}} \dot{\mu}_x \theta_x f(x)dx$ corresponds to the effect of increasing θ_y on the waiting times faced by all other individuals, and thus the effect it has on their incentive constraints. We prove in the appendix that $\dot{\mu}_y$ is negative for low incomes, $y_L \in [\underline{y}, y_k]$, and positive for high incomes, $y_H \in [y_k, \bar{y}]$ where $y_k \in [\underline{y}, \bar{y}]$ is such that

$$\frac{1}{w}\left(\lambda\left(p-k\right) - \frac{\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy}{\hat{x}}\right) = \Phi'\left(EU_{y_k}\right).$$

Hence, $\dot{\mu}_y < 0$ on the interval $[\underline{y}, y_k]$, so that the last term in (15) is positive for the agents with incomes inside this interval and their probability θ_{y_L} is pushed upwards with respect to the first best. On the contrary for agents with incomes $y_H \in [y_k, \bar{y}]$, this last term is negative so that θ_{y_H} is pushed downward. Such distortions on the probability of being assigned to the public hospital enables the social planner to relax the incentive constraints, and to redistribute

[13] To see this, compute the marginal utility of an agent with income y with respect to θ_y. It is decreasing in y, so that the marginal utility from patronizing the private sector is increasing in y.

some income from high- to low-income agents, by setting an increasing T_y function.

Let us finally say a word on the size of the waiting time. Replacing (12) into (14), one obtains

$$w^{SB} = \frac{u'(c_y)}{\lambda f(y) - \mu_y u''(c_y)} \left[\lambda (p-k) f(y) - \frac{f(y)}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \frac{\lambda f(y) - \mu_x u''(c_y)}{u'(c_y)} \theta_x dx \right],$$

and comparing it with its first best counterpart, (8), we find that the two expressions only differ by the term $\mu_y u''(c_y)$ which is related to the introduction of the incentive constraints. Since $\mu_y \leq 0$, we find that the waiting time is distorted up at the second best with respect to the first best. Distorting up the waiting times allows to avoid mimicking: if an agent mimics a lower income type, he will have to bear waiting times with a higher probability. If this time is high, the agent will have smaller incentives to do so.

Our results are summarized in the following proposition:

Proposition 3 *Assume that the social planner does not observe agents' income, y, but that he can assign agents to the public or the private hospital. The second-best optimum involves that:*

i) θ_y is decreasing in y.

ii) As compared to the first best, the waiting time is distorted up.

iii) Income taxation is such that consumptions are independent from whether the agent goes to the public or the private sector.

4.2 Lump sum taxation and subsidization of the private hospital

Let now assume that the social planner cannot use the probability θ_y as a policy instrument. This is a reasonable scenario, since in practice it is unlikely that the social has enough power to assign patients to hospital according to an income-dependent lottery. Such a system would require a high implementation cost. Furthermore, it would be incompatible with a certain view of fairness, and in particular with the principle of equal treatment of equals.[14] Hence, in this section, we assume that the social planner imposes a lump sum tax T to

[14]For a discussion, see Brito et al., 1995.

all individuals, but that agents going to the private hospital receive a subsidy τ on the fee paid in the private hospital. Like in Section 3.2, the idea behind proposing a subsidy on the private sector is to induce more people to use the private facility and thus, to reduce waiting times in the public hospital.[15]

Given the social planner's instruments, individuals can freely choose the hospital they patronize. An individual with income y chooses the public hospital if and only if $u(y-T) - w - u(y - T - (1-\tau)p)) \geq 0$. We define by $\tilde{y}(T,\tau)$ the income of the agent who is indifferent between patronizing the public or the private hospital when the government proposes the fiscal scheme (T,τ). This income is implicitly defined by:

$$u(\tilde{y}(T,\tau) - T) - w(\tilde{y}(T,\tau)) = u(\tilde{y}(T,\tau) - T - (1-\tau)p), \tag{16}$$

where

$$w(\tilde{y}(T,\tau)) = \frac{F(\tilde{y}(T,\tau)}{\hat{x}} - 1.$$

Since high-income agents have a higher willingness to pay to avoid the waiting time, they choose the private hospital. If $\tau < 0$, agents face a tax on the private hospital and $T^{PR} > T^{PU}$. In such case, high-income agents (who go to the private hospital) face higher overall taxation, even though incomes were unobservable in the first place. This would allow to redistribute income from high- to low-income individuals. However, a tax on the private sector would result in high waiting times. The social planner, when setting τ, has thus to trade-off redistribution concerns and waiting times containment.

This being said, the social planner's problem is now:[16]

$$\max_{T,\tau} \int_{\underline{y}}^{\tilde{y}(T,\tau)} \Phi\left(u(y-T) + h - w(\tilde{y}(T,\tau))\right) f(y)dy + \int_{\tilde{y}(T,\tau)}^{\bar{y}} \Phi\left(u(y - T - (1-\tau)p) + h\right) f(y)dy$$
$$s.to \quad T \geq F(\tilde{y}(T,\tau))k + \left[1 - F(\tilde{y}(T,\tau))\right]\tau p.$$

[15]Also Hoel and Sæther (2003) study the optimal subsidy on private treatments. However, in their model, there is no concern for income redistribution as agents all have the same income.

[16]This last section is very close to Besley and Coate (1991) who assume that the quality of public hospital is financed through a head tax.

The first-order conditions with respect to τ and T are, respectively

$$\frac{\partial \pounds}{\partial \tau} = -\frac{\partial \tilde{y}}{\partial \tau}\left[\frac{\partial w}{\partial \tilde{y}}\int_{\underline{\mathrm{y}}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) f(y)dy + \lambda(k-\tau p)f(\tilde{y})\right]$$
$$+\,p\int_{\tilde{y}}^{\bar{y}} \Phi'\left(u(y-T-(1-\tau)p)+h\right) u'(y-T-(1-\tau)p)f(y)dy - \lambda\left[1-F(\tilde{y})\right]p = 0 \tag{17}$$

$$\frac{\partial \pounds}{\partial T} = -\left[\begin{array}{l} \int_{\underline{\mathrm{y}}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) u'(y-T)f(y)dy \\ +\int_{\tilde{y}}^{\bar{y}} \Phi'\left(u(y-T-(1-\tau)p)+h\right) u'(y-T-(1-\tau)p)f(y)dy \end{array}\right]$$
$$-\frac{\partial \tilde{y}}{\partial T}\left[\frac{\partial w}{\partial \tilde{y}}\int_{\underline{\mathrm{y}}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) f(y)dy + \lambda(k-\tau p)f(\tilde{y})\right] + \lambda = 0, \tag{18}$$

where, for ease of notation, we set $\tilde{y} \equiv \tilde{y}(T,\tau)$ and where we replace for (16).

Our results are derived in Appendix B and summarized in the following proposition:

Proposition 4 *Assume that the social planner does not observe income, y, and that agents are free to choose their hospital. The optimal tax scheme consists in a linear tax on income, T and a linear subsidy τ, on the use of the private hospital, of the following form:*

$$\tau^* = \frac{-cov\left(\Phi' u', \mathbb{k}\right) + \frac{\partial \tilde{y}^c}{\partial \tau}\frac{1}{p}\left[\frac{1}{\hat{x}}\int_{\underline{y}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) f(y)dy + \lambda k\right] f(\tilde{y})}{\lambda f(\tilde{y})\partial \tilde{y}^c/\partial \tau},$$

where $cov\left(\Phi' u', \mathbb{k}\right)$ is the covariance between the social marginal utility of consumption and the use of the private hospital, and $\mathbb{k} = \{0,1\}$ with 0 (resp. 1) accounts for the demand for the public market (resp. private market).

The denominator is the standard efficiency term and is negative as it seems reasonable to assume that $\partial \tilde{y}^c/\partial \tau < 0$ where

$$\frac{\partial \tilde{y}^c}{\partial \tau} = \frac{\partial \tilde{y}}{\partial \tau} + \frac{\partial \tilde{y}}{\partial T}\frac{\partial T}{\partial \tau}. \tag{19}$$

Increasing the subsidy on the private sector distorts the choice between public and private hospitals: agents are more likely to choose the private hospital than in a situation where the subsidy is equal to zero. This has to be weighted by the number $f(\tilde{y})$ of agents at the threshold productivity and by the cost of public funds, λ. The first term in the numerator is the equity term and is positive since $cov\,(\Phi' u', \mathbb{k}) < 0$ (i.e. choosing the private hospital is negatively correlated with marginal utility of consumption as this corresponds to higher income levels). This first term is equivalent to Besley and Coate (1991): if the quality of the public sector is set exogenously by the social planner (translated to our model, this means that the public good cannot be congested), it is optimal to set a tax on the use of the private sector as this enables to increase the tax paid by high-income agents and thus to foster redistribution. The second term in the numerator accounts for the internalization of the effect of $\tilde{y}^c$ on the waiting cost, $F(\tilde{y})/\hat{x} - 1$ and on the government revenue. Since the bracket is positive and $\partial \tilde{y}^c / \partial \tau < 0$, this term is negative. Indeed, when τ becomes higher, more agents choose the private hospital so that first, it decreases the waiting cost in the public hospital and second it relaxes the budget constraint of the government since it reduces the total cost from running a public hospital. This second term in the numerator pushes toward subsidization of the private hospital so as to decrease the number of agents going to the public hospital. Depending on which effect dominates (redistributive concerns or the effect on $\tilde{y}^c$), τ^* could be either positive or negative. If the $\tilde{y}^c$-effect dominates the income redistribution effect, it is optimal to subsidize the private hospital. As an example, let us see what happens when $u\,(c)$ is linear and the social planner is utilitarian ($\Phi(.)$ is linear). In this case, there is no redistributive concern and $cov\,(\Phi' u', \mathbb{k}) = 0$ so that the subsidy on the private hospital simplifies to

$$\tau^* = \frac{1}{\lambda p} \left[\frac{1}{\hat{x}} \int_{\underline{\mathbf{y}}}^{\tilde{y}} \Phi' \left(y - T + h - w(\tilde{y}) \right) f(y) dy + \lambda k \right] > 0.$$

In this case, the subsidy simply corrects for the imperfect internalization by agents of the waiting costs and of the true cost of the public hospital. On the contrary, if redistributive concerns are very large, $-cov\,(\Phi' u', \mathbb{k}) > 0$ dominates the impact on $\tilde{y}^c$ and a tax on the private hospital, $\tau^* < 0$, would be optimal.

To sum up, we showed that, when taxation is linear, income redistribution is limited by the fact that the quality of the public hospital depends on the number of public patients, which is endogenous. This is different from Besley and Coate (1991), where the good that is publicly provided does not suffer congestion problems. In our case, taxation of the private hospital facility (that is $\tau^* < 0$), which is useful for redistribution purposes, is limited by the increase in the number of patients that would in turn be willing to use the public hospital and by the subsequent increase in waiting costs bared by low-income agents. Our result also differ from the one of Hoel and Sæther (2003): in their framework, the absence of redistribution motives always calls for a positive subsidy on the private treatment. In our model, we could as well have a subsidy or a tax.

5 Conclusion

This paper analyzes optimal income taxation when agents with different incomes all need a health treatment. We assume that they can choose between public and private provision. The public hospital is less expensive than the private one but exhibits waiting times depending on the number of patients treated. Differently from previous literature, we do not focus on the optimal size of the public health hospital, but on the role of endogenous waiting times as an instrument that can be used to ease income redistribution under asymmetric information.

We first find that, everything else being equal, at the laissez-faire the number of agents using the public facility is too high compared to the first-best optimum. This is the result of a simple externality problem: individuals do not internalize the effect of their hospital choice on the waiting time, which leads to higher than optimal waiting times at the laissez-faire.

We then turn to the asymmetric information problem. First, we analyze the second best allocation when the social planner can assign patients to hospitals. In this case, agents are proposed a menu of lotteries, describing the probability of being assigned to the public hospital and the income taxes they should pay if they go to the public or the private sector. With respect to the first best, low-income agents face a higher probability of being assigned to the public hospital than their high-income counterparts. Since high-income agents obtain higher marginal utility from the private sector, increasing the probability of being assigned to the public hospital (equivalently decreasing the probability of going to the private hospital) for low-income agents enables to relax incentive constraints and to impose a higher tax on high-income agents.

Second, we study a problem where the social planner cannot assign patients. In this case, only a linear income taxation together with a subsidy on the private hospital, can be used. We find that the optimal subsidy may be either positive or negative depending on the relative importance of redistribution and efficiency concerns. In that case, we have this somewhat interesting result: if public hospitals can be congested, their redistributive role can be undermined. Indeed, if waiting times are not too high, high-income agents should face a tax on the use of the private facility so as to redistribute resources toward low income-agents. On the opposite, if waiting times are high, it is optimal to encourage people to use the private facility. In such case, a subsidy on the private hospital fee turns out to be optimal and less income redistribution can be achieved.

Finally, our model relies on some important assumptions. The first one consists in assuming a fixed public capacity. As we already mentioned, we believe that it is a reasonable assumption in the short run. Moreover, our research question does not concern the optimal size of the public hospital, but the optimal taxation schemes when a public hospital already exists. This is certainly relevant for many countries in which the presence of a universal health system can be considered as a given, due to political constraints.

Another assumption concerns the modeling of individual utility functions. We use a utility function that is separable in the utility of consumption and in the benefit from the treatment, and thus on waiting times. Allowing for non separability would have been at the expense of increased complexity. Also, it is not clear neither theoretically nor empirically, whether the marginal utility of income should be increasing or decreasing with the health status.

Finally, in this paper, we assume that agents all suffer from the same illness and that the treatment can be obtained equally in one or the other system. In reality, this may not always be the case. Some public hospitals may be specialized in treating very peculiar illnesses that require specific and expensive technologies but that are not profitable for a private facility. The idea is that the public sector should intervene when fixed costs are so high that the private sector would never enter the market, even if there is a benefit for the community. On the other hand, private facilities may choose to specialize in highly profitable treatments, like plastic surgery. It would be interesting to see how our results would be modified when taking into account differences in hospital specialization. This is in our research agenda.

References

[1] Besley T. and S. Coate, 1991, Public provision of private goods and the redistribution of income, *American Economic Review*, 18(4), 979-984.

[2] Besley T., Hall J. and J. Preston, 1999, The demand for private health Insurance: do waiting lists matter?, *Journal of Public Economics*, 72(2), 155-181.

[3] Blomquist S. and V. Christiansen, 1995, Public provision of private goods as a redistributive device in an optimum income tax model, *Scandinavian Journal of Economics*, 97(4), 547-567.

[4] Boadway R. and M. Marchand, 1995, The use of public expenditure for redistributive policy, *Oxford Economic Papers*, 47, 45-59.

[5] Brito D.L., Hamilton J.H., Slutzky S.M. and J.E. Stiglitz, 1995, Randomization in optimal income tax schedules, *Journal of Public Economics*, 56, 189-223.

[6] Cremer H. and F. Gahvari, 1997, In-kind transfers, self-selection and optimal tax policy, *European Economic Review*, 41, 97–114.

[7] Hoel M. and E.M. Sæther, 2003, Public health care with waiting time: the role of supplementary private health care, *Journal of Health Economics*, 22, 599-616.

[8] Iversen T., 1997, The effect of a private sector on the waiting times in a national health service, *Journal of Health Economics*, 16, 381-396.

[9] Marchand M. and F. Schroyen, 2005, Can a Mixed Health Care System be Desirable on Equity Grounds?, *Scandinavian Journal of Economics*, 107(1), 1-23.

APPENDIX

A Proof of Proposition 2

Let us first define the disposable income as

$$D_y \equiv y - T_y$$

which is simply the income y net of taxes. The social planner proposes T_y, which can be positive or negative, to agents with income y and an additional Pigouvian subsidy, τ if they agree to go to the private hospital. In the following, we show how to choose the appropriate levels of (T_y, τ) in order to implement the first-best allocation $(\bar{c}, \theta^*)$. To do so, an agent with income y will be attributed D_y with probability θ^*, and $D_y^{'}$ with probability$(1-\theta^*)$. To be optimal, this allocation $\left(D_y, D_y^{'}, \tau\right)$ should satisfy three conditions. The first two conditions relate to the choice of optimal consumption levels while the last one relates to the optimal partition between the public and the private hospitals.

1. Consumptions should be equal to the first best level, namely $D_y = \bar{c}$ and $D_y^{'} - (1-\tau)p = \bar{c}$.

2. This allocation $\left(D_y, D_y^{'}, \tau\right)$ should also satisfy the government budget constraint, evaluated at the optimum,

$$E\left[y\right] - \theta^*(D_y + k) - (1-\theta^*)(D_y^{'} + \tau p) \geq 0.$$

3. Under this allocation $\left(D_y, D_y^{'}, \tau\right)$, agents with income y should actually choose the public hospital, that is

$$[u(D_y) - u(D_y - (1-\tau)p)] \geq \frac{\theta^*}{\hat{x}} - 1,$$

and agents with income y' should actually choose the private hospital, that is

$$\left[u(D_y^{'}) - u(D_y^{'} - (1-\tau)p)\right] \leq \frac{\theta^*}{\hat{x}} - 1.$$

Setting, as in condition 1,

$$\begin{aligned} D_y &= \bar{c} \\ D_y^{'} &= \bar{c} + (1-\tau)p \end{aligned}$$

trivially satisfies condition 2. This implies differentiated lump-sum taxes / transfers equal to

$$\begin{aligned} T_y &= y - \bar{c} \lesseqgtr 0 \\ T_y^{'} &= y' - (\bar{c} + (1-\tau)p) \lesseqgtr 0 \end{aligned}$$

for agents going respectively to the public hospital and to the private hospital.

Replacing then for the expressions of D_y, $D_y^{'}$ and θ^* in the inequalities of condition 3, we obtain

$$A \equiv u(\bar{c}) - u\left(\bar{c} - (1-\tau)p\right) \geq \frac{\theta^*}{\hat{x}} - 1 \tag{20}$$

and

$$B \equiv u\left(\bar{c} + (1-\tau)p\right) - u(\bar{c}) \leq \frac{\theta^*}{\hat{x}} - 1. \tag{21}$$

where for ease of notation in the following, we denote the LHS of these inequalities A and B respectively. Let us first mention that the above inequalities imply that necessarily $\tau \leq 1$ so that there can be at best a complete reimbursement of the private hospital fee. Since u is concave, $A \geq B$. Furthermore, both A and B increase when τ decreases. Thus, (20) is satisfied for any $\tau \leq \bar{\tau}$, where $\bar{\tau}$ is implicitly defined by

$$u(\bar{c}) - u\left(\bar{c} - (1-\bar{\tau})p\right) = \frac{\theta^*}{\hat{x}} - 1. \tag{22}$$

In a similar way, (21) is satisfied by any $\tau \geq \underline{\tau}$, where $\underline{\tau}$ is implicitly defined by

$$u\left(\bar{c} + (1-\underline{\tau})p\right) - u(\bar{c}) = \frac{\theta^*}{\hat{x}} - 1. \tag{23}$$

By concavity of u, $\bar{\tau} > \underline{\tau}$. Consequently, there exists a set of subsidies $[\underline{\tau}, \bar{\tau}]$ satisfying inequalities (20) and (21).

B Asymmetric information and possibility to assign agents to the hospitals

From condition (14), we find the value for α_y

$$\alpha_y = \frac{1}{w} f(y) \left[\lambda (p-k) - \frac{\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dx}{\hat{x}} \right]. \tag{24}$$

Thus, $\alpha_y = rf(y)$, where r is a constant. For (14) to hold, one needs $r \geq 0$, so that $\alpha_y \geq 0$.[17] Using (24) and (13), we can write

$$\dot{\mu}_y = \left[\frac{1}{w} \left(\lambda (p-k) - \frac{\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dx}{\hat{x}} \right) - \Phi' (EU_y) \right] f(y), \tag{25}$$

Moreover, the transversality conditions are $\mu_{\bar{y}} = \mu_{\underline{y}} = 0$ so that, together with the above equation, one obtains

$$\begin{aligned} \mu_y &= -\int_y^{\bar{y}} \dot{\mu}_x dx \\ &= \int_y^{\bar{y}} \left[\Phi' (EU_x) - \frac{1}{w} \left(\lambda (p-k) - \frac{\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy}{\hat{x}} \right) \right] f(x) dx \end{aligned}$$

Let us first study the sign of $\dot{\mu}_y$. The first term in parenthesis in (25) is a constant. Since EU_y is increasing in y and $\Phi''(.) < 0$, $\Phi'(EU_y)$ is decreasing in y and thus $\dot{\mu}_y / f(y)$ is necessarily increasing in y. Using the transversality conditions, this leads to having $\dot{\mu}_y$ negative up to y_k and positive for any $y > y_k$, where y_k is implicitly defined by

$$\frac{1}{w} \left(\lambda (p-k) - \frac{\int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy}{\hat{x}} \right) = \Phi' (EU_{y_k})$$

Given the shape of $\dot{\mu}_y$ described above, μ_y is decreasing up to y_k and then it is increasing. Since $\dot{\mu}_y$ is increasing, μ_y is convex. Overall, given the transversality conditions, one has that $\mu_y \leq 0$ for every $y \in [\underline{y}, \bar{y}]$.

[17] Suppose this was not true. Then, $-\alpha_y w - \frac{\partial w}{\partial \theta_y} \int_{\underline{y}}^{\bar{y}} \alpha_x \theta_x dx \geq 0$, and the condition would never be satisfied, since $p > k$.

The condition on θ_y can be rewritten as follows:

$$\frac{\partial \pounds}{\partial \theta_y} = -\alpha_y w - \frac{\partial w}{\partial \theta_y} \int_{\underline{y}}^{\bar{y}} \alpha_y \theta_y dy + \lambda (p - k) f(y) = 0 \tag{26}$$

Substituting for the first order condition on EU_y (13), it can be rewritten as

$$\begin{aligned} \frac{\partial \pounds}{\partial \theta_y} =& - w\Phi'(EU_y) + \lambda [p - k] - \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \theta_y \Phi'(EU_y) f(y) dy \\ & - \left[\frac{\dot{\mu}_y}{f(y)} w + \frac{1}{\hat{x}} \int_{\underline{y}}^{\bar{y}} \dot{\mu}_y \theta_y dy \right] \end{aligned} \tag{27}$$

Applying the implicit function theorem on the above equation, one has that

$$\frac{d\theta_y}{dy} =^s \frac{\partial^2 \pounds}{\partial \theta_y \partial y} = -\frac{d(\dot{\mu}_y / f(y))}{dy}$$

and we showed above that $\frac{d(\dot{\mu}_y/f(y))}{dy} > 0$. Hence, we have that $d\theta_y/dy < 0$.

C Proof of Proposition 4

Using the preceding first-order conditions and recalling that these are equal to zero at the equilibrium, we obtain

$$\frac{\partial \tilde{\pounds}}{\partial \tau} = \frac{\partial \pounds}{\partial \tau} + \frac{\partial \pounds}{\partial T} \frac{\partial T}{\partial \tau} \tag{28}$$

where $\tilde{\pounds}$ denotes the compensated Lagrangian and where $\partial T/\partial \tau = [1 - F(\tilde{y}(T, \tau))] p$ is obtained from the budget constraint of the government. This approach enables us to obtain tax formulas in terms of compensated tax derivatives. In the following we use the "compensated" derivative of $\tilde{y}$ with respect to τ, $\partial \tilde{y}^c/\partial \tau$

defined by (19). Replacing for (17), (18), and (19) , (28) can be rewritten as

$$\begin{aligned}
\frac{\partial \tilde{\mathcal{L}}}{\partial \tau} &= -\frac{\partial \tilde{y}}{\partial \tau}\left[\frac{\partial w}{\partial \tilde{y}}\int_{\underline{y}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) f(y)dy - \lambda(k-\tau p)f(\tilde{y})\right] \\
&\quad + p\int_{\tilde{y}}^{\bar{y}} \Phi'\left(u(y-T-(1-\tau)p)+h\right) u'(y-T-(1-\tau)p)f(y)dy - \lambda\left[1-F(\tilde{y})\right]p = 0 \\
&\quad + \frac{\partial T}{\partial \tau}\left[\begin{array}{l} -\int_{\underline{y}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) u'(y-T)f(y)dy \\ -\int_{\tilde{y}}^{\bar{y}} \Phi'\left(u(y-T-(1-\tau)p)+h\right) u'(y-T-(1-\tau)p)f(y)dy \\ -\frac{\partial \tilde{y}}{\partial T}\left[\frac{\partial w}{\partial \tilde{y}}\int_{\underline{y}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) f(y)dy - \lambda(k-\tau p)f(\tilde{y})\right] + \lambda \end{array}\right] \\
&= -\frac{\partial \tilde{y}^c}{\partial \tau}\left[\frac{\partial w}{\partial \tilde{y}}\int_{\underline{y}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) f(y)dy + \lambda(k-\tau p)f(\tilde{y})\right] \\
&\quad + p\left[\begin{array}{l} \int_{\tilde{y}}^{\bar{y}} \Phi'\left(u(y-T-(1-\tau)p)+h\right) u'(y-T-(1-\tau)p)f(y)dy \\ -\left[1-F(\tilde{y})\right]\int_{\underline{y}}^{\tilde{y}} \Phi'\left(u(y-T)+h-w(\tilde{y})\right) u'(y-T)f(y)dy \\ -\left[1-F(\tilde{y})\right]\int_{\tilde{y}}^{\bar{y}} \Phi'\left(u(y-T-(1-\tau)p)+h\right) u'(y-T-(1-\tau)p)f(y)dy \end{array}\right]
\end{aligned}$$

The last term in brackets is the covariance between marginal utility of consumption and the choice of the private hospital and is defined by

$$\begin{aligned}
cov\left(u'(c), \mathbb{k}\right) &= \int_{\tilde{y}}^{\bar{y}} 1\times\Phi'(U^{PR})u'(y-T-(1-\tau)p)f(y)dy + \int_{\underline{y}}^{\tilde{y}} 0\times\Phi'(U^{PU})u'(y-T)f(y)dy \\
&\quad -\left[0\times F(\tilde{y}) + 1\times(1-F(\tilde{y}))\right] \\
&\quad \times\left(\int_{\underline{y}}^{\tilde{y}} \Phi'(U^{PU})u'(y-T)f(y)dy + \int_{\tilde{y}}^{\bar{y}} \Phi'(U^{PR})u'(y-T-(1-\tau)p)f(y)dy\right)
\end{aligned}$$

where

$$\begin{aligned}
U^{PR} &= u(y-T-(1-\tau)p) \\
U^{PU} &= u(y-T)-w(\tilde{y})
\end{aligned}$$

and $\mathbb{k} = \{0,1\}$ with 0 (resp. 1) accounting for the demand for the public market (resp. private market). At the optimum, rearranging $\partial\tilde{\mathcal{L}}/\partial\tau = 0$ yields Proposition 3.