THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2008/29

DRG prospective payment system: refine or not refine?

By

Elin J.G. Hafsteinsdottir and Luigi Siciliani

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

DRG prospective payment system: refine or not refine?

Elín J. G. Hafsteinsdóttir* Luigi Siciliani†

9 August 2008

Abstract

We present a model of contracting between a purchaser of health services and a provider (a hospital). We assume that hospitals provide two alternative treatments for a given diagnosis: a less intensive one (for example a medical treatment) and a more intensive one (surgical treatment). We assume that prices are set equal to the average cost reported by the providers, as observed in many OECD countries (yardstick competition). The purchaser has two options: 1) to set one tariff based on the diagnosis only; 2) to differentiate the tariff between the surgical and the medical treatment (i.e. to refine the tariff). We show that when tariffs are refined, the provider has always an incentive to overprovide the surgical treatment. If the tariff is not refined, the hospital underprovides the surgical treatment (and overprovides the medical treatment) if the degree of altruism is sufficiently low compared the opportunity cost of public funds. Our main result is that price refinement might not be optimal.

Keywords: Hospitals; Refinement; Diagnosis-Related-Groups.

JEL Classification: H42; I11; I18; L13

---

*Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, UK; E-mail: elinhaf@landspitali.is.

†Department of Economics and Related Studies, and Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK; and C.E.P.R., 90-98 Goswell Street, London EC1V 7DB, UK. E-mail: ls24@york.ac.uk.

1 Introduction

In the last twenty five years many OECD countries have adopted some form of DRG (Diagnosis Related Group) prospective payment system to reimburse hospitals. The payment system was first introduced in 1983 by the US Medicare Programme to remunerate hospitals for treating the elderly. The DRG system is believed to be a powerful tool to induce hospitals to reduce costs, and under certain conditions to encourage increase in activity and quality. Hospitals are paid a fixed price for each patient treated. At admission every patient is assigned to one of about 500 DRG based on their registered primary diagnosis. Each DRG bears a weight reflecting the average cost of patients in the given DRG relative to that of the average patient reimbursed by the system (MedPAC, 2007). The price for each patient treated is obtained by multiplying the relevant DRG weight by a fixed monetary value.

Since Medicare adopted the DRG payment system several refinements have been introduced. Refinements involve splitting a single DRG category into two or more DRG categories relating to the same primary diagnosis. The split has mainly been based on the type of treatment the patient receives. Patients diagnosed with the same primary diagnosis are grouped into different DRG categories. For example, patients receiving treatment involving a surgical procedure are grouped into a different category than patients receiving a medical treatment for the same disease (McClellan, 1997; Gilman, 1999; 2000).

This study investigates the conditions under which refinement of DRG weights and therefore prices is optimal. If prices could be set by the purchaser of health services at any arbitrary level, then it would seem intuitive that refinement is always welfare enhancing. By splitting one DRG into two DRGs (i.e. moving from one price to two prices) the purchaser effectively gains one additional instrument (one degree of freedom), which should then increase welfare. However, in practice, DRG weights are based on a more restrictive average-cost rule (yardstick competition), and cannot be set by the purchaser at an arbitrary level. We show that if prices are set equal to the average cost reported by the providers (as for example in Shleifer, 1985) then refining DRGs (i.e. introducing an additional price) is not necessarily welfare improving.

More precisely, the results of the model suggest that if prices are refined, then the provider has an incentive to overprovide the more intensive treatment. In contrast, when prices are not refined, providers may overprovide or underprovide the more intensive treatment. If altruism is low compared to the opportunity cost of public funds, and prices are not refined, then the provider underprovides the more intensive treatment.

This study contributes to the literature on provider incentives in health care (see Ellis and McGuire, 1986, 1990; Pope, 1989; Levaggi, 1996; Ellis, 1998; Ma, 1994; Chalkley and Malcomson, 1998a, 1998b, 2000; Rickman and McGuire, 1999; DeFraja, 2000; Barros, 2003; Beitia, 2003; Eggleston, 2005; Jack, 2005; Mougeot and Naegelen, 2005; Brekke and Sørgard, 2007). Our main departing assumption from the existing literature is that we allow more than one type of treatment for a given diagnosis, and that these treatments can be reimbursed at different prices.

There appears to be only a small theoretical literature which investigates the optimal payment system when more than one type of treatment is available for the same diagnosis. Malcomson (2005) investigates the optimal payment system, when there is asymmetric information on patient type and the cost of treatment. He uses a framework with two possible treatments for the same diagnosis, where the payment made by the purchaser can only be conditioned on the type of treatment provided. In the model the provider of health care decides which treatment to provide and to whom. Thus, dumping of some patients is allowed. In general it is optimal to pay a different price for different treatments. However, the optimal provision of treatments differs from what would be efficient if the purchaser had full information.

Boadway, Marchand, and Sato (2004) analyse the incentive for the providers to overuse expensive treatments. They suggest that the incentives of hospitals and doctors to assign patients to the high-tech treatment can be controlled in part through the reimbursement system. Within a hierarchical model involving three decision makers, the government, the hospital manager, and the doctor, they derive the optimal hospital reimbursement when patients, who differ in severity of illness, can be treated by either an expensive high-tech therapy or a cheaper low-tech therapy.

Siciliani (2006) is concerned with the optimal payment scheme when there are two treatments available for the same diagnosis, a more intensive treatment and a less intensive treatment, and the average severity of illness differs across hospitals. If the average severity of illness is private information only known to the provider, the provider has an incentive to overprovide the surgical treatment to patients with low severity of illness. The optimal payment scheme, when information is asymmetric, involves paying hospitals which provide a higher share of surgical treatments a higher price for the surgical treatment, and a lower price for the medical treatment.

The model presented in this study is different from the above models in several respects. First, and most importantly, the price is determined according to the average cost within each DRG, as we observe in many health care systems (and in line with the seminal paper of Shleifer 1985). This is in contrast with Boadway, Marchand, and Sato (2004), Malcomson (2005), and Siciliani (2006) where the purchaser can set prices at any level and also make use of lump-sum transfers. Second, it differs from Siciliani (2006) and Malcomson (2005) in that the provider is semi altruistic and thus cares for the benefit of the patients.

The rest of the study is organised as follows. Section 2 describes the main assumptions of the model. Section 3 solves the maximisation problem of the provider. Section 4 derives the optimal solution from the purchaser's perspective and the welfare implications of refining prices. Section 5 discusses some extensions of the model. Section 6 concludes.

2 The model

Patients with a certain diagnosis can receive one of two possible treatments of different intensity from the provider of health services (a hospital). Denote $\underline{\theta}$ as the less intensive treatment, and $\overline{\theta}$ as the more intensive treatment (for example an invasive treatment or a surgical procedure). Patients are assumed to differ in the severity of illness $s \in [\underline{s}, \overline{s}]$.

Severity has density function $f(s)$ and cumulative distribution function $F(s)$. The number of patients treated by each hospital is normalised to one.

The benefit from treatment to the patient depends on which treatment the patient receives and the severity of the patient. The benefit from the more intensive treatment for a patient with severity s is $b(\overline{\theta}, s)$ while from the less intensive treatment is $b(\underline{\theta}, s)$. Patient's severity is observed by the provider, which decides what type of treatment to provide. Patients have a passive role and accept to undertake any treatment decided by the provider.

As a clinical example, consider a patient diagnosed with ischemic heart disease caused by the narrowing of coronary arteries in the heart, which results in diminished blood flow to the heart. The diminished blood flow can cause a lack of oxygen to the heart muscle, which can cause severe chest pain called angina. Patients suffering from mild angina often fare well on medication such as nitrates, beta-blockers or calcium-blockers. However, sometimes medication is not enough to alleviate or control the symptoms of the disease. In that case there are more intensive treatments available, like an operation, called coronary artery bypass graft (CABG), or an invasive nonsurgical treatment where the coronary arteries are widened from inside the blood vessels by using a tube with an inflatable balloon attached to it (this procedure is known as percutaneous transluminal coronary angioplasty or PTCA). The more severe the symptoms become, such as a persistent angina, the less benefit for the patient from medication such as nitrates or beta-blockers. The medication is no longer sufficient to alleviate the symptoms.

We assume that: 1) $b(\overline{\theta}, s) \geq b(\underline{\theta}, s)$: patient's benefit is higher when the more intensive treatment is provided;[1] 2) $b_s(\overline{\theta}, s) \geq b_s(\underline{\theta}, s)$: patients with higher severity benefit more from the more intensive treatment than patients with lower severity (see Figure 1).

The number of hospitals providing treatment to patients is H. Each hospital $i = 1, ..., H$, maximizes utility denoted with U_i. Hospitals are assumed to be partially altruistic in the sense that they care at least to some extent about the benefit of the treatment they

[1] In section 5 we discuss the case where patients with low severity might have lower benefit from the more intensive treatment.

provide to patients. We therefore assume that the hospital maximizes $U_i = \alpha B_i + T_i - C_i$, where $\alpha \in [0,1]$ is the degree of altruism, i.e. the weight that the hospital attaches to patients' benefit from treatment; B_i is the total benefit obtained from treating patients at hospital i (defined more precisely below); T_i is the transfer received by hospital i; and C_i is the total cost incurred by hospital i from the provision of treatment.

The cost of treating each patient is a function of the type of treatment and patient's severity of illness. The cost of the less intensive treatment is $c\left(\underline{\theta}, s\right)$ and the cost of the more intensive treatment is $c\left(\overline{\theta}, s\right)$. In order to keep the exposition of the model simple, the less intensive treatment $\underline{\theta}$ will hereafter be referred to as the medical treatment, and the more intensive treatment $\overline{\theta}$ as the surgical treatment. We assume that the cost of treatment increases with the severity of illness of the patient: $c_s(\theta, s) > 0$. Moreover, we assume: 1) $c_s(\underline{\theta}, s) \geq c_s(\overline{\theta}, s)$: the cost increases more quickly with severity for the medical treatment than with the surgical treatment; 2) $c(\overline{\theta}, \underline{s}) \geq c(\underline{\theta}, \underline{s})$: the cost of the surgical treatment is higher than the cost of the medical treatment for patients with lowest severity; 3) $c(\underline{\theta}, \overline{s}) \geq c(\overline{\theta}, \overline{s})$: the cost of the medical treatment is higher for patients with highest severity. Finally, we assume that the benefit is always higher than the cost for any level of severity. Figure 1 provides a graphical illustration of a benefit and a cost function which satisfies our assumptions.

Consider again the example of patients with ischemic heart disease. The cost of providing medical treatment to a patient with angina has a low fixed cost. However, if the disease progresses and the symptoms become more severe the patient is in need of increasing medical treatment such as intravenous medication requiring intensive monitoring with increased specialised staff and *rapidly* increasing costs. When patients with angina are provided with either one of the more intensive treatment, CABG or PTCA, the fixed cost of providing the treatment is higher than the fixed cost of providing the medical treatment due to fixed costs related to the CABG surgery or resources needed for the invasive PTCA treatment. All patients undergoing CABG or PTCA require certain monitoring and care following the procedures regardless of the severity of illness. However, more severely ill patients may require more intensive care or prolonged monitoring. In the cost structure applied in this section it is assumed that the monitoring and care following the more intensive treatments is increasing at a slower rate with severity than for the medical treatment. This assumption may not be realistic for all types of treatments. In section 5 we discuss the case where the cost of surgical treatment is always higher than the cost of the medical treatment for any level of severity: $c(\overline{\theta}, s) \geq c(\underline{\theta}, s)$ for any s (as we will show the results are not affected qualitatively by this assumption).

We define z_i as the cut-off severity point above which the more intensive treatment is provided. The number of medical treatments is $\underline{n}_i = \int_{\underline{s}}^{z_i} f(s)ds = F(z_i)$, and the number of surgical treatments is $\overline{n}_i = \int_{z_i}^{\overline{s}} f(s)ds = 1 - F(z_i)$. The total number of treatments provided at hospital i is one ($n_i = \underline{n}_i + \overline{n}_i = 1$).

The total cost of treatment at each hospital i, when the cut-off point is z_i, is $C_i\left(z_i\right)$. It consists of the sum of the cost of providing the medical treatment, $C_i^L(z_i)$, and the cost of providing the surgical treatment, $C_i^H(z_i)$:

$$C_i(z_i) = C_i^L(z_i) + C_i^H(z_i) = \int_{\underline{s}}^{z_i} c\left(\underline{\theta}, s\right) f(s)ds + \int_{z_i}^{\overline{s}} c\left(\overline{\theta}, s\right) f(s)ds \tag{1}$$

Similarly, the total benefit from treating $\underline{n}_i$ number of patients with the medical treatment, and $\overline{n}_i$ number of patients with the surgical treatment at hospital i is

$$B_i(z_i) = \int_{\underline{s}}^{z_i} b\left(\underline{\theta}, s\right) f(s)ds + \int_{z_i}^{\overline{s}} b\left(\overline{\theta}, s\right) f(s)ds \tag{2}$$

We assume that patients' severity of illness is observed by the provider but not by the purchaser. The treatment instead is known to both the provider and the purchaser. We assume that hospitals must treat all patients with severity $s \in [\underline{s}, \overline{s}]$, i.e. hospitals cannot dump patients.

When the purchaser contracts with hospital $i = 1, ..., H$, prices are set equal to the average cost of providing the treatment at the H hospitals. This price setting mechanism is in line with how the price for each DRG is determined in the health care systems of many OECD countries (see Shleifer, 1985 for a similar specification).

If prices are *unrefined* the purchaser pays the same price p for the medical and the surgical treatment. The price is equal to the average cost of providing the treatments at the H number of hospitals:

$$p(z_i) = \frac{\sum_i^H C_i(z_i)}{\sum_i^H n_i}. \tag{3}$$

If the purchaser instead decides to *refine* prices, the purchaser sets two different prices. The price for the medical and surgical treatments are:

$$\underline{p}(z_i) = \frac{\sum_i^H C_i^L(z_i)}{\sum_i^H \underline{n}_i(z_i)}; \quad \overline{p}(z_i) = \frac{\sum_i^H C_i^H(z_i)}{\sum_i^H \overline{n}_i(z_i)}. \tag{4}$$

3 Optimal cut-off severity point

Hospital $i = 1, ..., H$, maximizes utility, $U_i = \alpha B_i + T_i - C_i$, with respect to the optimal cut-off point z_i. The hospital provides the medical treatment to patients with severity below the optimal cut-off point, and the surgical treatment to patients with severity above the optimal cut-off point.

3.1 Unrefined price

The transfer to hospital i when prices are unrefined is the price multiplied by the number of patients treated

$$T_i(z_i) = pn_i = \frac{\sum_i^H C_i(z_i)}{\sum_i^H n_i} n_i \tag{5}$$

Hospital i maximization problem is

$$\underset{z_i}{Max}\ U_i(z_i, z_{-i}) = \alpha B_i(z_i) + p(z_i)n_i - C_i(z_i) \tag{6}$$

The first order condition (FOC) is $\partial U_i / \partial z_i(z_i, z_{-i}) = \alpha B_{z_i} + \left(n_i / \sum_i^H n_i - 1\right) C_{z_i} = 0$. In the symmetric equilibrium $z_i = z_j = z^u$, where z^u denotes the optimal cut-off point for the hospital under unrefined prices. Also $n_i = n_j = \mathbf{1}$, so that $n_i / \sum_i^H n_i = 1/H$. The FOC can be written as

$$\alpha B_z(z^u) - (\frac{H-1}{H})C_z(z^u) = 0 \tag{7}$$

where $B_z(z^u) < 0$. More explicitly, the optimal cut-off point is determined such that

$$\frac{H-1}{H}\left[c\left(\overline{\theta}, z^u\right) - c\left(\underline{\theta}, z^u\right)\right] f(z^u) = \alpha\left[b\left(\overline{\theta}, z^u\right) - b\left(\underline{\theta}, z^u\right)\right] f(z^u). \tag{8}$$

z^u is set at the level where the marginal benefit of an additional medical treatment $\underline{\theta}$ is equal to the marginal cost. When the hospital provides one more medical treatment, it substitutes a medical treatment for a surgical one. Since by assumption the surgical treatment is always more beneficial, an additional medical treatment reduces total patients benefit ($\alpha \partial B_i / \partial z_i = -\alpha\left[b\left(\overline{\theta}, z^u\right) - b\left(\underline{\theta}, z^u\right)\right] f(z^u) < 0$). On the other hand, one more medical treatment reduces costs as the surgical treatment is more expensive than the medical treatment for low-severity patients ($c\left(\overline{\theta}, z^u\right) > c\left(\underline{\theta}, z^u\right)$). At the optimum, the marginal savings from an additional medical treatment are equal to the hospital's value of the marginal reduction in patient's benefit.

The second order condition is

$$\frac{\partial^2 U_i(z_i, z_{-i})}{\partial z_i^2} \equiv \left(1 - \frac{n_i}{\sum_i^H n_i}\right)\left[c_s\left(\overline{\theta}, z_i\right) - c_s\left(\underline{\theta}, z_i\right)\right] f(z_i) - \alpha\left[b_s\left(\overline{\theta}, z_i\right) - b_s\left(\underline{\theta}, z_i\right)\right] f(z_i) < 0 \tag{9}$$

which is satisfied since by assumption cost increases with severity more quickly under the medical treatment than under the surgical treatment ($c_s\left(\underline{\theta}, z^u\right) > c_s\left(\overline{\theta}, z^u\right)$) and benefit increases more quickly with patients' severity under the surgical treatment than under the medical treatment ($b_s\left(\overline{\theta}, z_i\right) > b_s\left(\underline{\theta}, z_i\right)$).

Note that, for any positive α, by choosing the cut-off point the provider has also an influence on the price. However, if the number of hospitals is sufficiently large ($H \to \infty$) then the hospital's own effect on price becomes negligible as $lim_{H\to\infty} \frac{H-1}{H} = 1$.

As a special case, note that if the provider is profit maximiser, i.e. altruism is zero ($\alpha = 0$), then the cut-off point is the one which minimises total costs, so that $C_{z_i} = 0$ or equivalently $c\left(\overline{\theta}, z^u\right) = c\left(\underline{\theta}, z^u\right)$: the cut-off point $z^u(\alpha = 0)$ is chosen such that the cost of the medical and surgical treatment are equated (see point E in Figure 1).

Suppose now that altruism is at the highest level, i.e. $\alpha = 1$, and the number of hospitals H is large so that $\frac{H-1}{H} \approx 1$. Then in Figure 1, the optimal cut-off point $z^u(\alpha = 1)$ is such that the difference in the benefit from the two treatments (AB) is equal to the difference in cost (CD).

From the first order condition, we also obtain

$$\partial z^u/\partial\alpha = -U_{z\alpha}/U_{zz} = -\left[b\left(\overline{\theta}, z^u\right) - b\left(\underline{\theta}, z^u\right)\right] f(z^u)/(-U_{zz}) < 0.$$

More altruism implies a lower cut-off point, and a higher (lower) number of surgical (medical) treatments. In Figure 1 we have $z^u(\alpha = 1) < z^u(\alpha > 0) < z^u(\alpha = 0)$.

In summary, with unrefined prices the purchaser pays the same price for both treatments and the optimal cut-off point for the semi-altruistic hospital is greater than zero, resulting in the hospital providing the medical treatment to patients with severity below the optimal cut-off point and the surgical treatment to patients with severity above the optimal cut-off point. Finally, note that since price is set equal to the average costs, profits are zero for healthcare providers.

3.2 Refined prices

If prices are refined the purchaser pays two different prices for the two treatments, price $\underline{p}$ for the medical treatment and price $\overline{p}$ for the surgical treatment. The overall transfer to hospital i is $T_i = \underline{p}\underline{n}_i + \overline{p}\overline{n}_i$. Hospital i maximizes utility with respect to the optimal cut-off point z_i:

$$\underset{z_i}{Max}\ \ U_i(z_i, z_{-i}) = \alpha B_i(z_i) + \underline{p}\left(z_i\right)\underline{n}_i\left(z_i\right) + \overline{p}\left(z_i\right)\overline{n}_i\left(z_i\right) - C_i^H(z_i) - C_i^L(z_i) \tag{10}$$

Differentiating with respect to z_i, we obtain:

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = \alpha B_{z_i} + \left(\underline{p} \frac{\partial \underline{n}}{\partial z_i} + \overline{p} \frac{\partial \overline{n}}{\partial z_i} \right) + \overline{n} \frac{\partial \overline{p}}{\partial z_i} + \underline{n} \frac{\partial \underline{p}}{\partial z_i} - C^H_{z_i} - C^L_{z_i} \tag{11}$$

Since we assume identical hospitals, the optimal cut-off point for each hospital is the same in equilibrium and equal to $z_i = z_{-i} = z^r$. The above simplifies to:

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = \alpha B_z(z^r) - f(z^r)(\frac{H-1}{H}) \left\{ \left[\overline{p} - c\left(\overline{\theta}, z^r\right) \right] - \left[\underline{p} - c\left(\underline{\theta}, z^r\right) \right] \right\} < 0 \tag{12}$$

See proof in the Appendix. Note that $\alpha B_{z_i} < 0$ because by assumption the surgical treatment is always more beneficial. Also, $\overline{p} = \dfrac{C^H_i(z^r)}{\overline{n}_i(z^r)} = \frac{\int_{z^r}^{\overline{s}} c\left(\overline{\theta}, s\right) f(s) ds}{1-F(z^r)} > c\left(\overline{\theta}, z^r\right)$: since the cost of treatment increases with the severity of illness, the *average* cost of the surgical treatment over the interval $[z^r, \overline{s}]$ is always higher than the *marginal* cost of the surgical treatment at the lower end of the severity interval, z^r. Similarly, $\underline{p} = \dfrac{C^L_i(z^r)}{\underline{n}_i(z^r)} = \frac{\int_{\underline{s}}^{z^r} c(\underline{\theta}, s) f(s) ds}{F(z^r)} < c\left(\underline{\theta}, z^r\right)$: since the cost of treatment increases with the severity of illness, the *average* cost of the medical treatment over the interval $[\underline{s}, z^r]$ is always lower than the *marginal* cost at the upper end of the severity interval, z^r. Therefore, the marginal utility from an increase in the cut-off point is negative and $z^r = \underline{s}$: only surgical treatments are provided. The result holds for any positive levels of altruism, as well as when the provider is profit maximiser (when $\alpha = 0$).

Intuitively, since the difference in the two tariffs is always bigger than the difference in the marginal cost (as $\overline{p} - \underline{p} > c\left(\overline{\theta}, z^r\right) - c\left(\underline{\theta}, z^r\right)$), the provider has never an incentive to increase the cut-off point above the minimum severity $\underline{s}$. In summary, when DRG prices are refined, there are strong incentives to overprovide the surgical treatment. As in the previous case, since prices are set equal to the average costs, profits are zero.

4 The purchaser

Since hospitals are identical, we can focus on the welfare generated by a representative hospital, denoted by W. Total welfare is simply $H * W$.

The purchaser's welfare consists of the total benefit of treatment to patients $B_i(z_i)$ net of the social cost of transfer to the hospitals, $(1+\lambda)T_i(z_i)$. Since the purchaser raises revenues by imposing distortionary taxes, it generates a deadweight loss from taxation captured by the parameter λ. Welfare is then:

$$W = B_i(z_i) - (1+\lambda)T_i(z_i) \tag{13}$$

The purchaser maximizes welfare subject to the hospital participation constraint $U_i \geqslant 0$. Furthermore, we also assume that hospitals are subject to a limited-liability constraint, so that the hospital cannot operate with losses. We therefore assume that the total transfer to the hospital cannot be less than the cost of providing treatment: $T_i(z_i) - C_i(z_i) \geqslant 0$. With social welfare decreasing in transfer, $\partial W_i / \partial T_i < 0$, the purchaser meets this constraint with strict equality $T_i = C_i$. Note that this implies that the utility of hospital is non-negative, as $U_i = \alpha B_i \geq 0$ and the hospital participation constraint is therefore satisfied. Substituting into the above we obtain that welfare is: $W = B_i(z_i) - (1+\lambda)C_i(z_i)$.

An alternative to the above specification is that the purchaser maximises a utilitarian welfare function, so that welfare is instead $W = B_i - (1+\lambda)T_i + U$, the sum of patients benefit and provider's utility net of the transfers to the provider. Given $U = \alpha B_i + T_i - C_i$, we obtain: $W = B_i - (1+\lambda)T_i + \alpha B_i + (T_i - C_i)$. From the limited-liability constraint, as before we obtain $T_i = C_i$, which implies $W = (1+\alpha)B_i - (1+\lambda)C_i$. It has been argued (see Chalkley and Malcomonson, 1999; Hammond, 1980) that this specification with altruistic providers leads to double-counting of the benefits of the patients, and that the altruistic component αB_i should be removed from the welfare. If we follow this suggestion we obtain that the welfare is again: $W = B_i(z_i) - (1+\lambda)C_i(z_i)$.

The purchaser maximizes welfare with respect to z_i

$$\underset{z_i}{Max} \quad W = B_i\left(z_i\right) - (1+\lambda)C_i(z_i) \tag{14}$$

The first order condition is $\partial W\left(z_i\right)/\partial z_i = 0$ or

$$B_{z_i}(z^*) - (1+\lambda)C_{z_i}(z^*) = 0 \tag{15}$$

with z^* denoting the first-best solution. At the optimum the marginal benefit from an additional medical treatment is equal to the social marginal cost. More extensively,

$$(1+\lambda)\left[c\left(\overline{\theta}, z^*\right) - c\left(\underline{\theta}, z^*\right)\right] f(z^*) = \left[b\left(\overline{\theta}, z^*\right) - b\left(\underline{\theta}, z^*\right)\right] f(z^*) \tag{16}$$

The optimal cut-off point is such that the cost savings from reimbursing one less surgical treatment and one more medical treatment is equal to the benefit foregone by one less patient receiving a surgical treatment.

The second order condition is $\partial^2 W/\partial z_i^2 < 0$ or, more extensively,

$$\partial^2 W/\partial z_i^2 = (1+\lambda)\left(c_s\left(\overline{\theta}, z_i\right) - c_s\left(\underline{\theta}, z_i\right)\right) f(z_i) - \left[b_s\left(\overline{\theta}, z_i\right) - b_s\left(\underline{\theta}, z_i\right)\right] f(z_i) < 0 \tag{17}$$

and is always satisfied. Since

$$\partial z^*/\partial\lambda = -\left(\partial^2 W/\partial z_i \partial\lambda\right)/\left(\partial^2 W/\partial z_i^2\right) = \left[c\left(\overline{\theta}, z^*\right) - c\left(\underline{\theta}, z^*\right)\right] f(z^*)/\left(-\partial^2 W/\partial z_i^2\right) > 0,$$

the purchasers optimal cut-off point is higher when the deadweight loss from raising public funds is higher, which implies a lower number of surgical treatments.

How does the optimum for the purchaser (z^*) compare with the solutions obtained when the tariff is respectively refined or unique?

When prices are refined the answer is clear. The hospital provides only the surgical treatment. There is therefore an overprovision of surgical treatments as $z^* > z^r = \underline{s}$.

If prices are unrefined, whether the cut-off point is too high or too low depends on the relative magnitude of the degree of altruism and the opportunity cost of public funds. Comparing the FOC in equation (7) with (15), it is straightforward to establish that the optimal cut-off point from the purchaser's perspective is higher than the one chosen by the provider, i.e. $z^u > z^*$, when $\frac{1}{1+\lambda} > \alpha \frac{H}{H-1}$.

Note that for a number of hospitals H sufficiently large, $H/(H-1) \approx 1$, so that $z^u > z^*$ when $1/(1+\lambda) > \alpha$: there is underprovision of surgical treatments when altruism is sufficiently small. Underprovision of surgical treatments always arises when the opportunity cost of public funds is zero ($\lambda = 0$). Then for any degree of altruism strictly less than one, $\alpha \in [0, 1)$, the provider chooses too few surgical treatments as $z^u > z^*$ when $1 > \alpha$.

In contrast, suppose that the provider is perfectly altruistic ($\alpha = 1$) and the opportunity cost of public funds is strictly positive ($\lambda > 0$), then there is certainly overprovision of surgical treatments as $z^u < z^*$ when $1/(1+\lambda) < 1$.

The following proposition summarises the results.

Proposition 1 *When the purchaser refines the prices, there is an overprovision of more intensive treatments: $z^* > z^r = \underline{s}$. When the purchaser does not refine the price, there is an underprovision of the more intensive treatment when the degree of altruism is sufficiently small $(1/(1+\lambda) > \alpha)$, so that $z^u > z^*$.*

In the following section we establish the conditions under which it is optimal for the purchaser to refine prices.

4.1 Welfare comparison

The purchaser obtains a higher welfare when prices are refined if

$$W(z^r) > W(z^u) \tag{18}$$

or more extensively if

$$
\begin{aligned}
& \int_{\underline{s}}^{z^r} \left[b\left(\underline{\theta}, s\right) - (1+\lambda)c\left(\underline{\theta}, s\right)\right] f(s)ds + \int_{z^r}^{\overline{s}} \left[b\left(\overline{\theta}, s\right) - (1+\lambda)c\left(\overline{\theta}, s\right) f(s)\right] ds \\
> & \int_{\underline{s}}^{z^u} \left[b\left(\underline{\theta}, s\right) - (1+\lambda)c\left(\underline{\theta}, s\right)\right] f(s)ds + \int_{z^u}^{\overline{s}} \left[b\left(\overline{\theta}, s\right) - (1+\lambda)c\left(\overline{\theta}, s\right)\right] f(s)ds
\end{aligned}
$$

which, after rearranging, gives:

$$
\int_{z^r}^{z^u} \left[b\left(\overline{\theta}, s\right) - (1+\lambda)c\left(\overline{\theta}, s\right)\right] f(s)ds \geq \int_{z^r}^{z^u} \left[b\left(\underline{\theta}, s\right) - (1+\lambda)c\left(\underline{\theta}, s\right)\right] f(s)ds \tag{19}
$$

Intuitively, this condition suggests that welfare is higher under refinement when the difference between benefit and cost for the marginal patients is higher when the surgical treatment is provided rather than the medical treatment. Alternatively, the above can be written as:

$$
\int_{z^r}^{z^u} \left[b\left(\overline{\theta}, s\right) - b\left(\underline{\theta}, s\right)\right] f(s)ds \geq (1+\lambda) \int_{z^r}^{z^u} \left[c\left(\overline{\theta}, s\right) - c\left(\underline{\theta}, s\right)\right] f(s)ds \tag{20}
$$

Refinement is optimal when the additional benefit for the patients with severity $s \in [z^r, z^u]$ receiving a surgical treatment rather than a medical treatment is higher than the corresponding additional cost.

Suppose that when prices are not refined, there is an overprovision of surgical treatments: $z^* > z^u$. Furthermore, when prices are refined, we have $z^r = \underline{s}$. It then follows that $z^* > z^u > z^r = \underline{s}$. Consequently, since welfare is concave in z and is highest at z^*, in this case welfare is higher when prices are unrefined and $W(z^*) > W(z^u) > W(z^r = \underline{s})$. Refining prices would only increase the number of patients receiving the surgical treatment, moving the hospital's choice of treatment further away from the purchaser's optimum. In summary, the purchaser should not refine.

Suppose instead that when prices are not refined, there is under provision of surgical treatments: $z^* < z^u$. The purchaser in this case faces a trade-off, as refinement leads to overprovision, while no refinement to underprovision. The loss of welfare under refinement is $W(z^r = \underline{s}) - W(z^*) < 0$, where $z^r < z^*$. The loss of welfare under no refinement is $W(z^u) - W(z^*) < 0$, where $z^u > z^*$. The purchaser should refine only if $W(z^r) - W(z^*) < W(z^u) - W(z^*)$, i.e. when the welfare loss from deviating from the optimal cut-off point are smaller under refinement.

5 Extensions

5.1 Surgical treatment is always more expensive

Suppose that the surgical treatment is always more expensive than the medical one for any level of severity, i.e. $c(\overline{\theta}, s) \geq c(\underline{\theta}, s)$ for any s. The other assumptions are the same as above. Figure 2 provides a graphical illustration.

As in the previous section, the optimal cut-off point under refinement is the lowest possible level of severity: $z^r = \underline{s}$. The FOC is

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = \alpha B_{z_i} - f(z^r)(\frac{H-1}{H})\left\{\left[\overline{p} - c\left(\overline{\theta}, z^r\right)\right] - \left[\underline{p} - c\left(\underline{\theta}, z^r\right)\right]\right\} < 0 \tag{21}$$

which is still negative for any cut-off point, since the difference in price is larger than the difference in the cost of the two treatments ($\overline{p} - \underline{p} > c\left(\overline{\theta}, z^r\right) - c\left(\underline{\theta}, z^r\right)$), and the surgical treatment is more beneficial ($\alpha B_{z_i} < 0$).

If the purchaser does not refine, the optimal cut-off point is

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = f(z^u)\left\{\frac{H-1}{H}\left[c\left(\overline{\theta}, z^u\right) - c\left(\underline{\theta}, z^u\right)\right] - \alpha\left[b\left(\overline{\theta}, z^u\right) - b\left(\underline{\theta}, z^u\right)\right]\right\} \geq 0 \tag{22}$$

Notice that for sufficiently low altruism, we have $\partial U_i(z_i, z_{-i})/\partial z_i > 0$: there is no guarantee that an interior solution is obtained. When altruism is low, a higher cut-off point may always be in the interest of the provider. Since the cost of surgical treatment is higher than the medical one, it is always in the provider interest to replace a surgical treatment with a medical one when altruism is sufficiently low. Indeed, if altruism is zero then this is always the case, and the optimal cut-off point is $z^u = \overline{s}$.

Under this latter scenario, refinement is beneficial when $W(z^r = \underline{s}) > W(z^u = \overline{s})$. The purchaser faces a stark trade-off: under refinement only the surgical treatment is provided (overprovision), while under no refinement only the medical treatment is provided (underprovision).

5.2 Medical treatment is more beneficial for low severity

Suppose that for some low levels of severity, the medical treatment is more beneficial than the surgical treatment. More precisely, we assume: $b(\underline{\theta}, \underline{s}) \geq b(\overline{\theta}, \underline{s})$: the benefit of the medical treatment is higher for patients with lowest severity; $b(\overline{\theta}, \overline{s}) \geq b(\underline{\theta}, \overline{s})$ the benefit of the surgical treatment is higher for patients with highest severity.

The FOC when prices are not refined is unchanged. When prices are refined, a corner solution might not arise anymore, and the cut-off point may be set above the minimum severity. This might arise when altruism is sufficiently high and the difference between the benefit of medical and surgical treatment is also sufficiently high (for lower levels of severity). Figure 3 illustrates such case. Analytically, since $\alpha B_{z_i}(z_i = \underline{s}) = \alpha \left[b(\underline{\theta}, z_i = \underline{s}) - b(\overline{\theta}, z_i = \underline{s}) \right] f(z_i = \underline{s}) > 0$, we might have that

$$\left.\frac{\partial U_i(z_i, z_{-i})}{\partial z_i}\right|_{z_i = \underline{s}} = \alpha B_{z_i} - f(\underline{s})(\frac{H-1}{H}) \left\{ \left[\overline{p} - c\left(\overline{\theta}, \underline{s}\right) \right] - \left[\underline{p} - c\left(\underline{\theta}, \underline{s}\right) \right] \right\} > 0. \tag{23}$$

The optimal cut-off point $z^r > \underline{s}$ satisfies the FOC

$$\alpha \left[b(\underline{\theta}, z^r) - b\left(\overline{\theta}, z^r\right) \right] f(z^r) = f(z^r)(\frac{H-1}{H}) \left\{ \left[\overline{p} - c\left(\overline{\theta}, z^r\right) \right] - \left[\underline{p} - c\left(\underline{\theta}, z^r\right) \right] \right\} \tag{24}$$

The cut-off point is such that the marginal benefit for the patients from an additional medical treatment, rather than a surgical one, is equal to the reduction in revenues. In summary, if prices are refined overprovision still arises but is not complete (both types of treatment are provided).

6 Conclusions

We have analysed hospitals incentives when more than one treatment is available for a given diagnosis, and when the purchaser faces the choice between setting a price which is identical across the different treatments (price is not refined), and setting a different price for each treatment (price is refined). Throughout the study, we have also assumed that prices are determined according to an average-cost rule, as observed in many OECD countries.

We find that if prices are refined, then the provider has always an incentive to overprovide the more intensive treatment. In contrast, when prices are not refined, providers may overprovide or underprovide the more intensive treatment. If altruism is low compared to the opportunity cost of public funds, then the provider underprovides the more intensive treatment when prices are not refined.

If providers overprovide the more intensive treatment when prices are not refined, then introducing refinement is not optimal. Refinement generates even stronger incentives to overprovide the more intensive treatment and therefore reduces welfare. However, if the providers underprovide the more intensive treatment when prices are not refined, but overprovide it when prices are refined, then the provider faces a trade-off. Whether refinement is optimal depends on the specific shape of the benefit and cost function.

Our main policy implication from this study is that in a second-best setting where prices are set equal to the average cost within each DRG, refiniment is not always optimal, and indeed may be welfare reducing.

7 References

Barros P. 2003. Cream-skimming, incentives for efficiency and payment system. *Journal of Health Economics* **22**: 419–443.

Beitia A. 2003. Hospital quality choice and market structure in a regulated duopoly. *Journal of Health Economics* **22**: 1011–1036.

Boadway R, Marchand M, Sato M. 2004. An optimal contract approach to hospital financing. *Journal of Health Economics* **23**: 85-110.

Brekke KR, Sørgard L. 2007. Public versus private health care in a national health service. *Health Economics* **16**: 579-601.

Chalkley M, Malcomson JM. 1998a. Contracting for Health Services with unmonitored Quality. *The Economic Journal* **108**: 1093-1110.

Chalkley M, Malcomson JM. 1998b. Contracting for Health Services when patient demand does not reflect quality. *Journal of Health Economics* **17:** 1-19.

Chalkley M, Malcomson JM. 2000. Government purchasing of health services. *Handbook on Health Economics*, Chapter 15. Culyer AJ, Newhouse JP (eds). North Holland; Elsevier: Amsterdam.

De Fraja G. 2000. Contracts for health care and asymmetric information. *Journal of Health Economics* **19**: 663-677.

Eggleston K. 2005. Multitasking and mixed systems for provider payment. *Journal of Health Economics* **24**: 211-223.

Ellis RP. 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics* **17**: 537-555.

Ellis RP, McGuire TG. 1986. Provider Behavior under Prospective Reimbursement. *Journal of Health Economics* **5:** 129-51.

Ellis RP, McGuire TG. 1990. Optimal payment systems for health services. *Journal of Health Economics* **9**: 375-396.

Gilman BH. 1999. Measuring hospital cost-sharing incentives under refined prospective payment. *Journal of Economics and Management Strategy* **8**: 433–452.

Gilman BH. 2000. Hospital response to DRG refinements: the impact of multiple reimbursement incentives on inpatient length of stay. *Health Economics* **9**: 277–294.

Hammond P. 1987. Altruism. In: Eatwell J, Milgate M, Newman P, (Eds.). TheNew Palgrave: A Dictionary of Economics. Macmillan, London: 85-87.

Jack W. 2005. Purchasing health care services from providers with unknown altruism.

Journal of Health Economics **24**: 73-93.

Levaggi R. 1996. NHS contracts: An agency approach. *Health Economics* **5**: 341-352.

Ma CA. 1994, Health Care Payment Systems: Cost and Quality Incentives. *Journal of Economics and Management Strategy* **3**: 93-112.

Malcomson JM. 2005. Supplier discretion over provision: theory and an application to medical care. *RAND Journal of Economics* **36**: 412-432.

MedPAC. 2007. Hospital acute inpatient services payment system. MedPAC, Washington, DC. http://www.MedPAC_Payment_Basics_07_hospital.pdf [5 August 2008].

McClellan M. 1997. Hospital reimbursement incentives: an empirical analysis. *Journal of Economics and Management Strategy* **6**: 91–128.

Mougeot M, Naegelen F. 2005. Hospital price regulation and expenditure cap policy. *Journal of Health Economics* **24**: 55-72.

Pope G. 1989. Hospital Nonprice Competition and Medicare Reimbursement Policy. *Journal of Health Economics* **8**: 147-72.

Rickman N, McGuire A. 1999. Providers' reimbursement in the health care market. *Scottish Journal of Political Economy* **46**: 53-71.

Shleifer A. 1985. A theory of yardstick competition. *RAND Journal of Economics* **16**: 319-327.

Siciliani L. 2006. Selection of treatment under prospective payment systems in the hospital sector. *Journal of Health Economics* **25**: 395-598.

8 Appendix. First order condition under refined prices

Differentiating with respect to z_i, we obtain:

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = \alpha B_{z_i} + \left(\underline{p}\frac{\partial \underline{n}_i}{\partial z_i} + \overline{p}\frac{\partial \overline{n}_i}{\partial z_i}\right) + \overline{n}_i\frac{\partial \overline{p}}{\partial z_i} + \underline{n}_i\frac{\partial \underline{p}}{\partial z_i} - C^H_{z_i} - C^L_{z_i} \tag{25}$$

Recall $\underline{p}(z_i) = \frac{\sum_i^H C_i^L(z_i)}{\sum_i^H \underline{n}_i(z_i)}$ and $\overline{p}\,(z_i) = \frac{\sum_i^H C_i^H(z_i)}{\sum_i^H \overline{n}_i(z_i)}$. We obtain

$$\frac{\partial \underline{p}}{\partial z_i} = \frac{C^L_{z_i}}{\sum_i^H \underline{n}_i} - \frac{\sum_i^H C_i^L(z_i)}{\left(\sum_i^H \underline{n}_i\right)^2}\underline{n}_{z_i} = \frac{C^L_{z_i}}{\sum_i^H \underline{n}_i} - \underline{p}\frac{\underline{n}_{z_i}}{\sum_i^H \underline{n}_i} \tag{26}$$

$$\frac{\partial \overline{p}}{\partial z_i} = \frac{C^H_{z_i}}{\sum_i^H \overline{n}_i} - \overline{p}\frac{\overline{n}_{z_i}}{\sum_i^H \overline{n}_i} \tag{27}$$

After substitution we obtain:

$$\begin{aligned}\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} &= \alpha B_{z_i} + \left(\underline{p}\underline{n}_{z_i} + \overline{p}\overline{n}_{z_i}\right) + \overline{n}_i\left(\frac{C^H_{z_i}}{\sum_i^H \overline{n}_i} - \overline{p}\frac{\overline{n}_{z_i}}{\sum_i^H \overline{n}_i}\right) \\ &\quad + \underline{n}_i\left(\frac{C^L_{z_i}}{\sum_i^H \underline{n}_i} - \underline{p}\frac{\underline{n}_{z_i}}{\sum_i^H \underline{n}_i}\right) - C^H_{z_i} - C^L_{z_i}\end{aligned} \tag{28}$$

or

$$\begin{aligned}\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} &= \alpha B_{z_i} + \underline{p}\underline{n}_{z_i}\left(1 - \frac{\underline{n}_i}{\sum_i^H \underline{n}_i}\right) + \overline{p}\overline{n}_{z_i}\left(1 - \frac{\overline{n}_i}{\sum_i^H \overline{n}_i}\right) \\ &\quad - \left(1 - \frac{\underline{n}_i}{\sum_i^H \underline{n}_i}\right)C^H_{z_i} - \left(1 - \frac{\overline{n}_i}{\sum_i^H \overline{n}_i}\right)C^L_{z_i}\end{aligned} \tag{29}$$

Since we assume identical hospitals, the optimal cut-off point for each hospital is the same in equilibrium and equal to $z_i = z_{-i} = z^r$. Also $\underline{n}_i = \underline{n}_j$ and $\overline{n}_i = \overline{n}_j$. Noting that $\underline{n}_{z_i} = f(z_i)$ and $\overline{n}_{z_i} = -f(z_i)$, the above simplifies to:

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = \alpha B_{z_i} + \left(1 - \frac{1}{H}\right)\left(f(z_i)\left(\underline{p} - \overline{p}\right) - C^H_{z_i} - C^L_{z_i}\right). \tag{30}$$

Substituting $C^H_{z_i} = -c\left(\overline{\theta}, z_i\right) f(z_i)$, $C^L_{z_i} = c\left(\underline{\theta}, z_i\right) f(z_i)$, we obtain

$$\frac{\partial U_i(z_i, z_{-i})}{\partial z_i} = \alpha B_{z_i} - \left(1 - \frac{1}{H}\right) f(z_i) \left(\overline{p} - \underline{p} + c\left(\underline{\theta}, z_i\right) - c\left(\overline{\theta}, z_i\right)\right) < 0. \tag{31}$$

Figure 1. Optimal cut-off point

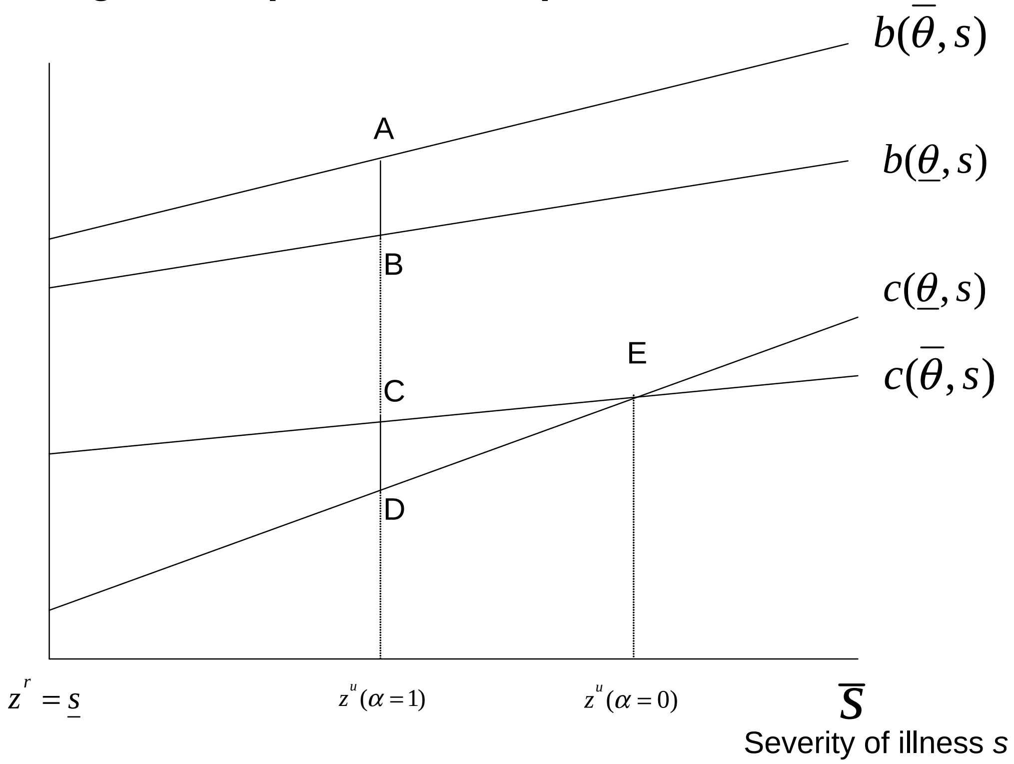

Figure 2. Surgical treatment is always more expensive

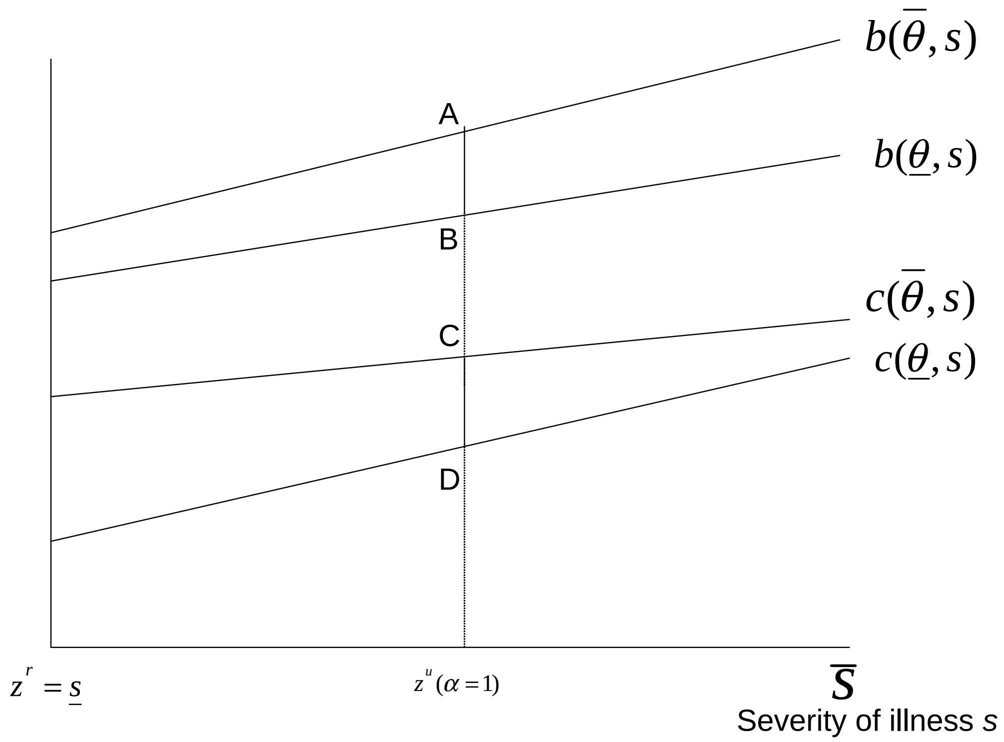

Figure 3. Medical treatment is more beneficial for low-severity patients