Cream-skimming, incentives for efficiency and risk adjustment*

Pedro Pita Barros
Universidade Nova de Lisboa
and *CEPR (London)*

This version: October 2000

Abstract

Reform proposals of health care systems in several countries have advocated variations of a risk adjustment/capitation system. These proposals face a serious objection: incentives to risk selection are prevalent in the system. By now, considerable literature has been devoted to finding ways of mitigating, if not eliminating, this problem, while at the same time preserving incentives to efficiency. We contribute to this debate presenting a transfer system that, under some circumstances, attains both provider efficiency and no risk selection. The transfer system extends typical linear capitation formulas. It can be interpreted as a fixed transfer in the beginning of the period plus an ex-post fund at the end of the period. The novelty rests in the way contributions to this fund are defined.

JEL numbers: I11, I18

Keywords: capitation, health care financing

Correspondence address:

Faculdade de Economia

Universidade Nova de Lisboa

Travessa Estêvão Pinto

P-1099-032 Lisboa

Portugal

Fax: 351-213 886 073

Email: ppbarros@fe.unl.pt

*I benefited from the comments of Fernando Branco, Bill Encinosa, Miguel Gouveia, John Hassler, Izabela Jelovac and Diogo de Lucena. Financial support from PRAXIS XXI/CEG/15/96 is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

The increasing relevance of the health care sector in modern economies, and the numerous reform attempts around the world, are, by now, well documented. One reform proposal in several instances is the development of sophisticated financing schemes involving capitation transfers from a funds' collector (the Government, for example) to health care purchasers and/or providers (the purchaser can coincide with the provider, like in HMOs and fund-holding GPs). Such transfers call for adequate risk adjustment. Capitation systems and risk adjustment have been used in the Netherlands, Israel, and the United Kingdom, among other countries. A similar system appears in proposals for the US, characterized by a single payer contracting with competing health plans (Newhouse, 1994). The issue of risk adjustment is also important in the US context, as clear incentives for risk selection in health plans have been empirically identified (Newhouse et al., 1997).[1]

Two main problems with this approach have been exposed in the literature, and several remedies to mitigate them have been put forward. The problems are risk selection, on the one hand, and providing incentives to efficiency of health care delivery, on the other hand.

Cream-skimming means selection by providers (or entities responsible for helth care provision) of those consumers expected to be profitable, given the system of risk-adjusted capitation payments. A central element is a framework already in place in some countries: a capitation system against which providers can play.

Avoidance of cream-skimming has been discussed along two main lines: adequate risk adjustment and pro-competitive regulation. In the latter, we typically see open enrollment rules and definition of standardized benefits. In the former, the two aspects more extensively investigated have been the refining of the risk-adjusted capitation, and the imposition of some sort of high-risk pooling.[2] The risk selection issue is far from being settled. In a recent account by Newhouse (1998) a dark picture of the future prospects is drawn, as the following quote illustrates:

> The physician treating the patient will have more information about the patient's likely future spending than the risk-adjustment formula will incorporate. As a result, the incentives to cream and dump will remain.

The recent paper by Van de Ven *et al.* (1998) discusses the current difficulties in im-

[1] For a detailed analysis of risk adjustment mechanisms in several countries, see Van de Ven and Ellis (2000), Tables 5 and 6.

[2] The literature on risk adjustment is too vast to be fully reviewed here. See the recent overview by Van de Ven and Ellis (2000).

proving capitation formulas by estimation of average risks. Marginal improvements in the capitation formula are obtained at a considerable research cost. Moreover, it is argued that avoidance of cream-skimming requires strong regulations and possibly some sort of mandatory high-risk pooling. In the papers by Shmueli *et al.* (1998) and Smith (1998), the same type of analysis is carried out, in the sense that both attempt to econometrically approximate risk adjusted capitations.[3] Newhouse (1996b) states that current adjustments to capitated payments do leave substantial "between-person variance in expected health care costs unexplained." Nonetheless, in a recent work, Van de Ven et al. (2000), based on simulations of risk-adjusted premiums, conclude that such risk-adjustment is the appropriate strategy to avoid risk-selection and dumping of patients.

Another way to proceed is to recognise the difficulties in estimating risk-adjusted capitations, which must not be vulnerable to superior information by recipients. This leads to the economic analysis of risk adjustment. In this vein, Glazer and McGuire (forthcoming, 1998) propose a different adjustment rule. This rule essentially establishes that above average payments must be associated with higher risk individuals. In a similar spirit, but in a different direction, Encinosa (1999) addresses the economic fundamentals behind risk adjustment in capitation rates. The major point raised by Encinosa is the importance of cost savings from treating more than one type of patients. A different way to proceed is to offer a menu of reimbursement policies to suppliers of health care, as explored by Sappington and Lewis (1999). They find a common result in the literature: both capitation (prospective) payment and cost-sharing (reimbursement) are present in the optimal transfer system.

An alternative to a simple capitation transfer has been considered: each insurance company contributes with a given percentage of its insureds to a common pool (a 5% value was used in the simulations conducted by Banerveld et al. (1995)). Typically, each firm contributes with the worst risks in the portfolio to this fund. The existence of the pool reduces the incentive to risk selection. The theoretical foundations of this approach seem to rest on a couple of assumptions. First, risk selection is increasingly costly as we move from the right-end tail of risk characteristics to the left. This means that cream-skimming with all risks is less costly than cream-skimming after the worst 5% of each company have been taken out to a compulsory pool. Second, the incentive for doing cost-reducing effort is not much affected by this 5% contribution to the pool.

The argument has not been formalized, up to our knowledge, although the reasoning behind it seems compelling. The merits of the approach have been discussed in the context of

[3] See also Van Vliet and Van de Ven (1992).

simulations. Besides these options, another one was explored by the Swiss federal government, which introduced a risk-equalization fund.[4] The idea of a fund is another avenue to counteract on cream-skimming. In this paper we explore its merits. The paper makes precise the desired characteristics of the fund. We focus on how transfers to health plans can be set such that provider-level efficiency is achieved, without giving rise to cream-skimming. The discussion is not only applicable to private health insurance plans, but also to sickness funds (Israel), Health Maintenance Organizations (US) and general practitioner fundholders (UK).

We propose a payment rule which, whithin the class of linear payments, enlarges the usual set of instruments. The payment rule, acting on routine information available, avoids the costs of very fine risk adjustment and does not promote an "information race" between the payer and the provider.

The reason why the system works is simple. We start, in the usual context, with three objectives: patients should pay an equal amount, whatever their initial condition is (equity considerations), moral hazard and risk selection problems should be minimised. However, only two instruments are available in the typical capitation/prospective payment story: a fixed amount and a proportion of actual costs. The payment rule we propose adds one more instrument – a reimbursement component based on the overall population risk mix.

Moreover, this seemingly complex rule turns out to have a fairly simple implementable form. It can be applied as a direct payment by patients (equal for each and every patient) plus a capitation from a central fund (again, just based on counts of enrollees), and an end-of-the-year adjustment fund.[5] The contribution to the fund will be determined by both risk selection and moral hazard effects. A provider that selects good risks will be a net contributor, thus reducing incentives to cream skim the market. A provider with high costs per patient will have an incentive to be more efficient. If the company has a lot bad risks, higher inefficiency means higher transfer. However, the actual costs of delivering health care to patients counteracts the incentive and keeps an interest of the provider in achieving efficiency.

The paper unfolds in the following way. First, in Section 2, the basic model is presented. The selection issue is highlighted in the model, as well as the incentive problem in the choice of cost-reducing effort. Next, in Section 3, the imposition of open-enrollment rules is addressed. The results confirm that to provide incentives to efficiency, capitation payments based on expected medical expenses must be partial and complemented with (equal) direct payments

[4] "Insurers having an age and sex structure favorable contribute, while others with unfavorable structure receive it." Beck and Zweifel (1998).

[5] We can also interpret it as a reimbursement system with a different ex-post fund. See discussion below for more details.

to insurers. This confirms the adequacy of the model to investigate the proposed issues. We proceed to the analysis of the optimal transfer system, linear on the expected medical expenses. In the absence of risk selection but with incentive problems present, the result accords with conventional wisdom. It prescribes a partial capitation system, supplemented by direct contributions. Then, in Section 4, we introduce risk-selection on the basis of an informational advantage by insurers, and show that an extended linear capitation system is able to solve both risk selection and cost-reduction incentives problems. This extended linear system has an interpretation in terms of a capitation, plus an ex-post fund. Its characteristics make, in principle, simple to implement this financing structure. Section 5 considers a different formulation for the efficiency-effort cost and shows how the transfer system has to be adjusted. Section 6 extends the model to the case of multiple health risks. Finally, Section 7 concludes.

2 The Model

For the sake of concreteness, we consider the organizational architecture described in Van de Ven and Van Vliet (1994), Van de Ven and Van Vliet (1992) and Van Vliet (1992), among others. As it is obvious the same ideas apply to many other settings where capitation transfers are an issue (like fund-holding GPs and health plans).

A paper close to ours is Ma (1994), in the sense that Ma looks at the cost-reduction and quality-enhancing incentive effects from different payment systems. The main result of his paper is that a mixed payment system (combining prospective payment and reimbursement) can give the appropriate incentives to achieve the socially optimal choices of quality and cost-reducing efforts, even if there is the possibility of dumping the worst cases. Our setting differs from Ma (1994) in several respects, although the main essential elements are present: discussion of payment systems and the issue of giving incentives to cost-reducing efforts.

The main differences are that we extend the possible payment system in one direction, different from the optimal rule derived in Ma (1994) (a piecewise linear rule), and no quality choices are considered. We introduce the cost-reducing effort as a moral hazard variable, as the transfer recipient is not the provider of health care, as in Ma (1994).

The on-going debate seems to indicate that competition coupled with risk adjustment is a promising strategy to contain health care rising costs. Newhouse (1996a) argues in this direction, and supports his view by a model which explicitly considers a sunk cost of developing new contracts. In the presence of sunk costs, a pooling equilibrium may be sustainable. The essence of the argument is that sunk costs make unprofitable to fully discriminate across consumers. We contribute to this debate by presenting a model that both (i) formalizes the arguments of risk selection and incentives to efficiency; and (ii) allows us to study the optimal transfer system.

The essential features of the model are the following. There is a central fund, endowed with a compulsory power to collect contributions from the population. These contributions are a function of income alone and denoted by $T(y)$, where y is individual income. In tax-financed systems, the function $T(y)$ is the tax system; in health systems financed by compulsory social insurance, this function denotes the way contributions are linked to earned income. The function $T(y)$ incorporates all the distributive (equity) concerns in the financing of health care systems.

For expositional purposes, we consider a homogeneous population with respect to health needs: in the event of illness, the amount of health care provided is same to all individuals. This assumption centers our attention on the essential features of the transfer system.

The central fund makes a payment to purchasers of health care (insurance firms, in our context), denoted by S. This capitation payment may not cover the full insurance premium. In this case, an extra payment is made by individuals directly to insurers. The exact value of this payment is determined by health care purchasers. Sticking the closest possible to real-life facts, we impose this direct payment to insurance companies to be the same for all consumers that choose a particular health insurance company. That is, insurance companies are not allowed to price discriminate consumers according to their risk characteristics. This type of constraint on the behavior of insurance companies has been recognized as inducing opportunities for risk selection.[6] The constraint of uniform prices across individuals is widespread. Not only several European countries have this feature in their national health systems; health plans in the US have it too. Thus, this is a constraint to be included in the model, also because it is at the roots of incentives to select risks.

The extensive form of the game is the following. In the first stage, the Government or the sickness fund defines the elements/parameters of the financing system, that is, the contribution function and the capitation rule. Next, in the second stage, insurance companies set a flat rate contribution to be paid directy by consumers. Then, consumers choose the insurance company to contract with. In the third stage, insurance companies set their effort level. Finally, Nature determines whether each consumer is ill, with probability p. Health care is demanded by sick consumers, and the contract terms are applied.

The insurance system provides full insurance. The assumption is intended to simplify the

[6] See Newhouse (1984) and Pauly (1984).

exposition. The introduction of contracts with less than full insurance has been advocated as a way to discipline demand (moral hazard on the consumption decision taken by sick individuals facing zero price at this point), and no extreme departure from relatively small copayments has been seriously proposed. These copayments are not regarded as main financing instruments. We choose not to clout our analysis of the financing system, and we leave out of the model these well-known arguments for positive copayments.[7]

The insurance company has to contract with health care providers the supply of health services. In case of illness, an individual must receive some given units of care (fixed across the population), which entail a monetary cost of X.

This monetary cost is dependent on the effort of insurance companies to get the best possible conditions (e.g. value for money). Denote by e the effort exerted by insurance companies. Then $X = X(e)$, with $X'(e) < 0, X''(e) > 0$: more effort decreases total expenses per patient but does so at a decreasing rate.

The effort is our efficiency variable and it is, by assumption, not observable (more rigorously, it is not contractible). The assumption of unobservable effort by insurance firms to achieve better provider performance is reasonable (and even more when the purchaser is integrated with the provider). It seems to be quite hard to specify in a contract, and monitor compliance with it, what should be the exact stance of insurance firms in negotiation processes with providers, for example, or in the specification, implementation and monitoring of medical protocols or guidelines. On the other hand, we assume that medical expenses $X(e)$ are observable and contractible.

Actual expenditures are determined by random elements and by the efficiency effort.[8] Implicitly, we assume that a given value of X can result from a choice of e but also from other non-observable elements. A word of caution is in order. The traditional principal-agent theory uses payment schedules to extract information regarding effort from the realized value of expenses. Here, we assume the third-party payer implements reward structures that are linear in realized treatment cost. As we restrict payment transfers to be linear, contracts of the sort "Pay S if X equals some value; pay nothing otherwise" are not possible. The set of feasible payments are more restrictive than it is usual in the principal-agent literature. However, the capitation transfers correspond to what we observe in reality, and simplicity of linear rules suggests that applied theory should use this type of assumptions if informative

[7] There is also the well-known argument of Rotschild and Stiglitz (1976) for offering contracts with less than full insurance. This does not appear here, as no discrimination between consumers is allowed.

[8] For a recent use of this assumption in health care settings, see Sappington and Lewis (1999). See Holmstrom (1979) for the seminal contribution of this type of models.

policy conclusions are to be drawn from the analysis. The same approach has been followed in previous literature (Ma, 1994, for example). Constraining the set of possible payments is a way of incorporating policy-implementation constraints into the analysis. As it is well known, focusing on linear rules comes at a generality cost. However, non-linear sharing rules are not observed often due, to considerable extent, to the informational requirements needed for implementation.

Exerting effort for cost containment implies a cost $C(e)$ per insured. The marginal cost of effort is positive and increasing, that is, $C'(e) > 0$ and $C''(e) > 0$. Examples of this cost are monitoring the care provided, checking the costs incurred by the provider, etc. The incentive for insurance companies to perform check-ups and early screening can be seen as included in the effort e. A higher e means pressure of insurance companies to get physicians to perform such screening activities ot to act as gatekeepers to differentiated care.

Only two states of nature are considered: healthy and ill. The 'ill' state of nature occurs with probability p. Individuals in the population differ in the probability of being ill. Population has unit size and the probability p is distributed in the population according to a density function $g(p)$ and a distribution function $G(p)$. Consumers are risk averse. We use a strictly concave function $u(\cdot)$ to represent the preferences over consumption in each state of the nature. Expected utility of a consumer endowed with illness probability p is:

$$V = U(y - T(y) - F) - p\bar{B} \tag{1}$$

where $\bar{B}$ is the utility cost of being ill net of the benefits from consumption of health care (treatment) and we take it to be zero, without loss of generality. Individuals take the contribution system to the central fund as exogenous to their decisions, and choose the insurance company which announces a lower value for the insurance premium F paid directy by consumers to insurance companies.

The insurance company maximizes expected profits, which are defined as:

$$\begin{aligned} \Pi_i \;=\; & \int_{p\in\chi_i} S_i(p, X(e_i), \chi_i)\, \tilde{g}(p \mid \chi_i)\, dp + F_i \int_{p\in\chi_i} \tilde{g}(p \mid \chi_i)\, dp - C(e_i) \int_{p\in\chi_i} \tilde{g}(p \mid \chi_i)\, dp - \\ & - \int_{p\in\chi_i} pX(e_i)\, \tilde{g}(p \mid \chi_i)\, dp \end{aligned} \tag{2}$$

where χ_i is the set of probabilities of consumers that choose insurance company i and $\tilde{g}(p \mid \chi_i)$ is the distribution of p restricted to the set χ_i. We also use χ_i to denote the set of consumers that choose firm i. There is some abuse of notation here: $\tilde{g}(p \mid \chi_i)$ *is not* the conditional distribution of p on χ_i. An example illustrates the notation. Suppose that for some p, two firms share equally the density of consumers, $g(p)$. Then $\tilde{g}(p \mid \chi_1) = \tilde{g}(p \mid \chi_2) = 0.5g(p)$. The first term in equation(2) is the expected transfer payment (sum of per capita transfer over all clients of company i, possibly risk adjusted). The second term is the value of insurance premiums paid directly by consumers to the insurance company. The third term denotes the cost of efficiency effort exerted by the firm. For the moment, we assume constant returns to scale of effort cost with respect clients.[9] Finally, the fourth term denotes expected payments of health care costs to treat consumers that become ill in group χ_i.

The assumption of competition in the insurance market and the rule that precludes the practice of insurance premium discrimination according to risk characteristics of the individual means that the only way to make some price discrimination is through cream-skimming (choice of best risks), which in our model is reflected in a selective choice of χ_i.

One way to act against this type of behavior is to impose rules design so that insurance companies cannot refuse to contract with a consumer – open-enrollment rules. Another way to avoid risk selection is to design the transfer rule S in an appropriate way. The next sections investigate the properties of each of these alternatives.

3 The traditional argument

Under a system of open enrollment, all insurance companies have to charge the same insurance premium to individuals, and they cannot refuse a consumer that accepts to pay the publicly announced premium.

According to the timing of the game, we first characterize the optimal efficiency effort by insurance companies. The effort choice by the insurance company is made after insurance contracts are signed. The insurance company chooses the effort level that maximizes its expected profits. This choice is characterized by the following first-order condition, obtained by maximizing expected profit (2) with respect to effort, e:[10]

$$\frac{\partial \Pi_i}{\partial e_i} = \int_{p \in \chi_i} \left(\frac{\partial S_i}{\partial X} X'(e_i) - C'(e_i) - pX'(e_i) \right) \tilde{g}(p \mid \chi_i)\, dp = 0 \tag{3}$$

The assumption of a competitive insurance market implies that, in equilibrium, insurance companies must have zero economic profits. The premium set by the company in a competitive insurance market must satisfy:

$$F_i = C(e_i) + \frac{\int_{p \in \chi_i} (pX(e_i) - S_i(p, X(e_i)))\, \tilde{g}(p \mid \chi_i)\, dp}{\int_{p \in \chi_i} \tilde{g}(p \mid \chi_i)\, dp} \tag{4}$$

[9]This assumption is relaxed below.

[10]Standard regularity assumptions ensure that second-order conditions are met.

Take now the system where S is a proportion of expected health care costs. This is, of course, a strong restriction on the transfer system. While a linear system can be advocated on the basis of its simplicity, the assumption of simple proportionality may be too restrictive, even within the set of linear transfer systems. In the next section, we will argue in favor of the use of a more general linear transfer system.

Since open enrollment rules are usually discussed in the context of proportional payment systems, we maintain this assumption in order to evaluate conventional wisdom on the characteristics of this organization of health care sector. The open enrollment requirement is not free of problems. For example, incentives to provide poor quality of care to high-risk individuals may appear. See further discussion in Van de Ven et al. (2000, p. 319). Accordingly,[11]

$$S_i = \alpha p X(e_i), \alpha \leq 1 \tag{5}$$

The constraint on α ensures that no transfer exceeds expected health expenditures. The equilibrium value for the insurance premium charged by insurance firms to consumers is:

$$F_i = C(e_i) + X(e_i) \frac{(1-\alpha) \int_{p \in \chi_i} p\, \tilde{g}(p \mid \chi_i)\, dp}{\int_{p \in \chi_i} \tilde{g}(p \mid \chi_i)\, dp} \tag{6}$$

A first question to answer is to know whether in equilibrium firms must have the same risk-mix of patients or not. A second question is to design the optimal transfer system in the model, given the constraints imposed upon the transfer system.

Before attempting to answer these two questions, it is useful to establish the following result.

Lemma 1 *Under open enrollment and in a competitive insurance market, insurance firms will choose the same effort level for cost reduction.*

Proof: The optimal effort choice by insurance companies is governed by the condition (for $\alpha < 1$)

$$-C'(e_i) = X'(e_i) \frac{(1-\alpha) \int_{p \in \chi_i} p\, \tilde{g}(p \mid \chi_i)\, dp}{\int_{p \in \chi_i} \tilde{g}(p \mid \chi_i)\, dp} \tag{7}$$

Define $\bar{p}_i$ as the average probability of illnes of the set of consumers that choose to contract with company i:

$$\bar{p}_i = \frac{\int_{p \in \chi_i} p\, \tilde{g}(p \mid \chi_i)\, dp}{\int_{p \in \chi_i} \tilde{g}(p \mid \chi_i)\, dp} \tag{8}$$

[11] The analysis can be easily generalized to $S_i = \alpha(p) X(e_i)$, thus being non-linear on the probability.

The zero-profit condition for each insurance company implies:

$$F_i = C(e_i) + X(e_i)(1-\alpha)\bar{p}_i \tag{9}$$

Consumers select the insurance company with the lowest value F_i. Thus, in equilibrium, it must be the case that $F_i = F, \forall i$. Otherwise, firms with $F_j > \min F_i$ will not have consumers and exit the market.

From the two equilibrium conditions (zero-profit condition plus the first-order condition for the choice of e_i), we obtain:

$$F_i = C(e_i) - X(e_i)C'(e_i)/X'(e_i) \tag{10}$$

As $F_i = F$, it must be the case that $e_i = e, \forall i$ (and $\bar{p}_i = \bar{p}, \forall i$), as the right-hand side is monotone on e_i. The equilibrium value of e_i is given by equation (7). The adjustment to equilibrium is made through the set of consumers χ_i that a firm has. QED The risk mix of consumers in a company's portfolio can, thus, differ across companies. The only constraint implied by a competitive market is an equal value of $\bar{p}_i$. Of course, this has some implications for the admissible configurations of consumers in their risk distribution across companies. The most important one is that an equilibrium with companies picking up only consumers at one tail of the distribution of p is not possible. Nonetheless, this is quite less restrictive than requiring an equal distribution of individuals across firms.

This simple model reproduces some conventional wisdom results. In particular, *for a pure reimbursement system, there is no incentive for insurance companies to pursue cost-reducing effort. For incentives to cost-saving effort to emerge, a partial capitation system is necessary.*[12]

This can be readily seen making $\alpha = 1$ (full capitation), and recalling the first-order condition for the optimal effort choice:

$$\left.\frac{\partial \Pi_i}{\partial e_i}\right|_{\alpha=1} = -C'(e_i)\int_{p\in\chi_i} \tilde{g}(p \mid \chi_i)\, dp < 0 \tag{11}$$

Thus, e_i is set at the lowest possible value.

This is a result from the general structure of moral hazard problems. The corollary is the need to set $\alpha < 1$ for a positive effort level result, which can be seen by straightforward computation of the effect of decreasing α in the level of effort, e_i:[13]

$$\frac{\partial e_i}{\partial \alpha} = \frac{\bar{p}_i X'(e_i)}{(1-\alpha)\bar{p}_i X'(e_i) + C''(e_i)} < 0 \tag{12}$$

[12] Other papers that have proposed a partial capitation are Ellis (1998), Ellis and McGuire (1986), Newhouse (1996, 1998) and Selden (1990), among others. See also Sappington and Lewis (1999).

[13] The denominator has negative sign as $\partial^2\Pi_i/\partial e_i^2 \leq 0$ to satisfy second-order conditions of the maximization problem.

It means that the capitation does not cover the full expected health care expenses if incentives for cost-reducing effort are to be provided. The remaining amount must be obtained through the charge F, in a competitive environment.

Open enrollment rules are seldom fully enforceable. Insurance companies still have at their will some ways to select risks which are of quite difficult monitoring. The emergence of cream-skimming even under open enrollment rules is leading to proposals of changes aimed at reducing risk-selection activities. The proposals, so far, had mainly an empirical support and are based on the idea of "high-risk pooling".[14]

To investigate the optimal transfer system, we must first define the objective function of the Government. We take it to be the welfare of consumers in the society, aggregated according to a utilitarian view. The maximization of this social welfare function is constrained by three sets of conditions: (i) the behavior of insurance companies with respect to the choice of cost-reducing effort; (ii) the competitive insurance market equilibrium condition; and (iii) the budget constraint on the financing of the system.[15]

The third constraint requires that contributions raised are enough to pay for the expected health care costs. This modelling option makes an extra assumption: higher efficiency in the system is passed through to consumers by the lowering of contributions. Another alternative use of savings associated with efficiency gains would be to maintain contributions and increase services in other areas of the Government budget. This would, however, divert us from the main features of the health system. The other assumption is that the contribution system does not give rise to inefficiencies in the economy. It is widely recognized that tax systems usually have some distortion cost. Since this cost is general to the design of tax systems, not specific to the health care sector, we opt to set it aside in the analysis.

At the present stage, the set of admissible transfers includes only transfers proportional to the expected health spending per capita. The problem of a social planner can be formally stated as:

$$\left.\begin{array}{lr} \max_{\{\alpha,e_i,F\}} & \int_Y h(y)U(y-T(y)-F)dy \\ s.t. & -X'(e_i)\bar{p}_i(1-\alpha)=C'(e_i) \\ & F_i = F = C(e_i)+X(e_i)\bar{p}_i(1-\alpha), \forall i \\ & T(y)=\alpha\bar{p}X(e_i) \end{array}\right\} \texttt{Program A}$$

where $h(y)$ is the distribution of consumers by income level, independently distributed from

[14] See Banerveld et al (1995).

[15] We retain the assumption of full insurance.

health risks, with support in Y; and $\bar{p} = \int_0^1 pg(p)dp$ is the average probability of illness in the whole population. Define $\Theta_A = \{(\alpha, e_i, F) :$ solves Program A$\}$. This set contains all the triples than can implement the socially optimal allocation.

Proposition 1 *Consider $(\alpha, e, F) \in \Theta_A$. Then $\alpha < 1$. That is, the optimal transfer system in a proportional rule framework implies a partial capitation transfer.*

Proof: Substituting the last two constraints into the objective function, we can rewrite the program as:

$$\max_{\{\alpha, e_i\}} \quad E_y[U(y - \alpha\bar{p}X(e_i) - C(e_i) - X(e_i)(1-\alpha)\bar{p}_i]$$
$$s.t. \quad -X'(e_i)\bar{p}_i(1-\alpha) = C'(e_i)$$

The solution to this problem originates the result. To show this, it suffices to write down the constraint to the problem as

$$\alpha = 1 - \frac{C'(e)}{-\bar{p}_i X'(e)} < 1 \tag{13}$$

QED Due to the need to provide incentives to efficiency effort developed by insurance firms, the capitation payment cannot cover the full expected health care cost.[16] Thus, in the specified set of admissible linear transfer schedules, the optimal transfer is a less than complete transfer. Again, the essential feature is that insurance companies anticipate the effect of own decisions of effort on the transfer received. A full transfer would provide no incentives for the development of efficiency effort. Thus, the usual proportional transfer cannot induce both the efficient provision of effort and absence of cream-skimming, when the open enrollment rule is not fully enforceable.

This completes our digression on the conventional wisdom about health care systems. We have replicated in a formal way the standard arguments for partial transfers of central funds to insurance companies, as an incentive device to cost-reducing efforts. We are now in position to present our main result.

4 An improved transfer system

The rationalization provided in the previous section is still vulnerable to issues of risk-selection that cannot be ruled out by open enrollment rules. We now explicitly recognize the cream-skimming problem. The problem was referred to above in general terms, as selective choice

[16] Moreover, in equilibrium, $\bar{p}_i = \bar{p}, \forall i$.

of consumers. We now introduce it in a more specific way. However, by enlarging in a simple, and feasible, manner the set of linear transfers, we are able to propose a new transfer system that solves both the cream-skimming and the incentives-to-efficiency problems.

One feature that caracterizes real world health insurance markets is an informational advantage of insurance companies (or of providers) over the Government (or the central fund). The advantage of a better knowledge of risk characteristics of individuals can translate into a risk selection activity by insurance companies. This, of course, stems from the imposition of the constraint of no price discrimination across consumers.

In terms of economic modeling, one way of introducing this informational advantage is to assume that the central fund cannot identify the risk characteristic p, while the insurance company is able to observe it.

Maintaining the assumption of a transfer system proportional to health care expenses, under the assumed informational asymmetry between insurance companies and the Government, the value to be transferred must be the same for all individuals, as the Government is not able to differentiate. This implies that only the individuals with the lowest value of p would be chosen by insurance companies, which can mean that consumers endowed with a high probability of illness will not be able to get private insurance or do so at rather "unfavorable" terms, from a society's point of view. In the absence of an open enrollment rule, or if there is no way of monitoring compliance with it, the transfer system must provide the incentives for firms to not engage in cream-skimming activities.

An obvious way to eliminate the negative effect from cream-skimming that some people (the worst risks) are not able to get health insurance is to set a capitation transfer high enough, such that all individuals are profitable to insurance companies.[17] The distributive issues involved (high profits for insurance companies) as well as efficiency considerations (distortion costs arising from the taxes/contributions levied to get the necessary contributions) make it an undesirable solution.

We are then left with the alternative of determining the optimal transfer system. The natural option of refining the capitation formula, in a way such that no significant informational asymmetry exists between the central fund and insurance companies, has obvious difficulties. It may lead to a cost escalation in the monitoring and management of the system, with no significant effect in decreasing cream-skimming opportunities.

The search for the optimal transfer system must be done in a careful way. With the

[17] This does not avoid competition for the better risks. It, however, does eliminate the incentive to reject consumers.

assumption of informational asymmetry, the transfer must be a fixed payment (or, at least, based on a coarser risk discrimination than the one available to firms). And any fixed payment is not able to counteract on the incentive to cream-skim the market. However, through enlargement of the set of instruments available, we may consider the option of creating an ex-post clearing fund, which uses information available on a routine basis. The existence of financial flows with an ex-post clearing fund makes the transfer relevant to the insurance company to include not only the fixed capitation payment but also the ex-post adjustment, as the insurance company will anticipate and internalize the effect of risk selection on the ex-post contribution to this clearing fund. We now show formally this argument.

Due to its simplicity, we restrict our analysis to linear schedules. The linear transfer per capita we propose is of the form:

$$S_i = \alpha_0 + \alpha_1 \bar{p}_i X(e_i) + \alpha_2 \bar{p} dp X(e_i) \tag{14}$$

The transfer if defined by a fixed amount α_0, plus a proportion α_1 of expected health expenditures, where the expectation is based in the risk mix in the set χ_i, and a proportion of expected health expenses, where the expectation is based on average population risk. The definition of the optimal transfer rule consists in the specification of the parameters $(\alpha_0, \alpha_1, \alpha_2)$, which fully characterizes the transfer S_i. The main (linear) payment structures analyzed in previous literature are particular cases of equation (14). For $\alpha_1 = \alpha_2 = 0$, one has the pure prospective system, while for $\alpha_0 = \alpha_2 = 0, \alpha_1 = 1$, we have the pure full cost reimbursement model. The partial cost-sharing model emerges for $\alpha_0 > 0, \alpha_1 < 1$ and $\alpha_2 = 0$.

The profit of insurance company i is:

$$\Pi_i = \int_{p \in \chi_i} \left(F_i - C(e_i) + S_i - pX(e_i)\right) \tilde{g}(p \mid \chi_i)\, dp \tag{15}$$

Substituting for the transfer system in the profit of the insurance company:

$$\Pi_i = G_i \left(F_i - C(e_i) + \alpha_0 + \alpha_1 \bar{p}_i X(e_i) + \alpha_2 \bar{p} X(e_i) - \bar{p}_i X(e_i)\right) \tag{16}$$

where $G_i = \int_{p \in \chi_i} \tilde{g}(p \mid \chi_i)\, dp$ and $\bar{p} = \int_0^1 pg(p)dp$.

The transfer system must comply with two different sets of constraints: absence of cream-skimming and provision of incentives to cost-reducing effort. Cream-skimming usually means, in health insurance contexts, selection of good risks. That is, consumers who have expected health expenses far below the (risk-adjusted) capitation payment. We formalize this intuitive notion in terms of a contraint of “no cream-skimming” in the following way. Take a partition χ_i of consumers from the population. There is no incentive for cream-skimming provided that any other partition χ_j gives the same expected profit per capita to the insurance company. That is, $\Pi_i(\chi_i)/G_i = \Pi_i(\chi_j)/G_j$, where G_i(resp. G_j) is the number of consumers in partition χ_i(resp. χ_j). This definition is somewhat different from the intuitive notion, as it does not concentrate only on the idea of "good risks". This is so because we allow for configurations where an insurance company has both very good and very bad risks, but the risk-mix is, on average, similar to that of other insurance companies that take only intermediate risks (recall that the competitive equilibrium does not restrict consumers' distribution to be equal across companies). It also accommodates the issue that a bigger partition (more consumers) may imply more good-risk consumers (in absolute terms), and also more bad risks. It is, therefore, important to define the notion of "no cream-skimming" in terms of expected per capita profit to the insurance company.

The decision problem of the Government (or the health plan) is given by Problem B, defined by:

$$\left.\begin{array}{c} \max_{\{\alpha_0,\alpha_1,\alpha_2,F,e\}} \int_Y U(y-T(y)-F)h(y)dy \\ s.a. \int_Y T(y)h(y)dy = \sum_i \int_{p\in\chi_i} S_i\,\tilde{g}(p \mid \chi_i)\,dp \\ S_i = \alpha_0 + \alpha_1 \bar{p}_i X(e_i) + \alpha_2 \bar{p} X(e_i) \\ \Pi_i(\chi_i)/G_i = \Pi_i(\chi_k)/G_k, \forall i,k \\ \Pi_i(\chi_i) = 0, \forall i \\ -C'(e_i) + (\alpha_1\bar{p}_i - \bar{p}_i + \alpha_2\bar{p})X'(e_i) = 0, \forall i \end{array}\right\} \texttt{Program B}$$

Define the solution to `Program B` as $\Theta_B = \{(\alpha_0.\alpha_1, \alpha_2, e_i, F) :$ solves `Program B`$\}$.

Proposition 2 *Take $\theta \in \Theta_B$. Then, $\alpha_1 = 1$ and $\alpha_2 = -1$. That is, the optimal transfer is of the form:*

$$S_i = \alpha_0 + X(e)\,(\bar{p}_i - \bar{p}) \tag{17}$$

We proceed now to a sketch of the proof (a formal proof is provided in the Appendix). For cream-skimming to be unprofitable, it comes immediately that $\alpha_1 = 1$. In this case, there is no incentive of the insurance company to engage in risk selection. The individual profit earned in each consumer is given by the term in parenthesis in equation (16). This is independent of $\bar{p}_i$, the only variable determined by the set of consumers. There is no gain in making a selective choice of consumers. With $\alpha_1 = 1$, the private optimum for the choice of cost-reduction effort is given by:

$$-C'(e_i) + \alpha_2 \bar{p} X'(e_i) = 0 \tag{18}$$

To find the optimal value of α_2, which has the role of aligning the private incentives with the social incentives in the choice of effort, we need to look at the social planner's problem.

Note that the proportional rule restricts $\alpha_2 = 0$, which means that effort is set at its lowest possible value, unless $\alpha_1 < 1$ as shown previously.

Taking the social welfare function, the problem to be solved by the social planner is:

$$\begin{aligned} \max_e W &= \int_Y h(y)U(y - T(y) - F)dy \\ s.t. & \int_Y (T(y) + F)\, h(y)dy = \int_0^1 (C(e) + pX(e))\, g(p)dp \end{aligned}$$

Substituting the budget constraint into the objective function, the problem can be written as:

$$\max_e W = \int_Y U(y - T(y) + \int_Y T(y)h(y)dy - C(e) - \bar{p}X(e))h(y)dy \tag{19}$$

The first-order condition of this problem is

$$E_y\left(U'(\cdot)\right)\left(-C'(e) - \bar{p}X'(e)\right) = 0 \tag{20}$$

or simply

$$C'(e) = -\bar{p}X'(e) \tag{21}$$

as marginal utility is always positive. Going back to the problem of defining the optimal transfer system, to have the social optimum choice implemented, it suffices to make $\alpha_2 = -1$.

This implies a transfer formula given by:

$$S_i = \alpha_0 + X(e)\left(\bar{p}_i - \bar{p}\right) \tag{22}$$

We can interpret this transfer system as a lump-sum payment, α_0, plus an ex-post adjustment fund, $X(\bar{p}_i - \bar{p})$. The proposed transfer formula allows for a socially optimal choice of effort and at the same time it gives no incentive to cream-skimming. The value α_0 is determined, jointly with F, by the zero-profit equilibrium condition for the firm. In particular, either a value $\alpha_0 = 0$ or a value α_0 such that $F = 0$, are possible. The perfect substitutability between F and α_0 in the optimal solution is easily observed in the zero-profit condition for insurance companies.

The choice of a particular value for α_0 depends on the social planner's preferences about payment to insurance companies in a direct way and through the central fund. That is, the precise value α_0 hinges upon the social preferences associated with fund transfers and direct payments. The former may be associated with both income redistribution (positive effect) and heavier use of compulsory power to set contributions according to income, giving rise to other distortions in the economy (a negative effect). The latter has a negative equity effect but induces smaller distortions in the economy. We do not formalize this choice as it is not essential to our point.

Although we casted the payment structure in terms of a capitation followed by an ex-post fund, the same structure can be reinterpreted in the following way:

$$S_i = \alpha_0 + (\bar{p}_i X_i - \bar{p} X_i) = \bar{p}_i X_i + (\alpha_0 - \bar{p} X_i) \tag{23}$$

Thus, we can see it as a reimbursement system followed by an ex-post adjustment. This is so because a reimbursement component must be given to eliminate incentives for risk selection.

The ex-post adjustment fund has two desirable properties. First, it is a financially balanced scheme (except for the operating costs of the system).[18] Second, the system induces a socially optimal level for the cost-reducing effort.

The first term in the adjustment formula, $\bar{p}_i X(e_i)$, is simply the value of health care costs per capita within the insurance company consumers. The second term, $\bar{p} X(e_i)$ can be approximated, for implementation purposes, by the ratio of "sick" people in total population, multiplied by the average per capita cost of individuals that have been sick in the company's set of consumers. The information and computations required are of the same order of complexity as other ex-post funds.

This payment system implies that gains from cost reduction stay with the firm, provided it does not engage in cream-skimming. As the incentive to select consumers is eliminated, because the net profit per consumer does not depend on individual risk characteristics, the payment system provides incentives for efficiency. This constitutes the crucial difference to other adjustment rules based on average cost per capita in the whole health care sector.

Some comments are in order here. First, if there is some categorization of risks, by sex, age or other risk adjuster, the same idea of transfer system can be applied within each type of risk that is identified by the central fund.

This capitation system does not provide incentives for cream-skimming, retains incentives for efficiency and is of relatively easy computation. For the diseases to be explicitly covered by the capitation transfer, the information needed, at the company level, amounts to total expenditure per disease, the number of people treated of each disease and the number of insured consumers.

There is a recurrent issue in health care financing discussions: that lower quality of care may be induced by reforms. In our discussion, ensuring quality in provision may be seen

[18] A proof of this claim is provided in the Appendix.

as one of the tasks of insurance companies. Thus, obtaining quality is embodied in the effort variable. In this sense, also an optimal quality level will be provided, as long as direct contracting of the central fund with providers is not more informative in quality issues than using delegation to insurance companies.

The explicit introduction of quality in provision raises several issues. Up to our knowledge, in a similar context to ours, only Encinosa (1999) states its role. In his model, the problem of choice of quality is solved in a very simple way, as it is set at the maximum value to attract consumers. Of course, other papers have treated quality as an (un)observable element in the definition of the payment system. These discussions are complementary to our main point, which is related to incentives for efficiency by intermediate agents, like insurance companies and Health Maintenance Organizations.

Another feature of the extended transfer system we propose is administrative feasibility. The information needed to implement it is available on a routine basis, with a reasonable degree of accuracy and a (presumably) small time lag. Below, we provide an illustration of the transfer system.

5 Without constant returns to scale in effort cost

A fair criticism of the model in the previous section is the assumption that efficiency-effort costs are determined in a per capita basis. One readily find examples of efforts that result in cost savings and have a fixed-cost nature. A typical such case is the definition of a service protocol with a general practitioner. It will be applied to every consumer but the cost incurred (the protocol definition process) is independent of the number of individuals that will be treated. Therefore, it is of interest to assess the optimal transfer structure under the assumption that efficiency-effort costs are independent of the number of enrollees of each insurance company. It will be shown below that a straightforward generalization of the transfer mechanism will adjust for the fixed cost nature of efficiency-effort costs.

We first characterize the constrained optimal allocation, where the relevant constraint is the number of active firms (insurers or providers) in the market. The social welfare function is:

$$\int_Y U(Y - \sum_i C(e_i) - \sum_i G_i \bar{p}_i X(e_i)) h(y) dy \tag{24}$$

The optimal choice of effort for each firm is given by:

$$-C'(e_i) - G_i \bar{p}_i X'(e_i) = 0, \forall i \tag{25}$$

Although the number of firms, denoted by n, is taken as exogenous by the social planner, for the latter to be indifferent to the number of firms the following must be true. Every individual is indifferent, in utility terms, between insurance companies. Given that social welfare was defined as the sum of individual welfare (across the population), if the above condition is not met, then social welfare can increase by an appropriate change of one individual from one firm to another.

We now address whether a social planner would like to change the risk composition of an insurer. Define $q_i \equiv \int_{\chi_i} pg(p)dp = G_i\bar{p}_i$. A social planner does not want to change the risk mix of consumers if (taking two insurers, say firm 1 and firm 2):

$$-X(e_1)dq_1 - X(e_2)dq_2 = 0 \tag{26}$$

This requires $X(e_1) = X(e_2)$ or $e_1 = e_2$. To have equal $e_i, \forall i$, one needs $q_i = q, \forall i$. By definition, $\sum_i q_i = \bar{p}$. From which it follows $q = \bar{p}/n$. Therefore, the optimal effort choice is given by:

$$-C'(e_i) - \bar{p}X'(e_i)/n = 0 \tag{27}$$

We turn now to definition of the transfer system. The problem to be solved is:

$$\left.\begin{array}{c}
\max_{\{\alpha_0,\alpha_1,\alpha_2,F,e\}} \int_Y U(y - T(y) - F)h(y)dy \\
s.a. \int_Y T(y)h(y)dy = \sum_i \int_{p\in\chi_i} S_i\, \tilde{g}(p \mid \chi_i)\, dp \\
S_i = \alpha_0 + \alpha_1\bar{p}_i X(e_i) + \alpha_2\bar{p}X(e_i) \\
\Pi_i(\chi_i)/G_i = \Pi_i(\chi_k)/G_k, \forall i, k \\
\Pi_i(\chi_i) = S_iG_i - C(e_i) + G_iF_i - G_i\bar{p}_iX(e_i) = 0, \forall i \\
-C'(e_i) + G_i(\alpha_1\bar{p}_i - \bar{p}_i + \alpha_2\bar{p})X'(e_i) = 0, \forall i
\end{array}\right\} \texttt{Program C}$$

The no-selection constraint is again satisfied by $\alpha_1 = 1$. The no-selection requires that, for the same number of enrollees, the insurance company does not profit from selecting a different risk mix. The value $\alpha_1 = 1$ makes profits independent of the risk-mix of enrollees, as described previously.

Next, to induce the optimal efficiency effort, one needs to have

$$G_i\alpha_2\bar{p} = -\bar{p}/n \tag{28}$$

From this, results

$$\alpha_2 = -\frac{1}{nG_i} \tag{29}$$

This means that the payment system is firm-specific, and adjusts for the number of enrollees and for the number of operating firms, due to the existence of strong scale economies.

Define the solution to Program C as $\Theta_C = \{(\alpha_0, \alpha_1, \alpha_2, e_i, F_i) : \text{ solves Program C}\}$. The following proposition follows.

Proposition 3 *Take $\theta \in \Theta_C$. Then $\alpha_1 = 1$ and $\alpha_2 = -1/(nG_i)$. That is, the optimal transfer is of the form:*

$$S_i = \alpha_0 + X(e_i)\left(\bar{p}_i - \frac{\bar{p}}{nG_i}\right) \tag{30}$$

This proposition can be more easily interpreted if we take the total transfer to an insurance company:

$$G_iS_i = \alpha_0 G_i + G_i(\bar{p}_i - \bar{p})X(e_i) + \bar{p}\left(G_i - \frac{1}{n}\right)X(e_i) \tag{31}$$

The second term in the right-hand side is similar to the adjustment under constant returns to scale in the efficiency effort with relation to the number of enrollees. It corresponds to the selection component. An insurance with a better risk mix will be a contributor, according to this term.

The last term is a new one, and it results from the fixed cost nature of the effort. To give incentives for taking efficiency effort, under increasing returns to scale, the payment system must also correct for the activity level. The value G_i is the number of enrollees in firm i (or its market share, as total population has been normalized to one), while $1/n$ is the market share if the market is equally divided among the existing n firms. Firms larger than $1/n$ receive a higher transfer, as it pays to provide them with stronger incentives for efficiency.

As before, the payment system can be interpreted as an initial capitation payment, followed by an end-of.the-year compensation fund. It is straightforward to check that this transfer system is balanced, as in equilibrium the sum over all firms of contributions to the ex-post fund cancel out.

6 Multiple health risks

A natural extension of the model is to consider multiple healthrisks. Consider that an individual faces τ different health risks. The expected health expenditures of this individual are:

$$\sum_{\tau} p_\tau x_\tau(e_i) \tag{32}$$

where x_τ is the cost of treating illness τ and p_τ is the associated probability of occurrence. The expected payments of an insurance company with a set χ_i of consumers are given by

$$\int_{\underline{p}\in\chi_i} \sum_{\tau} p_\tau x_\tau(e_i)\tilde{g}(\underline{p} \mid \chi_i)d\underline{p} \tag{33}$$

where $\underline{p}$ denotes the vector of illness probabilities and $\tilde{g}(\underline{p} \mid \chi_i)$ is the joint density of probabilities in set χ_i (defined in a similar way to the single-illness case). We take again the initial assumption of constant returns to scale in the efficiency-effort technology. Accordingly, profits of the insurance company are:

$$\Pi_i = \int_{\underline{p} \in \chi_i} \left(S_i(\underline{p}, \underline{x}) + F_i - C(e_i) - \sum_{\tau} p_\tau x_\tau(e_i) \right) \tilde{g}(\underline{p} \mid \chi_i) d\underline{p} \tag{34}$$

The transfer S_i is defined by

$$S_i = \alpha_0 + \alpha_1 \frac{\int_{\underline{p} \in \chi_i} \sum_\tau p_\tau x_\tau(e_i) \tilde{g}(\underline{p} \mid \chi_i) d\underline{p}}{\int_{\underline{p} \in \chi_i} \tilde{g}(\underline{p} \mid \chi_i) d\underline{p}} + \alpha_2 \int_0^1 g(\underline{p}) \sum_\tau p_\tau x_\tau(e_i) d\underline{p} \tag{35}$$

These expressions are the natural generalizations of the single-illness case. The constraint of "no cream-skimming" is, in a similar way,

$$(\alpha_1 - 1) \left(\int_{\underline{p} \in \chi_i} \sum_\tau p_\tau x_\tau(e_i) \tilde{g}(\underline{p} \mid \chi_i) d\underline{p} - \right.$$
$$\left. - \int_{\underline{p} \in \chi_j} \sum_\tau p_\tau x_\tau(e_i) \tilde{g}(\underline{p} \mid \chi_j) d\underline{p} \right) = 0$$

Like before, this constraint is satisfied for $\alpha_1 = 1$. Turning to the optimal effort level, the insurance company chooses the efficiency level according to the following first-order condition:

$$\alpha_2 \int_0^1 \sum_\tau p_\tau \frac{\partial x_\tau}{\partial e_i} \tilde{g}(\underline{p} \mid \chi_i) d\underline{p} - C'(e_i) = 0 \tag{36}$$

It is relatively straightforward to establish the socially optimal effort level as the solution to

$$-C'(e) - \int_0^1 \sum_\tau p_\tau \frac{\partial x_\tau}{\partial e_i} \tilde{g}(\underline{p} \mid \chi_i) d\underline{p} = 0 \tag{37}$$

Direct comparison of the two first-order conditions shows that $\alpha_2 = -1$ leads again to the optimal effort level from the social point of view. Hence, the results of the previous section are generalized to multiple health risks in a natural way.

7 Concluding remarks

In the analysis of financing systems of the health care sector which rely, at least partially, on private health insurance markets, it is widely recognized that design of health financing structures that use risk adjustment must pay attention to two problems: moral hazard in the choice of efforts to achieve/enhance efficiency and risk selection (cream-skimming).

We present a model that addresses the two issues. It is first shown that, within a simple model of linear transfers on expected health care expenses, for incentives for effort to exist, capitation transfers must be partial ones. This result is, in itself, a particular case of the general result of incentive theory applied to health care financing issues. It shows that the proposed model is able to replicate the conventional wisdom arguments.

The existence of risk selection adds more constraints to the design of the payment system. Nonetheless, we are able to propose an expanded transfer rule that has two desirable features: it does not induce cream-skimming and maintains the incentives for insurance firms to exert cost-reducing effort. The proposal has also the advantage of relying only in information that is usually available and it is financially balanced.[19] The transfer system is defined by: (i) at the beginning of the year, the central fund pays to the insurance company a value α_0 per capita, independent of risk characteristics; the insurance company sets some value to collect directly from consumers. No price discrimation is allowed and open enrollment is enforced; (ii) at the end of the year, insurance companies participate in an ex-post clearing fund, where the contribution (positive or negative) of each company is defined as the difference of the company's risk mix to the population risk mix, times the cost per patient treated in the company. At the equilibrium, as firms are symmetric in our model, average risk and efficiency effort are equal across companies and no adjustment is necessary in expected value. The existence of the ex-post fund is necessary to guarantee the right incentives to firms. Its primary role is to work as a discipline device of deviations to the equilibrium values.

The proposed ex-post fund differs considerably from the "voluntary reinsurance pool" discussed, for example, in Van de Ven et al. (2000).[20] The disadvantage of that pool was clearly identified with reduced incentives for efficiency. We showed this is a feature of the particular form of that pool, not a general characterization of ex-post funds.

Of course, there are some caveats to the model. One of the issues not treated here was prevention effort. If prevention effects can, to some extent, be seen as smaller spending in the event of illness (eventually, even reducing it to zero), then prevention effort is included in a broad interpretation of the effort variable. The implications of considering prevention as a direct change in the risk distribution are left for future research.

Moral hazard on patients' demand is a well-known phenomenon in health care. The arguments to not to provide full insurance are well understood. We opted to let them out of the model. Moreover, other rationing efforts, namely acting on the supply side, and typically associated with managed care, like utilization review and implementation of protocols, can

[19] The interested reader can find an example of computation of this fund and how it compares with the more standard one in a technical appendix available at http://ppbarros.fe.unl.pt/papers.html, or upon request.

[20] "To finance the pool, each insurer pays to the pool, for each of its pooled clients, the average expenditures of all pooled individuals in the pool year." Van de Ven et al. (2000, p. 327).

be included in the effort variable. Similarly, it is widely recognized that physicians possess superior information on patients' characteristics and on patients' needs in each specific illness episode. The efficiency problems raised by this information asymmetry and possible measures taken to mitigate distortions are included, once again, in our broad definition of "effort".

Another aspect no explicitly mentioned was quality of care. In health care administration, quality has a definition closely related to the "efficiency" ideas from economics literature (e.g., reduced number of post-surgery complications). If we see quality in this sense, it is captured by our efficiency effort variable. Quality as a service (or good) that provides higher value to the consumer in the even of illness is a more complex issue, and it is left for future research.

Overall, we believe the approach in this paper to be a more relevant one, from an implementation viewpoint, to the issues at hand (payment systems, risk selection and incentives for efficiency) than the introduction of non-linear payment systems.

References

Banerveld, Erik, René Van Vliet and Wynand Van de Ven, 1996, Mandatory high risk pooling: an approach to reducing incentives for cream-skimming, Inquiry, 33: 133 – 143.

Beck, Konstantin and Peter Zweifel, 1998, Cream-skimming in deregulated social health insurance: evidence from Switzerland, In Peter Zweifel, editor, *Health, the Medical Profession, and Regulation*, Kluwer Academic Publishers, 211-227.

Encinosa, William, 1999, The economic theory of risk adjusting capitation rates, paper presented at the Fourth Biennial Conference on the Industrial Organization of Health Care.

Ellis, Randall, 1998, Creaming, skimping and dumping: provider competition on the intensive and extensive margins, Journal of Health Economics, 17: 537-556.

Ellis, Randall and Thomas McGuire, 1986, Provider behavior under prospective reimbursement, Journal of Health Economics, 5: 129-151.

Glazer, Jacob and Thomas McGuire, forthcoming, Optimal risk adjustment in markets with adverse selection: an application to managed care, American Economic Review

Glazer, Jacob and Thomas McGuire, 1998, Managed care and risk adjustment, pp. 195 – 200. In David Chinitz and Joshua Cohen, editors, Governments and Health Systems – Implications of Differing Involvements, John Wiley & Sons.

Ma, Ching-To Albert, 1994, Health care payment systems: cost and quality incentives, Journal of Economics and Management Strategy, 3(1): 93–112.

Newhouse, Joseph, 1984, Cream-skimming, asymmetric information, and a competitive insurance market, Journal of Health Economics, 3: 97-100.

Newhouse, Joseph, 1994, Patients at risk: health reform and risk adjustment. Health Affairs, 13(1): 132–146.

Newhouse, Joseph, 1996a, Reimbursing health plans and health providers: efficiency in production versus selection, Journal of Economic Literature, 34(1): 1236–1263.

Newhouse, Joseph, 1996b, Medicare, Journal of Economic Perspectives, 10: 159 – 167.

Newhouse, Joseph, Melinda Beewkes Buntin and John Chapman, 1997, Risk adjustment and Medicare: taking a closer look, Health Affairs, 16(5): 26-43.

Newhouse, Joseph, 1998, Risk adjustment: where are we now", Inquiry 35(2): 122-131.

OECD, 1997, OECD Economic Surveys - Portugal, Paris.

Pauly, Mark, 1984, Is cream-skimming a problem for the competitive insurance market?, Journal of Health Economics, 3(1): 87–95.

Rothschild, Michael and Joseph Stiglitz, 1976, Equilibrium in competitive insurance markets: an essay in the economics of imperfect information, Quarterly Journal of Economics, 90: 629-649.

Sappington, David E. and Tracy R. Lewis, 1999, Using subjective risk adjusting to prevent patient dumping in the health care industry, Journal of Economics and Management Strategy, 8(3): 351 – 382.

Selden, Thomas, 1990, A model of capitation, Journal of Health Economics, 9: 397-409.

Shmueli, Amir, Nira Shamai, Yoram Levi and Miri Abraham, 1998, In search of a natural capitation formula: the Israeli experience, pp. 201 – 208. In David Chinitz and

Joshua Cohen, editors, Governments and Health Systems – Implications of Differing Involvements, John Wiley & Sons.

Smith, Peter, 1998, The estimation, application, and experience of the new capitation formula in the United Kingdom, pp. 209 – 224. In David Chinitz and Joshua Cohen, editors, Governments and Health Systems – Implications of Differing Involvements, John Wiley & Sons.

Van de Ven, Wynand and Randall Ellis, 2000, Risk adjustment in competitive health plans, chapter 17 in A.J. Culyer and J.P. Newhouse, editors, Handbook of Health Economics, North-Holland, Elsevier: Amsterdam.

Van de Ven, Wynand and René Van Vliet, 1992, How can we prevent cream skimming in a competitive health insurance market? the great challenge for the 1990s, In Peter Zweifel and H.E. Frech III (eds.), Health Economics Worlwide, Kluwer Academic Press, pp.23–46.

Van de Ven, Wynand and René Van Vliet, 1994, Risk-adjusted capitation: recent experiences with the Netherlands, *Health Affairs*, 120–135.

Van de Ven, Wynand, René Van Vliet, Eric Van Banerveld and Leida Lamers, 1998, Towards a risk-adjustment mechanism in a competitive health insurance market, pp. 179–184. In David Chinitz and Joshua Cohen, editors, Governments and Health Systems – Implications of Differing Involvements, John Wiley & Sons.

Van de Ven, Wynand, René Van Vliet, Franz Schut and Erik Van Banerveld, 2000, Access to coverage for high-risks in a competitive individual health insurance market: via premium rate restrictions or risk-adjusted premium subsidies, Journal of Health Economics, 19: 311 – 339.

Van Vliet, René, 1992, Predictability of individual health care expenditures, *Journal of Risk and Insurance*, **54**(3): 443–460.

Van Vliet, René and Wynand Van de Ven, 1992, Towards a capitation formula for competing health insurers. An empirical analysis, *Social Science and Medicine*, **34**(9): 1035 – 1048.

Appendix

A formal solution to the choice of the transfer system

Formally, the social planner's problem can be written as:

$$\begin{aligned}
\max_{\{\alpha_0,\alpha_1,\alpha_2,F,e\}} \quad & \int_Y U(y - T(y) - F)h(y)dy \\
s.a. \quad & \int_Y T(y)h(y)dy = \sum_i \int_{p\in\chi_i} S_i\, \tilde{g}(p \mid \chi_i)\, dp \\
& S_i = \alpha_0 + \alpha_1 \bar{p}_i X(e_i) + \alpha_2 \bar{p} X(e_i) \\
& \Pi_i(\chi_i)/G_i = \Pi_i(\chi_k)/G_k, \forall i, k \\
& \Pi_i(\chi_i) = 0, \forall i \\
& -C'(e_i) + (\alpha_1 \bar{p}_i - \bar{p}_i + \alpha_2 \bar{p}) X'(e_i) = 0, \forall i
\end{aligned}$$

The first constraint establishes that the funding system must be balanced (a more generous system of transfers requires heavier contributions). The second expression is nothing more than the definition of the allowable transfers.

The third expression reproduces the set of conditions to be satisfied in order to eliminate incentives to cream-skimming (the profit of the insurance company is independent of the set of consumers selected by the insurer).

The fourth condition is a free-entry condition for a competitive health insurance market. Finally, the last constraint is the first-order condition that characterizes the optimal choice of effort. We assume that the regularity conditions for this approach to be valid hold.

The "no cream-skimming" condition can be re-written as:

$$\int_{p\in\chi_i} (S_i - pX(e_i))\ \tilde{g}(p \mid \chi_i)\, dp = \int_{p\in\chi_k} (S_i - pX(e_i))\ \tilde{g}(p \mid \chi_k)\, dp \tag{38}$$

or

$$\int_{p\in\chi_i} (\alpha_1 - 1)\, pX(e_i)\, \tilde{g}(p \mid \chi_i)\, dp = \int_{p\in\chi_k} (\alpha_1 - 1) pX(e_i)\, \tilde{g}(p \mid \chi_k)\, dp \tag{39}$$

It is straightforward to see that this condition holds for any two sets of consumers only if $\alpha_1 = 1$. This result can thus be substituted into the original program.

Take now the zero profit condition in a competitive health insurance market. Solving it in order to F yields:

$$F_i = C(E_i) + \bar{p}_i X(e_i) - S_i \tag{40}$$

The first constraint can also be written as :

$$\bar{T} = \int_Y T(y)h(y)dy = \alpha_0 + \alpha_1 \sum_i \bar{p}_i X(e_i) G_i + \alpha_2 \sum_i \bar{p} X(e_i) G_i \tag{41}$$

From lemma 1, $e_i = e$, and it is possible to write, using $\alpha_1 = 1$,

$$\bar{T} = \alpha_0 + (1 + \alpha_2)\bar{p}X(e) \tag{42}$$

From the imposition of open enrollment rules and from the existence of competition among insurance firms on the insurance premium directly charged to consumers, it must be the case that, in equilibrium, all firms will charge the same, $F_i = F, \forall i$. Moreover, as $G_i > 0$ and $\sum_i G_i = 1$ by definition, it follows that $F = \sum_i F_i G_i$. This characteristic of the model allows us to express social welfare in terms of a representative consumer, for each income level. Hence, the social welfare function can be written as:

$$W = \int_Y U(y - T(y) + \int_Y T(y)h(y)dy - \sum_i G_i C(e_i) - \sum_i G_i \bar{p}_i X(e_i))h(y)dy \tag{43}$$

Substitution for α_0 in the problem to be solved simplifies it to:

$$\begin{aligned} \max_{\alpha_2} \quad & \int_Y U\left(y - T(y) - C(e) + \bar{T} - \bar{p}X(e)\right) h(y)dy \\ s.t. \quad & -C'(e) + \alpha_2 \bar{p} X'(e) = 0 \end{aligned}$$

From this, it is clear that α_2 can be freely set to induce the optimal choice of effort level, which happens for $\alpha_2 = -1$ (in which case, the Lagrange multiplier associated with the incentive constraint has a zero value).

Therefore, a net transfer

$$S_i = \alpha_0 + X(e_i)\left(\bar{p}_i - \bar{p}\right) \tag{44}$$

allows the system to achieve the optimal level of cost-reducing effort, without creating incentives to risk selection.

The value α_0 is determined in order to ensure, together with F, that the participation constraint of the insurance firm is satisfied:

$$F + \alpha_0 = C(e) + \bar{p}_i X(e) \tag{45}$$

The proposed ex-post fund is financially balanced

We show now the claim made in the text that the ex-post adjustment process proposed is financially balanced.

The contribution of each firm to the fund is

$$A_i = (\bar{p}_i - \bar{p})\, X(e_i) G_i \tag{46}$$

Summing over all firms,

$$\sum_i A_i = \sum_i (\bar{p}_i - \bar{p})\, X(e_i) G_i \tag{47}$$

As $e_i = e, \forall i$,

$$\sum_i A_i = X(e) \left(\sum_i \left(\int_{i \in \chi_i} p g(p)\, dp \right) - \bar{p} \right) = 0 \tag{48}$$

The last equality sign follows from the fact that the set of consumers satisfies

$$\chi_i \cap \chi_j = \emptyset \text{ and } \cup_i \chi_i = 1. \tag{49}$$