# Purchasing health care services from providers with unknown altruism[☆]

## William Jack[*]

*Department of Economics, Georgetown University, Room ICC 580,
Washington, DC 20057, USA*

## Abstract

Cost-sharing rules for paying physicians have been advanced as a way of generating incentives for the provision of quality care, while recognizing their potential negative effects on production efficiency. However, the optimal sharing rate typically depends on the degree to which the physician acts in the interest of the patient, what we identify as the physician's altruism. Since the degree of altruism is likely to vary across physicians, and to be private information, the standard rules for setting the cost-sharing rate are unlikely to be optimal. This paper derives conditions for the optimal non-linear cost-sharing mechanism in the presence of asymmetric information about altruism, and shows how it can sometimes be implemented through a menu of linear cost-sharing schemes. The model can be used to rationalize the design of the fund-holder system for general practictioners that operated in the 1990s in the United Kingdom.
© 2004 Elsevier B.V. All rights reserved.

[*] Tel.: +1 202 687 0773; fax: +1 202 687 6102.
*E-mail address:* wgj@georgetown.edu.
*URL:* www.georgetown.edu/faculty/wgj.

## 1. Introduction

It is now widely recognized among health economists that in budgeting resources for health care services, it matters as much how resource flows are implemented as how much of a resource envelope is available. That is, research has focused increasingly on how to purchase services, in addition to the standard question of how much health care should be provided. The latter kind of question can be addressed with tools of cost-benefit and cost-effectiveness analysis. This paper contributes to the literature on the first question: how to purchase medical care.

Much of the existing literature pertaining to this question has recently been thoroughly reviewed by Newhouse (2002). A recurring theme in the design of physician and hospital payment mechanisms is the need to provide incentives for the provision of quality care without cream-skimming, while also maintaining incentives for cost-control. Cost-reimbursement is good for the former, while fixed budgets are good for the latter, but it has proven perennially difficult to achieve both objectives simultaneously, except when consumers themselves are good judges of quality and can induce the delivery of high quality by threatening to switch providers (Ma, 1994; Glazer and McGuire, 1994).

An important element of a number of models of physician payment mechanisms is the extent to which providers value the benefits of care that otherwise accrue to patients. For example, Ellis and McGuire (1986) showed that if a medical care provider cares about net revenue and patient benefits in the same way that a social planner does (so the physician is a "perfect agent"), then fully prospective payment is optimal. There is no problem with under-provision of quality simply because it is assumed that providers care about it as much as the social planner does. If, on the other hand, providers place a smaller weight on patient benefits, then under prospective payment they will lean towards under-provision of quality. This distortion can be corrected by reimbursing some of the incurred costs: since there is a one-to-one relationship between costs and quality, subsidizing costs is the same as subsidizing quality, and the social optimum can be implemented. Thus, Ellis and McGuire's contribution has the flavor of corrective Pigouvian taxation—quality provision imposes a positive externality on consumers (because their benefits are under-valued by the provider) and so is subsidized.

Chalkley and Malcomson (1998) introduced the idea that there may not be a one-to-one relationship between costs and quality, so that if the physician places less than full weight on consumer well-being, the corrective subsidy inherent in Ellis and McGuire's cost-sharing rule would not yield the full optimum. In particular, costs depend on both quality (in the usual way) and effort exerted by the physician. The full optimum could be implemented by subsidizing quality directly, but only if this is contractible. Subsidizing costs, however, is not a perfect substitute for subsidizing quality, as it reduces incentives to exert effort. The size of the subsidy optimally employed depends on the extent to which the provider values consumer benefits. If the physician places any value on patient benefits, then a prospective payment will induce optimal cost-reducing effort, and some positive, but sub-optimal, level of quality. A small shift to cost-reimbursement is welfare improving, as the positive impact of increased quality is of first order, while the negative impact through higher costs is second order. Intuitively, as the physician values patient well-being more,

quality choices under prospective payment move towards the social optimum, and the size of the corrective cost-reimbursement rate falls.[1] Just as the Ellis and McGuire model is reminiscent of Pigou's (1947) analysis of externalities, Chalkley and Malcomson's account can be seen as being in the spirit of the Greenwald and Stiglitz (1986) examination of the second best.

In both papers, the optimal cost-sharing rate (which achieves the first best in Ellis and McGuire, and the second best in Chalkley and Malcomson) depends on the extent of agency – that is, on the extent to which the provider incorporates patient benefits when making decisions about quality and effort. It is arguable, however, that such preference parameters are at least as difficult to discern as the quality of services provided, if not more so. Indeed, it is very likely that providers vary widely in their altruism towards patients, so it would be necessary to pay different providers differently—the more benevolent having a smaller share of incurred costs reimbursed. There is no guarantee, however that faced with a menu of cost-sharing schemes designed for the range of provider types, each type would choose the cost-reimbursement scheme meant for him.

We employ standard techniques from the optimal regulation literature (Laffont and Tirole, 1993) to examine the way a purchaser should optimally contract with providers with heterogeneous and unknown degrees of altruism. These techniques have been used in the health economics literature previously, but when heterogeneity is with respect to productivity or cost. For example, Chalkley and Malcomson (2002) used the techniques to study the role of cost reimbursement when patients with a given administratively defined condition (i.e., a given DRG) differ in the severity of their illnesses. In their model, providers were identical, but there was demand-side heterogeneity in the sense that the severity of patient conditions within a given diagnostic related group, could differ across providers. Similarly, Jack (2002) assumed providers differed directly according to a measure of productivity in a model of the institutional design of a health system when decisions about hospital closures are included in contract design.

The next section sets up the structure of the model, including the technological relationship between physician effort, the quality of care, and costs, as well as a description of the objectives of the purchaser. In Section 3 the first best outcome is illustrated, when provider characteristics and actions are contractible. Section 4 examines the second best policy when just provider characteristics, but not quality, are observable, similar to Chalkley and Malcomson's (1998) analysis. Section 5 presents what we call the third best—the optimal payment mechanism when neither provider characteristics nor quality of care are observable. Section 6 presents an illustration of the use of the model in rationalizing the design of the UK's fund-holder experiment, wherein general practices were given the option of adopting fund-holder (high cost-sharing) status and continuing with the status quo (low cost-sharing). Section 7 briefly concludes.

---

[1] Glaeser and Shleifer (2001) present a model that provides a motivation for why firms might choose non-profit status. Their model is closely related to Chalkley and Malcomson's in that reducing the financial reward to cost reduction (i.e., having some degree of cost-sharing) reduces incentives to stint on quality.

## 2. Model set-up

In this section, we first describe the technological environment relating physician effort, quality of services, and financial costs, and then specify social objectives that are pursued by the purchaser.

### 2.1. Effort, quality, and costs

A health care purchaser (for example, the government) contracts with a population of medical care providers. Providers can choose the quality, $q > 0$, of care delivered, and can affect the financial cost, $C$, by exerting effort, $e > 0$. We assume $C(q, e) > 0$ satisfies the convexity assumptions $C_1 > 0$, $C_{11} > 0$, $C_2 < 0$, $C_{22} > 0$, and $C_{12} < 0$, where subscripts denote partial derivatives. It is helpful to think of $C$ as the cost the provider incurs when buying inputs (including the labor services of other staff), so that higher quality services cost more, but effort reduces these costs. Notice that the provider can "reduce costs" by simply providing a lower quality. There is no heterogeneity in $C(\cdot,\cdot)$ across physicians, so they all have the same case mix and labor productivity (or else differences in these factors have been accounted for).[2]

As well as the financial costs of provision, the provider suffers some extra disutility associated with her choice of quality and effort. Usually, we think of this as just the personal costs of effort that do not show up in purchased inputs, what Laffont and Tirole (1993) for example have denoted $\psi(e)$, where $\psi(\cdot)$ is positive, increasing, and strictly convex. However, in this paper, we admit the possibility that providers care about the quality of the services they provide. It is suggested that they derive some direct benefit $\theta\varphi(q)$ from providing quality $q$. $\theta \in [\theta_0, \theta_1] \subset \mathbb{R}^+$ parameterizes the strength of a provider's preference for quality, and $\varphi(\cdot)$ is positive, increasing, and weakly concave. Thus, the net disutility of providing quality $q$ and exerting effort $e$ is:

$$\gamma(q, e, \theta) = \psi(e) - \theta\phi(q). \tag{1}$$

Although, we shall sometimes refer to $\theta$ as a measure of benevolence or altruism, it should be noted that consumer well-being does not fully enter the provider's utility function. This is for two reasons: first, although the provider values quality (the direct benefits of which accrue to patients), she might not value it in the same way as consumers. That is, $\varphi(\cdot)$ might not necessarily measure the benefit of care to consumers. The second, and more consequential reason is that, while the provider values quality, she does not value cost savings in a similar way. Because, the financial costs of service delivery are borne by consumers through taxes, cost savings benefit patients, but the physician does not recognize these benefits when choosing the level and quality of medical care. This aspect of the formulation is consistent with much of the literature on physician objectives, as surveyed for example by McGuire (2000), although we recognize that, as McGuire notes (page 521–22), there is some empirical

---

[2] Other models in the literature (Ma and McGuire, 1997; McGuire, 2000) formulate the quality effort by writing quality as a function of effort and other inputs $z$, e.g., $q(z, e)$. If the prices of other inputs are fixed, this function can be inverted to yield $C(q, e)$ as above.

evidence suggesting that physicians respond also to the financial costs borne by consumers (Eisenberg, 1986).[3]

## 2.2. Social objectives

The purchaser cares about consumer and provider utility (i.e., consumer and producer surplus). Net consumer utility is gross benefits from quality, $S(q)$, increasing and concave, less the financial costs of paying for services. Following the standard literature, we assume the provider is reimbursed its financial costs $c$, and let $t$ be the transfer (if any) to the provider in excess of these costs. Thus, the gross transfer received by the provider is $t + c$. The consumer, however, must forgo an amount $(1 + \lambda)(t + c)$, where $\lambda > 0$ is the marginal excess burden of taxation. Thus, net consumer utility, or consumer surplus, is $S(q) - (1 + \lambda)(t + c)$.

The provider's net utility is $u = t - \gamma$, which is decomposed into net profit plus benefits from quality,

$$u = t - \gamma = t - \psi + \theta\phi \equiv \pi + \theta\phi.$$

Utility is decomposed in this fashion because, following Hammond (1987) and Chalkley and Malcomson (1998), we do not wish to include the direct benefits derived by the provider in social welfare. That is, the purchaser's objective function is:

$$W = S(q) - (1 + \lambda)(t + c) + t - \psi = S(q) - (1 + \lambda)(\psi + c) - \lambda\pi. \qquad (2)$$

Leaving profits with the provider is costly, simply because it costs money to transfer resources to them from consumers. A provider of type $\theta$ will agree to sign a contract only if her net utility is non-negative, $u \geq 0$. This in turn means that the purchaser must ensure that $\pi \geq -\theta\phi$.[4]

It is instructive to note the consequences of including the altruistic component of provider utility in social welfare. The welfare function then becomes:

$$W_+ = S(q) - (1 + \lambda)(t + c) + t - \gamma = S(q) - (1 + \lambda)(\gamma + c) - \lambda u.$$

Leaving utility rents to providers is again costly. If in both cases providers can be kept to their reservation utility levels, then the only difference between the formulations is that the quality of care optimally purchased under $W_+$ (and the health care budget allocated to providers) will be greater than that purchased under $W$. In both cases, the participation constraint is $u \geq 0$, and even in the third best, when this cannot be made to bind for all types of provider (i.e., for all $\theta$), the qualitative features of the optimal contract, such as the

---

[3] We note, however that in the model of this paper, all costs financed by consumers are channelled through the tax system, so it is arguable that providers would be much less responsive to increases in them than if they were paid directly by consumers at the point of service.

[4] It might be necessary in practice to add the constraint that $t > 0$, since, otherwise the provider would be required to pay the purchaser for the privilege of working.

direction of distortions to budgets and quality, are unaffected by the inclusion of altruistic benefits in the welfare function.

## 3. The first best

If a purchaser could write a contract based on quality and cost, and if it knew the type of provider it was purchasing from, it would set $t = \psi - \theta\varphi$ (i.e., $\pi = -\theta\phi$) and solve the following maximization problem:

$$\max_{q,e} S(q) - (1 + \lambda)(\psi(e) + C(q, e, \theta)) + \lambda\theta\phi(q),$$

with associated first order conditions for an interior solution:

$$S'(q) - [(1 + \lambda)C_q - \lambda\theta\phi'(q)] = 0, \tag{3}$$

and

$$-(\psi'(e) + C_e) = 0. \tag{4}$$

Denote the optimal values by $q^*(\theta)$ and $e^*(\theta)$. Quality should be expanded until the marginal benefit to consumers is equal to the marginal cost they bear (Eq. (3)). The marginal cost of quality borne by consumers is composed of a gross marginal payment to cover additional input expenses $(1+\lambda)C_q$, offset by the marginal saving on distortionary costs because providers are effectively partially paid in kind, $\lambda\theta\varphi'(q)$. Provider effort should be exerted to equate the marginal disutility, $\psi'(e)$, with the marginal cost savings, $-C_e$ (Eq. (4)).

To examine how the first best solution varies with $\theta$, write Eqs. (3) and (4) as $g(q, e, \theta) = 0$ and $h(q, e) = 0$, respectively. The second order condition for a maximum, which is assumed to be satisfied, requires in part that $\Delta = g_1 h_2 - g_2 h_1 > 0$. Optimal quality and effort choices vary according to:

$$\begin{pmatrix} q^{*'}(\theta) \\ e^{*'}(\theta) \end{pmatrix} = \frac{\lambda\phi'}{\Delta} \begin{pmatrix} -h_2 \\ h_1 \end{pmatrix}.$$

As $h_1$ is positive and $h_2$ is negative, quality and effort both increase with $\theta$ at the first best. These two effects mean that at the optimum, the variation of input expenditures $c^*(\theta)$ and the optimal transfer to the provider $t^*(\theta)$ with $\theta$ is not determined without further structural assumptions. The net utility earned is identically zero for all provider types, $u^*(\theta) = 0$.

## 4. The second best: non-contractible quality and effort

Suppose the purchaser attempts to implement the first best optimum by giving a provider of type $\theta$ a budget $c^*(\theta)$ with which to provide services, and a transfer or salary $t^*(\theta)$, to

compensate her for her effort costs.[5] If quality and effort choices are not directly observable, then this transfer will in general not yield the first best outcome. In particular, given any transfer $c$, as long as she is willing to participate, the provider will choose quality and effort to solve the following problem:

$$\min_{q,e} \gamma(q, e, \theta) \quad s.t. \quad C(q, e) = c. \tag{5}$$

Let $\hat{q}(c, \theta)$ and $\hat{e}(c, \theta)$ be the provider's optimal choices (assumed unique), which satisfy:

$$-\theta\phi' + \mu C_q = 0 \tag{6}$$

and

$$\psi' + \mu C_e = 0 \tag{7}$$

for some non-negative Lagrange multiplier $\mu(c, \theta)$. The convexity assumptions on $C$, $\varphi$, and $\psi$ imply that $\hat{q}$ and $\hat{e}$ are increasing in $\theta$. We further make the following assumptions regarding the dependence of $\hat{q}$ and $\hat{e}$ on $c$:

**Assumption 1.**

$$\frac{d\hat{q}}{dc} > 0 \qquad \frac{d\hat{e}}{dc} < 0,$$

$$\frac{d^2\hat{q}}{dc^2} < 0 \qquad \frac{d^2\hat{e}}{dc^2} > 0,$$

and

$$\frac{d^2\hat{q}}{d\theta dc^2} > 0 \qquad \frac{d^2\hat{e}}{d\theta dc} < 0.$$

These assumptions ensure that indifference curves in $(c, t)$-space are downward sloping and convex. The first pair of assumptions say that a higher budget leads to higher quality, but crowds out some effort. The second pair represent a kind of decreasing returns assumption. Further increases in a budget lead to diminishing quality improvements, and smaller crowding out effects on the supply of effort. Finally, the third pair say that the higher is $\theta$, the larger the effect on quality and effort of incremental budget resources. The first order effects of provider type and budgets are illustrated in Fig. 1.

Eqs. (6) and (7) are the same as the first order conditions for the first best optimum if and only if the budget is the right size (so that $\mu(c, \theta) = 1$), and if the provider values quality the same way as the purchaser, up to an additive constant: that is, only if $\theta\varphi(q) = S(q) + k$, for some constant $k$. To see this, note from (6) that if $c$ is set to ensure $\mu = 1$, the provider chooses quality to satisfy $C_q = \theta\varphi'(q)$. Substituting this into (3), we find that the social optimum is
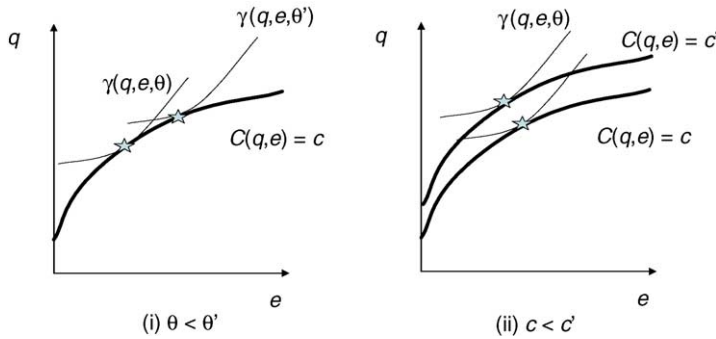
Fig. 1. Bold lines represent effort-quality pairs yielding constant financial costs, and thin lines represent constant disutility of provision lines. (i) More altruistic providers put in more effort and produce higher quality for the same budget. (ii) A larger budget leads to higher quality and lower effort.

attained only if $\theta\varphi(q) = S(q) + k$. Ellis and McGuire (1986) referred to a provider with such a coincidence of preferences as a "perfect agent".

If the provider is not a perfect agent for the purchaser (the interesting case), what is the best the purchaser can do? This is essentially the question addressed by Ellis and McGuire (1986) and Chalkley and Malcomson (1998). The purchaser's decision variables are now the budget $c$ and transfer $t$ made to the provider. We assume in this section that the purchaser knows the provider's type, $\theta$.

For a given budget $c$, denote the minimized disutility in (5) by $\hat{\gamma}(c, \theta)$, and let:

$$\hat{\gamma}(c, \theta) = \hat{\psi}(c, \theta) - \theta\hat{\phi}(c, \theta).$$

That is, $\hat{\psi}(c, \theta)$ is the cost of effort optimally incurred by the provider when given a budget $c$, and $\theta\hat{\phi}(c, \theta)$ is the direct benefit from quality optimally derived. The net disutility function, $\hat{\gamma}$, has the following properties, which are implied by our underlying assumptions:

$$\hat{\gamma}_2 = \gamma_3 = -\phi(\hat{q}(c, \theta) < 0$$
$$\hat{\gamma}_1 = -\mu(c, \theta) < 0$$

In addition, we make the following single crossing assumption:[6]

## Assumption 2.

$$\hat{\gamma}_{12} < 0 \tag{8}$$

---

[6] The single crossing property does not quite follow from our previous assumptions. To see this, note that:

$$\hat{\gamma}_{12} = \frac{d}{d\theta}\left(\frac{d\hat{\gamma}}{dc}\right) = \frac{d}{d\theta}\left(\psi'\frac{d\hat{e}}{dc} - \theta\phi'\frac{d\hat{q}}{dc}\right) = \psi'\frac{d^2\hat{e}}{dcd\theta} + \psi''\frac{d\hat{e}}{d\theta}\frac{d\hat{e}}{dc} - \phi'\frac{d\hat{q}}{dc} - \theta\phi'\frac{d^2\hat{q}}{dcd\theta} - \theta\phi''\frac{d\hat{q}}{d\theta}\frac{d\hat{q}}{dc}$$

Under our maintained assumptions, all the terms in this expression are negative, except the last. Note that if $\varphi$ is linear, then the single crossing property holds automatically.

Similarly, define the social benefit associated with a $\theta$-type provider providing services with a budget $c$ as:

$$\hat{S}(c, \theta) \equiv S(\hat{q}(c, \theta)).$$

Because, it knows $\theta$ and can thereby infer the provider's quality and effort choices, the purchaser can also infer disutility costs, $\hat{\gamma}(c, \theta)$. And because leaving rents to providers is inefficient, the purchaser sets $t = \hat{\gamma}(c, \theta)$. Redefine welfare from (2) in terms of $c$ and $\theta$ as:

$$\hat{W}(c, \theta) = \hat{S}(c, \theta) - (1 + \lambda)(\hat{\psi}(c, \theta) + c) + \lambda\theta\hat{\phi}(c, \theta). \tag{9}$$

From our assumptions on providers' supply of effort and quality, we are guaranteed that $\hat{W}$ is concave in $c$.[7] In addition, for later reference, we make a final assumption that the marginal welfare impact of budget resources delivered to more altruistic physicians is larger. That is,

**Assumption 3.**

$\hat{W}_{12} > 0.$

Since the level of effort (and hence, its marginal cost) is higher when $\theta$ is larger, this assumption is consistent with extra budget resources leading to a larger reduction in effort costs amongst more benevolent physicians.[8]

The optimal budget to give the provider then satisfies the first order condition:

$$\hat{W}_1(c, \theta) = \hat{S}_1(c, \theta) - (1 + \lambda)(\hat{\psi}_1(c, \theta) + 1) + \lambda\theta\hat{\phi}_1(c, \theta) = 0. \tag{10}$$

---

[7] Concavity of $\hat{W}$ is assured, since:

$$\hat{W}_{11} = \hat{S}_{11} + \lambda\theta\hat{\phi}_{11} - (1 + \lambda)\hat{\psi}_{11} = (S'' + \lambda\theta\phi'')\left(\frac{\mathrm{d}\hat{q}}{\mathrm{d}c}\right)^2 + (S' + \lambda\theta\phi')\left(\frac{\mathrm{d}^2\hat{q}}{\mathrm{d}c^2}\right)$$

$$- (1 + \lambda)\left(\psi''\left(\frac{\mathrm{d}\hat{q}}{\mathrm{d}c}\right)^2 + \psi'\left(\frac{\mathrm{d}^2\hat{e}}{\mathrm{d}c^2}\right)\right) < 0.$$

[8] Formally, a sufficient condition for $\hat{W}_{12} > 0$ is $\hat{S}_{12} > 0$ and $\hat{\psi}_{12} < 0$. Differentiating (but omitting function arguments):

$$\hat{S}_{12} = S''\frac{\mathrm{d}\hat{q}}{\mathrm{d}\theta}\frac{\mathrm{d}\hat{q}}{\mathrm{d}c} + S'\frac{\mathrm{d}^2\hat{q}}{\mathrm{d}\theta\mathrm{d}c}$$

and

$$\hat{\psi}_{12} = \psi''\frac{\mathrm{d}\hat{e}}{\mathrm{d}\theta}\frac{\mathrm{d}\hat{e}}{\mathrm{d}c} + \psi'\frac{\mathrm{d}^2\hat{e}}{\mathrm{d}\theta\mathrm{d}c}$$

Given our assumptions on the functions $\hat{q}$ and $\hat{e}$, $\hat{\psi}_{12} < 0$ as required. On the other hand, the first term in the expression for $\hat{S}_{12}$ is negative, while the second is positive. Assumption 3 is satisfied as long as the second term dominates the first.

Denote the solution to this condition by $\hat{c}(\theta)$. Under Assumption 3, the optimal budget is increasing in $\theta$. On the other hand, the additional transfer or salary, $\hat{\imath}(\theta)$ falls with $\theta$ since,

$$\hat{\imath}(\theta) = \frac{\mathrm{d}\hat{\gamma}}{\mathrm{d}\theta} = \hat{\gamma}_1 \hat{c}'(\theta) + \hat{\gamma}_2 < 0.$$

Thus, a more benevolent provider is given a larger budget because she makes it go further, but she is given a smaller salary as the value of the payment she receives in kind (derived from quality) is larger.

### 4.1. Implementation through cost-sharing

The second best optimum is a budget $\hat{c}(\theta)$ and transfer $\hat{\imath}(\theta)$. The provider is not thought of choosing $c$ in this interpretation—she just takes $c$ as fixed, and decides how much effort to exert. However, the purchaser can implement the same outcome using a cost-sharing rule. To this end, suppose the provider is able to choose the size of her budget, and that when she chooses $c$ she receives a transfer $\hat{T}(c)$. Let us write the provider's net utility as:

$$\hat{u} = (c, \theta) = \hat{T}(c) - \hat{\gamma}_1(c, \theta).$$

She simply chooses $c$ to maximize this expression, thereby, choosing $c$ to satisfy:

$$\hat{T}'(c) = \hat{\gamma}_1(c, \theta),$$

which is negative. Evaluate the right hand side of this expression at the second best optimum budget, $\hat{c}(\theta)$, and denote its absolute value by:

$$\hat{\tau}(\theta) = -\hat{\gamma}_1(\hat{c}(\theta), \theta).$$

If the provider is reimbursed costs $c$, and given an additional transfer:

$$\hat{T}(c) = \hat{\alpha}(\theta) - \hat{\tau}(\theta)c$$

for some constant $\hat{\alpha}(\theta)$, she will indeed choose the desired budget, $\hat{c}(\theta)$. Now just choose $\hat{\alpha}(\theta)$ to satisfy,

$$\hat{\alpha}(\theta) - \hat{\tau}(\theta)\hat{c}(\theta) = \hat{\gamma}(\hat{c}(\theta), \theta),$$

and the provider will earn zero net utility from the contract. Finally, the budget reimbursement and extra transfer components can be combined to yield a gross payment to the provider of:

$$\hat{P}(c) = \hat{T}(c) + c = \alpha(\theta) + (1 - \hat{\tau}(\theta))c \equiv \alpha(\theta) + \hat{\sigma}(\theta)c.$$

That is, the gross transfer to the provider is a fixed sum plus a share, $\hat{\sigma}(\theta) = 1 - \hat{\tau}(\theta)$, of realized costs. The parameters of these sharing schemes vary across providers. In particular,

$$\hat{\sigma}'(\theta) = -\hat{\tau}'(\theta) = \hat{\gamma}_{11}\hat{c}'(\theta) + \hat{\gamma}_{12} < 0, \tag{11}$$

so a more altruistic provider has a smaller share of realized costs reimbursed. This is consistent with the Chalkley and Malcomson result alluded to in the introduction—the

more the provider acts in the interests of the patient, the less should be the subsidy to costs (and hence, indirectly, quality).

As a special case, suppose $S(q) = \varphi(q)$ and $\theta = 1$, so the physician cares about quality in the same way as the purchaser, and cares about it with full weight, so is a perfect agent in the sense of Ellis and McGuire (1986). Condition (10) for the optimal budget reduces to:

$$\hat{S}_1(c, 1) = \hat{\psi}_1(c, 1) + 1,$$

and the cost-sharing rate that implements the optimum is,

$$\hat{\tau}(1) = \hat{S}_1(c, 1) - \hat{\psi}_1(c, 1) = 1.$$

A perfect agent should thus be given a budget and be made a residual claimant for all incurred cost savings. On the other hand, if $\theta \neq 1$ (note that we have not precluded the case $\theta > 1$), then it is straightforward to show:

$$\hat{\tau}(\theta) = 1 - \left(\frac{1 - \theta}{1 + \lambda}\right) \hat{S}_1(\hat{c}(\theta), \theta),$$

which is increasing in $\theta$. The more altruistic an agent is, the higher the rate of cost-sharing should be. Conversely, the less the physician cares about patient benefits (the lower is $\theta$), the more costs should be covered at the margin by the payer.

## 5. Third best contracts

We continue to preclude contracting on quality and effort, but now introduce the realistic assumption that provider type is not contractible, that is, that budgets and transfers cannot depend on $\theta$. This does not mean that it will be impossible to induce providers with different types to provide services at different financial cost, but it does impose costs on the purchaser from inducing such separation.

To see what might go wrong when contracts cannot be based on provider type, suppose the menu of second best contracts defined in the previous section is presented to physicians. There is now no guarantee that a given provider will choose the contract that is meant for her. This is most easily seen in Fig. 2, where we have depicted the optimal cost-sharing contracts for two different types of provider, with types $\theta^-$ and $\theta^+$, where $\theta^- < \theta^+$. The vertical axis in this figure measures the transfer in excess of financial costs made to the provider. There are two transfer schedules, $T(c, \theta^-)$ and $T(c, \theta^+)$, meant for the $\theta^-$ and $\theta^+$ types, respectively. Also shown are provider indifference curves, along which net utility is zero for each type. The second best budgets and transfers/salaries are at points $X^-$ and $X^+$, respectively. However, it is clear that, given the choice, $\theta^+$-type providers will opt to be paid under the $T(c, \theta^-)$ schedule and not the one designed for them, $T(c, \theta^+)$.

To investigate the effect that unobservability of $\theta$ has on the kinds of cost-sharing contracts that should be employed, we search for the optimal direct mechanism $(t(\theta), c(\theta))$, which represents a menu of transfer/budget pairs. A provider is asked to announce her type, and then given a budget $c(\theta)$ to spend on inputs, and a transfer $t(\theta)$ to compensate her for
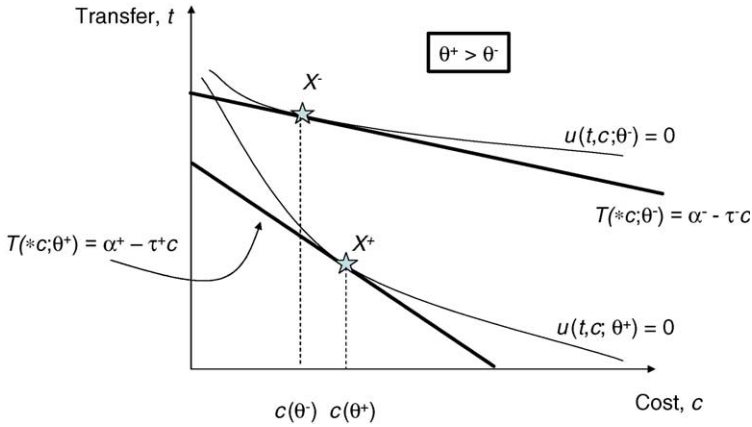
Fig. 2. The second best optimal contracts leave providers with zero net utility, and give a larger budget $c$, but smaller transfer $t$, to more benevolent providers.

disutility costs. However, there is no guarantee that this transfer will not exceed the net disutility suffered by the provider, that is, she might earn positive rents.

Indeed, let

$$\rho(\theta, \theta') = t(\theta') - \hat{\gamma}(c(\theta'), \theta),$$

be the rent earned by a provider of type $\theta$ when she announces her type to be $\theta'$. We invoke the revelation principle (Myerson, 1979) and confine ourselves to direct mechanisms that induce truthful reporting, that is, incentive compatible mechanisms. It is well understood (Laffont and Tirole, 1993) that the truth telling constraint, which implies $\rho_2(\theta, \theta) = 0$, reduces to the monotonicity condition:[9]

$$c'(\theta) \geq 0. \tag{12}$$

---

[9] Following the exposition of Laffont and Tirole (page 63), let $\theta > \theta'$. A mechanism $(t(\theta), c(\theta))$ induces truthful reporting if:

$$t(\theta) - \hat{\gamma}(c(\theta), \theta) \geq t(\theta') - \hat{\gamma}(c(\theta'), \theta)$$

and

$$t(\theta') - \hat{\gamma}(c(\theta'), \theta') \geq t(\theta) - \hat{\gamma}(c(\theta), \theta').$$

Add these two inequalities, eliminate the transfers and divide by $(\theta - \theta')$, yielding:

$$\hat{\gamma}_2(c(\theta'), \theta) \geq \hat{\gamma}_2(c(\theta), \theta),$$

which in turn implies:

$$\hat{\gamma}_{12} c'(\theta) \leq 0,$$

as $\theta > \theta'$. From (8) we infer the monotonicity condition, $c'(\theta) \geq 0$.

Net utility earned by the provider under an incentive compatible mechanism is:

$$u(\theta) = \rho(\theta, \theta) = t(\theta) - \hat{\gamma}(c(\theta), \theta),$$

which varies with $\theta$ according to:

$$u'(\theta) = \rho_1(\theta, \theta) + \rho_2(\theta, \theta) = \rho_1(\theta, \theta) \quad \text{(by incentive compatibility)}$$

$$= -\hat{\gamma}_2(c(\theta), \theta) = \hat{\phi}(c(\theta), \theta) \geq 0. \tag{13}$$

That is, under an incentive compatible mechanism, higher-$\theta$ providers, earn higher net utility.

Social welfare is the expected value of (9). We assume $\theta$ is distributed on $[\theta_0, \theta_1]$ with density function $f$, and cdf $F$. The welfare attained under an incentive compatible mechanism $(c(\theta), t(\theta))$ is thus:

$$\tilde{W} = \int_{\theta_0}^{\theta_1} [\hat{S}(c(\theta), \theta) - (1 + \lambda)(\hat{\psi}(c(\theta), \theta) + C(\theta)) - \lambda\pi(\theta)] \, f(\theta)\mathrm{d}\theta$$

$$= \int_{\theta_0}^{\theta_1} [\hat{S}(c(\theta), \theta) - (1 + \lambda)(\hat{\psi}(c(\theta), \theta) + c(\theta)) - \lambda\{u(\theta) - \theta\hat{\phi}(c(\theta), \theta)\}] \, f(\theta)\mathrm{d}\theta$$

$$= \int_{\theta_0}^{\theta_1} [\hat{W}(c(\theta), \theta) - \lambda u(\theta)] \, f(\theta)\mathrm{d}\theta, \tag{14}$$

where $\hat{W}(c(\theta), \theta)$ is defined in (9).

Use (13) to express the utility of provider $\theta$ as:

$$u(\theta) = u(\theta_0) + \int_{\theta_0}^{\theta} \hat{\phi}(c, \theta')\mathrm{d}\theta'. \tag{15}$$

From (14), provider utility should be as low as possible at the optimum, and since it must be increasing in $\theta$, the purchaser should set $u(\theta_0) = 0$. We use the standard trick of substituting (15) into (14) and integrating by parts to eliminate the dependence of welfare on provider profits, to get,

$$\tilde{W} = \int_{\theta_0}^{\theta_1} [\hat{W}(c, \theta) - \lambda H(\theta)\hat{\phi}(c, \theta)] f(\theta) \, \mathrm{d}\theta,$$

where $H(\theta) = (1 - F(\theta))/f(\theta)$. The purchaser's problem is to maximize this expression by choosing a cost function $c(\theta)$, subject to the monotonicity condition (12). Differentiating the integrand pointwise, we find,

$$\hat{W}_1(c, \theta) - \lambda H(\theta)\hat{\phi}_1(c, \theta) = 0. \tag{16}$$

Even though we have assumed $\hat{W}$ to be concave, the concavity of $\hat{\phi}$ means that the solution to (16), which we denote $\tilde{c}(\theta)$, need not be the pointwise maximizer of the integrand in $\tilde{W}$. Additionally, the solution to (16) is not necessarily non-decreasing. Both of these problems can be addressed by making suitable assumptions regarding the distribution of the altruism parameter. However, if it is assumed that $\phi(q) \equiv S(q)$ for all $q$ (so $\hat{\phi}(c, \theta) = \hat{S}(c, \theta)$), that $H(\cdot)$ is

non-increasing in $\theta$, and that there are sufficiently many physicians with the lowest possible value of altruism, then $\tilde{c}(\theta)$ will indeed solve the purchaser's maximization problem.[10]

**Proposition 1.** *Suppose $\varphi(q) \equiv S(q)$ for all $q$. If $f(\theta_0) \geq \lambda/(1 + \lambda\theta_0)$ and $H(\cdot)$ is non-increasing in $\theta$, then $\tilde{c}(\theta)$ solves the purchaser's maximization problem.*

**Proof.** Totally differentiating (16) we find:

$$\frac{d\tilde{c}}{d\theta} = -\left( \frac{\hat{W}_{12} - \lambda(H'\hat{\phi}_1 + H\hat{\phi}_{12})}{\hat{W}_{11} - \lambda H\hat{\phi}_{11}} \right).$$

If $H' \leq 0$, the numerator is positive. With $\hat{\phi} = \hat{S}$, the denominator is $\hat{S}_{11}(1 + \lambda(\theta - H)) - (1 + \lambda)\hat{\psi}_{11}$. Since $\hat{S}$ is concave and $\hat{\psi}$ is convex, if $1 + \lambda(\theta - H(\theta)) \geq 0$, the denominator is negative (and the second order condition is satisfied). If $H$ is non-increasing, then a sufficient condition for the denominator to be negative is that $1 + \lambda(\theta_0 - H(\theta_0)) \geq 0$, or,

$$f(\theta_0) \geq \frac{\lambda}{1 + \lambda\theta_0}.$$

□

In particular, if $\theta_0 = 0$, then as long as the first two conditions are met and $f(0) \geq \lambda$, $\tilde{c}(\theta)$ is the solution to the purchaser's problem. Comparison with (10) shows that the budget assigned to a provider of type $\theta$ is distorted away from its second best level, due to the extra term $\lambda H(\theta)\hat{\phi}_1$ in (16), except at $\theta = \theta_1$. Indeed, the expenditure budgets for all providers with $\theta < \theta_1$ are restricted below their second best optimal level, that is, $\tilde{c}(\theta) < \hat{c}(\theta)$. The intuition for this result is illustrated in Fig. 3. The second best contracts are not implementable, since the $\theta^+$ provider will choose contract $X^-$. The optimal pair of contracts $(Y^-, Y^+)$ involves an increased payment to the $\theta^+$ provider (to contract $Y^+$), who now earns a rent, and a reduction in the budget of the $\theta^+$ provider (to contract $Y^-$).

The provider's net utility under the third best optimal scheme (using (15)) is:

$$\tilde{u}(\theta) = \int_{\theta_0}^{\theta} \hat{\phi}(\tilde{c}(\theta'), \theta') \, d\theta',$$

which is increasing in $\theta$. The transfer, $t$, in excess of the budget must satisfy,

$$\tilde{t}(\theta) = \tilde{u}(\theta) + \hat{\gamma}(\tilde{c}(\theta), \theta). \tag{18}$$

---

[10] To interpret the optimality condition (16), rewrite it as:

$$[\hat{S}_1(c, \theta) - \hat{\psi}_1(c, \theta) + \lambda(\theta\hat{\phi}_1(c, \theta) - \hat{\psi}_1(c, \theta))]f(\theta) = (1 + \lambda)f(\theta) + \lambda(1 - F(\theta)\hat{\phi}_1(c, \theta). \tag{17}$$

A marginal increase in the budget of $\theta$-type providers (of whom there are $f(\theta)$) yields a direct margial benefit to consumers of $\hat{S}_1$, and a benefit to providers that is counted in social welfare of $-\hat{\psi}_1$. In addition, the purchaser can guarantee these providers the same utility by reducing the transfer they receive by the reduction in disutility, $-\hat{\gamma}_1 = \theta\hat{\phi}_1 - \hat{\psi}_1$ thereby saving $\lambda(\theta\hat{\phi}_1 - \hat{\psi}_1)$ in distortionary costs. On the other hand, the extra budget for all $\theta$-types imposes a direct social cost of $(1 + \lambda)f(\theta)$. Finally, the small increase in the budget raises the rent of each $\theta$-type provider by $\hat{\phi}_1$, and by (15), of all $(1 - F(\theta))$ providers with higher $\theta$'s. The social cost of this increased rent is the last term on the right hand side of (17).
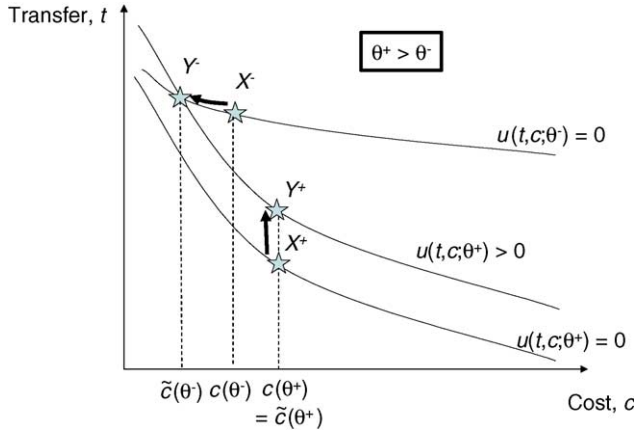
Fig. 3. Third best contracts distort the budget of low-$\theta$ providers downwards, and permit high-$\theta$ providers to earn a positive rent.

While under the second best contractual arrangement providers with higher values of $\theta$ received unambiguously lower salary transfers, that is $\hat{t}'(\theta) < 0$, at the third best optimum there is a tension between paying more benevolent providers less, because they are "cheaper," and paying them more, because of the incentive constraint. That is, differentiating (18) yields:

$$\tilde{t}'(\theta) = \underbrace{\tilde{u}'(\theta)}_{>0} + \underbrace{[\hat{\gamma}_1 \tilde{c}'(\theta) + \hat{\gamma}_2]}_{<0}. \tag{19}$$

### 5.1. Physician choice

Under the conditions of the previous section, $\tilde{c}(\theta)$ is strictly increasing in $\theta$, so define the inverse function $\tilde{\theta}(c)$, which gives the type of provider that should, at the optimum, choose budget $c$. The net transfer $\tilde{t}(\theta)$ can now be written as a function of the observed cost,

$$\tilde{T}(c) = \tilde{t}(\tilde{\theta}(c)) = \tilde{u}(\tilde{\theta}(c)) + \hat{\gamma}(c, \tilde{\theta}(c)).$$

Using the implicit function theorem, it can be shown that,

$$\tilde{T}'(c) = (\tilde{u}' + \hat{\gamma}_2)\tilde{\theta}'(c) + \hat{\gamma}_1 = \hat{\gamma}_1 < 0. \tag{20}$$

On the other hand, it is not possible in this model to establish the convexity properties of the transfer as a function of cost, since

$$\tilde{T}''(c) = \hat{\gamma}_{11}(c, \tilde{\theta}(c)) + \hat{\gamma}_{12}(c, \tilde{\theta}(c))\tilde{\theta}'(c),$$

and the first term is positive, while the second is negative.

If $\tilde{T}'' > 0$, so $\tilde{T}$ is convex, then it is possible to implement the optimal contract as a menu of linear contracts of the form,

$$\tilde{T}(c;\theta) = \tilde{\alpha}(\theta) - \tilde{\tau}(\theta)c, \tag{21}$$

where $\tilde{\alpha}(\theta)$ is a fixed salary component, increasing in $\theta$, and $\tilde{\tau}(\theta)$ is a cost-sharing component, also increasing, in $\theta$ (corresponding to the behavior of $\hat{\tau}(\theta)$ in condition (11) in Section 4.1). Providers are given a choice among the continuum of contracts $(\tilde{\alpha}(\theta), \tilde{\tau}(\theta))$, being paid a lump-sum $\tilde{\alpha}(\theta)$ and having a share $\tilde{\sigma}(\theta) = 1 - \tilde{\tau}(\theta)$ of expenses reimbursed. More altruistic providers choose contracts with higher fixed salaries and have a lower share of costs reimbursed. The intuition is that reimbursing costs is bad for cost-control incentives, but necessary to induce quality provision. Since, it takes less of an incentive to induce quality provision from more altruistic providers, their cost-control incentives can be strengthened by having them bear a larger share of costs, while not compromising quality too much. Having to pay more of the incurred costs of medical care means that, to ensure participation, they must be paid a larger salary component.

If $\tilde{T}$ is not convex, it is not possible to implement the third best using a menu of linear contracts. However, it may be possible to approximate the third best with such a menu. Let $\tilde{T}^H(c)$ be the function that describes the lower boundary of the convex hull of the set of points $\{(c, t): t \geq \tilde{T}(c)\}$, where $\tilde{T}$ is the third best optimum transfer function. By (20), $\tilde{T}^H(c)$ is downward sloping, and by definition it is convex, and so can be implemented with a menu of linear contracts of the form $\tilde{\alpha}^H(\theta) - \tilde{\tau}^H(\theta)$. An example is shown in Fig. 4.
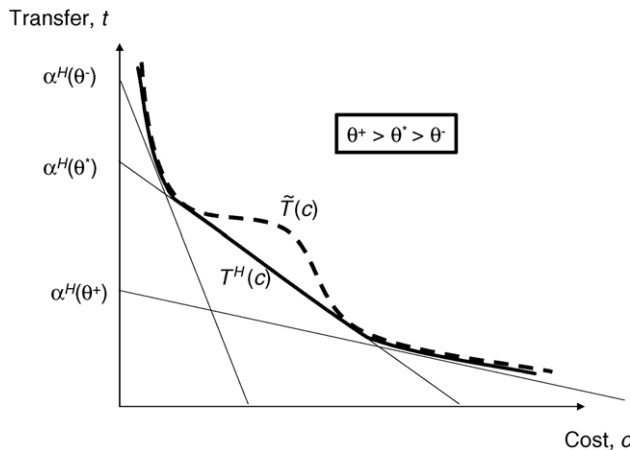


Fig. 4. In this example, the third best transfer function $\tilde{T}(c)$ is decreasing but not convex, so cannot be implemented with a menu of linear contracts. The convex function $\tilde{T}^H(c)$ approximates the third best, and can be implemented by offering providers a choice of payment schedules as shown.

## 6. An example and extensions

The general idea that agents can be given optimal incentives by allowing them to choose from a menu of linear incentive schemes is familiar from the regulation and procurement literature (Laffont and Tirole, 1993). An example in the sphere of medical care is the fund-holder experiment in the UK (Glennerster et al., 1994; Dusheiko et al., 2004), itself part of the creation of an "internal market" in the 1990s that aimed to increase the efficiency of resource use in the health sector. The reforms aimed broadly at improving the quality of services provided, while maintaining incentives for cost-control. Public financing of medical care was maintained, but the way providers were paid was changed so as to pursue these quality and cost goals. The objectives assumed of the purchaser of services in this paper correspond closely to these policy goals.

Between 1991 and 1999, larger general practices in the UK were afforded the option of becoming fund-holders, or continuing with the status quo. GPs who adopted fund-holder status were given a budget from which to finance certain non-emergency hospital-delivered secondary care. Unused budgetary resources could be used by fund-holders to purchase new equipment, and to even purchase additional services from themselves (Dusheiko et al., 2004). While perhaps not equivalent to a dollar of personal consumption, a dollar saved arguably thus contributed non-negligibly to net physician income. General practices that opted to continue under the existing arrangements neither bore the costs of secondary care directly, nor appropriated the savings that might have arisen from economizing on such care. By 1997, over half of the UK population was registered with a fund-holding general practice (Dusheiko et al., 2004).

Because GPs were given the option to become fund-holders, any empirical attempt to examine the effect of fund-holder status on referrals and costs must account for the obvious selection issues. For example, Dusheiko et al. (2004) assume physicians are heterogeneous with respect to a taste parameter, $\theta$. Those with higher $\theta$'s admit more patients for secondary care, because they perceive the gross benefit of admission to be higher—this means $\theta$ has similar behavioral implications as it would if it were interpreted as a measure of altruism. In the analysis of Dusheiko et al. (2004), physicians with higher values of $\theta$ are also less likely to choose fund-holder status, because of a fixed cost associated with such status (e.g., "... direct transaction costs and any distaste felt by GPs under the fund-holding regime about having to make explicit tradeoffs between income and elective admissions" (p. 8)), which is increasing in $\theta$. These two separate roles of the taste parameter induce a negative correlation between fund-holder status and admissions, which is useful for identification purposes. The authors find that accounting for selection effects, fund-holder status was associated with reduced admission rates and lower secondary care expenditures.

While the fund-holder experiment presented a unique opportunity to test the responsiveness of physicians to alternative financial incentives, it was clearly not an exercise in randomization. The model of this paper provides a normative rationale for why the policy allowed physicians to choose the financial constraints under which they operated: unobservable physician heterogeneity implies that optimal cost-sharing schemes that trade off rent extraction against effort incentives are typically non-linear, and they can be implemented (sometimes only approximately) by offering a menu of options. The two options in the UK system, which, using the notation of (21), correspond to cost-sharing rates of $\tilde{\tau} = 0\%$

(non-fund-holder) and $\tilde{\tau} = 100\%$ (fund-holder, with a correspondingly larger budget), represent such an approximation. This approximation could conceivably have been fine-tuned by introducing a third status, say with an intermediate budget and 50% cost-sharing.

This illustrative example of course does not definitively confirm the theoretical ideas explored in this paper, although it is indicative. The model supports the idea that GPs were given the choice of adopting fund-holder status not because it was thought that some of them should face weaker cost-control incentives on the margin, but rather that it would have been too expensive (in terms of the rents paid to other, more altruistic, physicians) to provide such incentives to all GPs. That is, if the government had mandated universal fund-holder status, the budget allocations to general practices needed to induce participation by all (especially low-$\theta$) physicians would have been inefficiently large.

A number of alternative explanations for including cost-sharing components in physician payment can be considered. For example, we have assumed an environment in which physicians face no uncertainty regarding costs and/or case loads. However, in a stochastic setting there is a role for cost-sharing to balance incentives against risk exposure (Gaynor, 1994), even if physicians are homogeneously altruistic.[11,12] On the other hand, while moral hazard rationalizes the use of cost-sharing mechanisms, it does not provide a good motivation for allowing physicians to sort amongst alternative contracts. If physicians differ in a relevant characteristic, for example, risk aversion, then such sorting, effected by offering a menu of options, could be efficient. We would expect to see more risk averse physicians being paid on (or close to) a fee-for-service basis, while less risk averse providers would be paid by (or close to) capitation. It is plausible that exhibited risk aversion could be linked to wealth, and that an empirical strategy would be to investigate the degree to which specialists sort themselves by wealth across cost-sharing regimes. It is less clear that altruism is empirically correlated with wealth, so that the absence of a correlation between the degree of cost-sharing chosen and wealth would, abstracting from obvious endogeneity problems, constitute evidence consistent with our theory.

Our model of physician heterogeneity can also be extended to understand patterns of physician compensation in markets with non-government payers, such as in the private insurance market in the US. The objectives of insurance plans (the buyers of physician services) would need to be specified differently to our representation of welfare, since the plans would have some profit motive. Even not-for-profit plans would be financed in part by prices (premiums) paid by individuals, and not solely through taxes

---

[11] The particular form of an optimal payment policy under conditions of moral hazard depends on the precise way in which physician effort affects costs. In a general principal-agent setting, a deductible policy (one with 100% cost-sharing up to a limit, and zero percent cost-sharing thereafter) is optimal if the agent can influence the ex ante probability of a costly event (e.g., an accident), but not the loss associated with that event (e.g., the extent of damage). On the other hand, if the agent can affect the ex post size of the loss, some kind of marginal (perhaps non-linear) cost-sharing is optimal. Since, it is likely that GPs have more control over the cost of treating patients who present with given conditions than reducing the chance that such conditions are developed, optimal payment mechanisms that traded off risk-protection against cost control incentives would tend to exhibit cost-sharing between zero and 100 percent.

[12] There was some degree of risk protection for GPs in the UK scheme, as fund-holders were only required to finance certain secondary care procedures known as "chargeable electives". Payments for "non-chargeable electives" and emergency care were made by the geographically defined health authorities (Dusheiko et al., 2004).

as in the current model. However, substituting a revenue function $R(q)$ for the surplus function $S(q)$ would lead to qualitatively similar predictions for paying heterogeneous physicians.

A second modelling choice arises when there is competition among plans, so that a plan is not a monopsonist in the physician labor market. One approach would be to assume that plans competitively bid for the labor services of physicians by offering employment contracts stipulating a budget ($c$) and compensation ($t$), and to find an equilibrium that satisfies zero profit and incentive compatibility constraints in the spirit of Rothschild and Stiglitz (1976). In her survey, Glied (2000) notes the wide variety of methods for paying physicians (salary, fee-for-service, etc.), based partly on research by Gold et al. (1995) who survey managed care plans. The model in this paper, and the alternatives suggested above, suggest that different methods of payment can be an efficient way of sorting physicians according to underlying (and possibly unobserved) heterogeneity (their altruism, risk aversion, or ethical standards). What is less obvious is whether this sorting should be across plans, in which case each health plan would pay its physicians on a single schedule (e.g., salaried physicians in HMOs versus fee-for-service providers in indemnity plans), or if each plan would have a mix of reimbursement schedules from which physicians could choose, but with less heterogeneity across plans. This line of enquiry awaits further research.

Finally, our model can be seen, in the spirit of Glaeser and Shleifer (2001) as a rationalization for allowing plans to voluntarily choose non-profit status. The model clearly does not address the revenue generating activities of non-profits (pricing, fund raising, etc.). However, since a profits tax can be interpreted as a tax on revenues plus a subsidy to costs (at the marginal tax rate), a plan that opts for non-profit status effectively bears the full burden of its financial costs, whereas the government shares some of the costs of a plan that selects for-profit status. The selection operates then at the plan level, and we might expect plans that have a direct concern for quality, as suggested by Newhouse (1970) adopt non-profit status.

## 7. Conclusions

Physician payment mechanisms need to generate incentives for the potentially conflicting goals of quality provision and production efficiency, at reasonable financial cost to the payer. Subsidizing costs can be a useful way to induce higher quality, as Ellis and McGuire (1986) pointed out. Chalkley and Malcomson (1998) observed that subsidizing costs can compromise production efficiency (discourage cost-reducing effort), thereby mitigating the role of cost-sharing to some degree. In both models, the appropriate cost-sharing rate depends on the extent to which consumer benefits enter the provider's utility function. This paper contributes to this literature by relaxing the assumption that the purchaser of care knows the extent of provider altruism. The variation in altruism across the population of providers is likely to be large, and the ability of a purchaser to base payment incentives directly on such information, were it available, would surely be limited. If contracts cannot be tailored explicitly to provider characteristics, the purchaser needs to design an incentive mechanism that induces providers to reveal their types. The result is, in general, a non-linear

scheme that relates payment to incurred costs. If this payment function is convex, it can be implemented by allowing physicians to choose from a menu of linear payment schedules, each characterized by a different cost-sharing parameter. In our model, however, it is not guaranteed that the optimal transfer function will be convex, in which case it can be approximated by such a menu of linear contracts.

## Acknowledgements

## References

Chalkley, M., Malcomson, J., 1998. Contracting for health services when patient demand does not reflect quality. Journal of Health Economics 17, 1–19.

Chalkley, M., Malcomson, J., 2002. Cost sharing in health service provision: an empirical assessment of cost savings. Journal of Public Economics 84, 219–249.

Dusheiko, M., Hugh, G., Rowena, J., Peter, S., 2004. The effect of budgets on gatekeeping doctor behaviour: evidence from a natural experiment. In: Paper presented at the Sixth Industrial Organization of Health Care Conference, Hyannis, MA.

Eisenberg, J.M., 1986. Doctors' Decisions and the Cost of Medical Care. Health Administration Press, Ann Arbor, MI.

Ellis, R., McGuire, T., 1986. Provider behavior under prospective reimbursement: cost sharing and supply. Journal of Health Economics 5, 129–152.

Gaynor, M., 1994. Issues in the industrial organization of the market for physician services. Journal of Economics and Management Strategy 3 (1), 211–255.

Glaeser, E., Shleifer, A., 2001. Not-for-profit entrepreneurs. Journal of Public Economics 81, 99–115.

Glazer, J., McGuire, T., 1994. Payer competition and cost shifting in health care. Journal of Economics and Management Strategy Spring 3 (1), 71–92.

Glied, S., 2000. Managed care. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. North-Holland (Chapter 13).

Glennerster, H., Matsaganis, M., Owens, P., 1994. Implementing GP Fundholding: Wild Card or Winning Hand? Open University Press, Buckingham.

Gold, M.R., Hurley, R., Lake, T., Ensor, T., Berenson, R., 1995. A national survey of arrangements managed care plans make with physicians. New England Journal of Medicine 333 (25), 1678–1683.

Greenwald, B., Stiglitz, J., 1986. Externalities in economies with imperfect information and incomplete markets. Quarterly Journal of Economics 101 (2), 229–264.

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

Jack, W., 2002. Institutional design and the closure of public facilities in transition economies. Economics of Transition 10 (3), 619–635.

Laffont, J.-J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press.

Ma, A.C.-T., 1994. Health care payment systems: cost and quality incentives. Journal of Economics and Management Strategy Spring 3 (1), 93–112.

Ma, Albert C.-T., McGuire, Thomas, 1997. Optimal health insurance and provider payment. American Economic Review 87 (4), 685–704.

McGuire, T., 2000. Physician agency. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. North-Holland (Chapter 9).

Myerson, Roger, 1979. Incentive compatibility and the bargaining problem. Econometrica 47, 61–73.

Newhouse, J., 1970. Toward a theory of non-profit institutions: an economic model of a hospital. American Economic Review 60, 64–74.

Newhouse, J., 2002. Pricing the Priceless: A health Care Conundrum. MIT Press, Cambridge and London.

Pigou, A.C., 1947. A Study in Public Finance, third ed. Macmillan, London.

Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Quarterly Journal of Economics 90, 629–649.