The design of insurance coverage for medical products under imperfect competition[1]

David Bardey[2], Helmuth Cremer[3]
and Jean-Marie Lozachmeur[4]

January 2015

[1] We thank Philippe Choné for his very helpful and constructive comments. The authors acknowledge support by the Health Chair—a joint initiative by PSL, Université Paris-Dauphine, ENSAE and MGEN under the aegis of the Fondation du Risque (FDR).

[2] University of Los Andes (Bogota, Colombia) - Department of Economics (CEDE) and Toulouse School of Economics (France).

[3] Toulouse School of Economics (IDEI and GREMAQ).

[4] Toulouse School of Economics (IDEI and GREMAQ-CNRS).

Abstract

This paper studies the design of health insurance with *ex post* moral hazard, when there is imperfect competition in the market for the medical product. Various *scenarios,* such as monopoly pricing, price negotiation or horizontal differentiation are considered. The insurance contract specifies two types of copayments: an *ad valorem* coinsurance rate and a specific (per unit) copayment. By combining both copayment rates in an adequate way the insurer can effectively control the producer price, which is then set so that the producer's revenue just covers fixed costs. Consequently, a suitable regulation of the copayment instruments leads to the *same* reimbursement rule of individual expenditures as under perfect competition for medical products. Additional rationing of coverage because of imperfect competition as advocated by Feldstein (1973) is thus *not* necessary. Interestingly the optimal policy closely resembles a *reference price mechanism* in which copayment rates are low (possibly negative) and coinsurance rates are high.

JEL Codes: I11, I13, I18.

Keywords: *ex post* moral hazard, health insurance contracts, copayments, imperfect competition.

1 Introduction

According to conventional wisdom, the level of health insurance provided by a competitive insurance industry is "too generous" when there is imperfect competition in the market for medical products. This property has been first documented in the classical paper by Feldstein (1973), and more recently by Feldman and Dowd (1991). A direct implication of this property is that the regulator should restrict insurance coverage[1] or/and encourage price competition.[2] This paper challenges the first view, namely that insurance coverage should be lower under imperfect than under perfect competition for medical products. Instead, we show that when there is imperfect competition, a suitable regulation of copayment instruments would eventually lead to the *same* reimbursement rule of individual expenditures as under perfect competition for medical products. In other words, while full insurance is typically not desirable, imperfect competition in the market for medical products does *not* require an *extra* reduction in insurance coverage, as long as the reimbursement scheme is properly designed.

The need to restrict health insurance coverage arises because of the so-called *ex post* moral hazard in health care systems. This phenomenon refers to the increase of health expenditures, which is due to health insurance coverage; Arrow (1963) and Pauly (1968). Expenditures increase for two reasons. First, health insurance typically reduces the share of health expenditures that is paid by policyholders. This tends to increase total expenditures, as long as patients' demand is not totally price inelastic. Second, health care providers, which evolve in an imperfect competition setting, respond to this higher demand by increasing prices. To mitigate this moral hazard problem, insurance coverage is typically less than 100%. Copayments for medical products take various forms in developed countries. On top of (possibly capped) deductible, patients may face a copayment rate (a fixed amount per unit of the medical good consumed)[3] and/or a traditional coinsurance rate (a proportion of the medical bill). Very often, public or private insurance plans combine these two kinds of copayments.

[1] See Varithianathan (2006), for a formal analysis of this issue.
[2] See, *e.g.,* Gaynor *et al.* (2000).
[3] Exemples of copayments are payment per drug prescription, per pills, per doctor's visit, per hospital's admission *etc...*

We characterize the "optimal" reimbursement of medical goods under *ex post* moral hazard, and imperfect competition in the providers' market. The novelty of this paper is that the insurer may use simultaneously two types of instruments that are widely used in practice: a linear *ad valorem* copayment rate (also called coinsurance rate), which is a share of expenditure for the health care product, plus a *specific* (per unit) copayment rate which is proportional to the quantity of the health care product consumed. Doing this, we build on the framework developed by Zeckhauser (1970) and later by Besley (1988), but we relax the assumption that the price of medical services is exogenously fixed. Instead, we study health insurance design when the price of medical products is determined in an imperfectly competitive market. We consider various *scenarios,* such as monopoly pricing, price negotiation or horizontal differentiation with free entry.

Our model consists of a three stage game. In a first stage, the insurer chooses the levels of the premium and the two out-of-pockets rates. In a second stage, the price of medical good is determined. Finally, in a third stage, the state of nature is revealed and patients buy the quantity of medical care products to maximize their utility given the state of their health and (net) prices.

We show that the insurer can always extract the full surplus of the monopolist by using an appropriate combination of coinsurance and copayment. In other words, under this reimbursement policy, the producer's net revenue is just sufficient to cover the fixed cost of the medical product. This remains true *mutatis mutandis* in the different competitive *scenarios* we consider. In either case, the insurer can restore a second-best reimbursement rule corresponding to the one used under perfect competition. This combination involves the lowest possible level of copayment rate (possibly negative) associated with the highest possible level of coinsurance rate such that the participation constraint of health providers is binding.[4] As a result, the optimal policy resembles very much the reference price mechanism in which copayment rates are very low and effectively negative, while coinsurance rates are high. To be more precise, the idea of reference pricing is that patients are required to pay the difference between the actual

[4] In the Bertrand competition case with free entry, the level of this profit is endogenous to the reimbursement system (see Subsection 6.3.2).

price received by the supplier and the insurer's contribution limit (the reference price). As such, in its extreme form, reference pricing can be seen as a combination of a 100% coinsurance rate combined with a negative copayment rate (equal to minus the reference price).

This result and the conditions on which it relies are particularly interesting as it sheds light on some of the issues that are at the heart of the American health care debate. The Obama administration has recently given the go-ahead for insurers and employers to use reference prices as cost-control strategy in the health sector.[5] Our results provide support for this policy. However, for our policy to be fully effective, both instruments have to be regulated; controlling the *average* reimbursement rate as envisioned by the Affordable Care Act is *not* enough. Instead the type of contract and the form of the copayments would have to be regulated. Such a system would resemble that of so called "Social Health Insurance" countries in Europe like Germany, The Netherlands or Switzerland where insurance companies offer standardized contracts but compete in premiums. Our results imply that a suitable regulation of the forms of copayment would ultimately solve the problem of excessive insurance without altering the level of individuals protection.

In addition to the literature already mentioned, our paper is related to a recent article by Lakdawalla and Sood (2013). These authors also show that a second-best insurance scheme can be restored when the medical product is sold by a monopolist. However, their result holds only when the monopolist can choose *both* quantity and price. In this case, everything works exactly as when the monopolist can charge a two-part tariff which leads to a market price equal to marginal cost, while surplus is extracted through the fixed part. In our framework, there is no need that the monopolist chooses both quantity and prices; put differently, the policy we describe is effective under linear pricing; no two-part tariffs are needed.

Finally our paper is also related to the literature on tax incidence under imperfect

[5]Since the success of CalPERS (California Public Employees' Retirement System) in implementing reference pricing for hip and knee replacement (see Robinson and Brown, 2013), the Mercer benefits consulting firm reported that 12% of the largest employers were using reference pricing in 2013, nearly twice the 7% who relied on it in 2012.

competition which has been developed in public finance.[6] In particular this literature compares the respective properties of *ad valorem* taxes and specific taxes and shows that the former is in general welfare superior. To be more precise, the *ad valorem* tax yields more tax revenue for any given output level. Intuitively this result arises because the *ad valorem* tax is effectively equivalent to a specific tax supplemented by a profit tax, the latter raising revenue without affecting quantity. Most closely related to our analysis is probably the paper by Delipalla and Keen (1992). They establish the welfare dominance of *ad valorem* taxes in a Ramsey framework where the government has to finance a given level of public expenditure and has to rely only on consumption taxes. However, they restrict all taxes to be positive so that producer prices cannot be fully controlled. Recall that our result typically requires a negative copayment rate. In the health sector this is a realistic option as the reference pricing example shows, while otherwise taxes *are* most often restricted to be positive.

In the next section we present the model. Section 3 studies the benchmark case with perfect competition in the market for medical products. Next, Section 4 examines the case of a monopoly. Section 5 analyzes the optimal insurance scheme when the insurer and the monopoly bargain over the producer price. Finally, Section 6 studies the optimal contract when different variants of the medical product are sold in a horizontally differentiated oligopoly. We use a setting based on the discrete choice approach, without and with free entry.

2 The Model

A mass 1 of agents are endowed with an identical income w. They each face a risk distributed over $\Theta \equiv \left[\underline{\theta}, \overline{\theta}\right]$ to become of type θ. The density and distribution functions of θ are respectively given by $g(\theta)$ and $G(\theta)$. Individuals maximize their expected utility. Their preferences are represented by a Von-Neumann-Morgenstern utility function $u_\theta \equiv u(x, q, \theta)$, where x denotes the consumption of a numeraire composite good, while q represents the consumption of a medical product (for instance a prescription drug) the (producer) price of which is denoted by P. The function u is strictly increasing and

[6] See Myles (1995), ch.11 for a detailed survey.

strictly concave in its x and q.

The consumption of q may be covered by a health insurance scheme, which can be seen as a partial subsidy to P. The copayment can take two forms. The first type of copayment is "*ad valorem*" in the sense that the out-of-pocket amount paid by the individual is a share of total expenses on the medical product, namely Pq. Let t denote the "*ad valorem*" copayment *rate*, which in the health insurance literature is often referred to as "coinsurance" rate. The second type of copayment is "specific" so that the out-of-pocket payment is simply proportional to the quantity q; this corresponds to what the health economics literature refers to as "copayment" rate. Let c denotes this copayment rate. As a result, the net (or consumer) price of the medical product is given by

$$\tilde{P} = c + tP. \tag{1}$$

Observe that we do *not* restrict c to be positive nor t to be less than one. As an example, consider the case in which $c < 0$ and $t = 1$. This special case represents the reference price mechanism in which patients are entitled to pay the difference between the producer price P and the reference price (here equal to $-c$). To buy this insurance protection, policyholders pay a premium π. In state θ, the consumption of the composite good is thus given by

$$x_\theta = w - \pi - \tilde{P}q_\theta, \tag{2}$$

where q_θ denotes the consumption of the medical product in state θ, and the expected utility can be rewritten as

$$E_\theta u_\theta = \int_\theta u(w - \pi - \tilde{P}q_\theta, q_\theta, \theta)dG(\theta),$$

where E_θ denotes the expectation operator over Θ.

2.1 Timing

Decisions are made in three stages according to the following timing. First, the insurer chooses the triplet (π, t, c). Second, the gross price P (and thus the net price $\tilde{P}$) is determined by the market of the medical product. Finally, the state θ is realized and each individual of type θ chooses q_θ. We consider different scenarios which differ mainly in the underlying specification of stage 2, that is how the market for the medical product is represented. These include perfect competition, monopoly, differentiated oligopoly as well as a bargaining procedure over the price.[7] In either case, we determine the "subgame perfect" equilibrium and solve the game by backward induction.[8]

The optimal health insurance contract is determined by the following problem

$$\max_{\pi,c,t} \quad E_\theta u_\theta = \int_\theta u(w - \pi - \tilde{P}q_\theta, q_\theta, \theta) dG(\theta), \tag{3}$$

$$\text{s.t.} \quad B = \pi - \left(P - \tilde{P}\right) E_\theta q_\theta \geq 0, \tag{4}$$

where the constraint (4) states that B, which is equal to the premium minus expected coverage of expenses on the medical product, must be positive or equal to 0. In a subgame perfect solution, the insurance provider anticipates the equilibrium induced by its strategy in the subsequent stage: the levels of q_θ in stage 3 and the producer price P in stage 2.

2.2 The individual problem

As policyholders are price takers, the solution to stage 3 does not depend on the scenario adopted for the market of the medical product, i.e. the determination of its price. Once the state of nature θ is revealed policyholders choose their consumption of q that solves

$$\max_q u(w - \pi - \tilde{P}q, q, \theta),$$

taking $\tilde{P}$ and π as given. Assuming an interior solution yields the following first-order condition for all $\theta \in \Theta$

$$\tilde{P}\frac{\partial u(\tilde{w} - \tilde{P}q, q, \theta)}{\partial x} = \frac{\partial u(\tilde{w} - \tilde{P}q, q, \theta)}{\partial q}, \tag{5}$$

where $\tilde{w} = w - \pi$ represents the income net of the premium paid. Equation (5) simply defines the standard Marshallian demand $q_\theta^*\left(\tilde{P}, \tilde{w}\right)$, as a function of the (consumer) price and the income. Aggregate (expected) demand is denoted by $Q^*\left(\tilde{P}, \tilde{w}\right) = E_\theta q_\theta^*\left(\tilde{P}, \tilde{w}\right)$.

[7] In the differentiated oligopoly case, stage 3 is also modified to account for consumers preferences over the differentiated medical products.

[8] Even though stage 2 is not always explicitely as a "proper" game (for instance in the case of perfect competition).

3 Optimal insurance under perfect competition

Assume first that there is perfect competition in the market of the medical product, with firms having a constant marginal cost equal to k. The producer price P of the medical product is then equal to k and the problem of the insurance provider amounts to choosing $\left(\pi, \tilde{P}\right)$ that solves

$$\max_{\pi,\tilde{P}} \quad E_\theta u_\theta = \int_\theta u\left[w - \pi - \tilde{P}q^*_\theta\left(\tilde{P},\tilde{w}\right), q^*_\theta\left(\tilde{P},\tilde{w}\right), \theta\right] dG(\theta). \tag{6}$$

$$\text{s.t.} \quad B = \pi - \left(k - \tilde{P}\right) Q^*\left(\tilde{P},\tilde{w}\right) dG\left(\theta\right) \geq 0 \tag{7}$$

Since the producer price is given, it immediately follows that coinsurance or copayment rates are equivalent. Only $\tilde{P} = c + tk$ matters and all combination of c and t that bring about a given level of $\tilde{P}$ yields exactly the same outcome. This does not come as a surprise! It is perfectly in line with the standard result, obtained in tax incidence, that *ad valorem* and specific taxes are equivalent under perfect competition. And as far as their impact on consumers are concerned, copayment rates are exactly similar to taxes.[9]

Appendix A shows that the optimal reimbursement rate is determined by

$$\frac{\left(P - \tilde{P}\right)}{\tilde{P}} = \frac{\left(k - \tilde{P}\right)}{\tilde{P}} = -\frac{\operatorname{cov}\left(q^*_\theta, \Lambda_\theta\right)}{E_\theta q^*_\theta S^*_\theta}, \tag{8}$$

where S^*_θ is the price elasticity of the compensated (Hicksian) demand for q while Λ_θ is defined by

$$\Lambda_\theta = \frac{\frac{\partial u_\theta}{\partial x}}{\lambda} - \left(\tilde{P} - k\right)\frac{\partial q^*_\theta}{\partial \tilde{w}} = \frac{1}{\lambda}\left[\frac{\partial u_\theta}{\partial x} - \lambda\left(\tilde{P} - k\right)\frac{\partial q^*_\theta}{\partial \tilde{w}}\right], \tag{9}$$

where λ is the multiplier associated with the insurance provider's revenue constraint (7). The terms Λ_θ is the counterpart to the "net social marginal utility of income" used in taxation models; see Diamond (1975). The expression in brackets in the last term of (9) represents the expected marginal utility gain of a marginal *decrease* in the premium.[10] The term $\operatorname{cov}(q^*_\theta, \Lambda_\theta)$ measures the benefits of providing insurance. It is positive when policyholders with a higher θ, who consume a higher level of q have a larger marginal utility of income. This is necessarily true when the illness causes a monetary cost as is

[9] Similarly, the net revenue of the insurer, B, plays the same role as tax revenue in taxation models.

[10] Dividing by λ, the shadow cost of the insurer's net revenue, converts utility into monetary terms.

often assumed in the health economics literature. With a more general representation of the impact of illness on utility, the covariance term is in principle ambiguous. However, we shall concentrate on the case where it is positive because otherwise health insurance is not desirable. This rule expresses the optimal reimbursement rate (LHS) as being a trade-off between risk sharing (the numerator in the RHS) and efficiency loss due to moral hazard *ex post* (the denominator in the RHS).[11]

In the remainder of the paper, we consider alternative models of the market for the medical product and determine the optimal policy of the insurance provider in each case. Section 4 considers the case of an (unregulated) monopoly. In Section 5 we continue to assume that there is a single producer but that the price is determined through a Nash bargaining process between the regulator and the seller. Finally Section 6 considers a model of Bertrand competition with N profit maximizing producers selling horizontally differentiated products.

4 Profit maximizing monopoly

4.1 Price formation

Assume that the medical product is supplied by a profit maximizing monopoly. Its profit function is given by

$$\Pi = (P - k)\,Q - F,$$

where k and F respectively denote the marginal and the fixed cost, which includes all the sunk costs incurred during the development of the medical product. Aggregate (expected) output Q is given by $Q = Q^*\left(\tilde{P}, \tilde{w}\right)$. Rewriting the profit in terms of the consumer price $\tilde{P}$ defined by (1) yields

$$\Pi = \frac{1}{t}\left(\tilde{P} - \tilde{k}\right) Q^*\left(\tilde{P}, \tilde{w}\right) - F, \tag{10}$$

[11] Since copayment rates are formally equivalent to taxes, it is not surprising that this rule is very similar to the many-person Ramsey rule derived by Diamond (1975). In taxation models, the trade-off is between redistributional benefits and deadweight loss. In our insurance model, "redistribution" occurs between states of nature and is referred to as risk sharing, while the efficiency loss, measured by the impact of *ex-post* moral hazard, is essentially equivalent to the expected deadweight loss of reimbursement rates.

where $\tilde{k} = c + tk$. Recall that at this stage c, t as well as π, and therefore $\tilde{w} = w - \pi$ are given. Expression (10) shows that the problem of the monopoly amounts to choosing the net price $\tilde{P}$. The monopoly's problem is then to maximize its profits given by equation (10) with respect to $\tilde{P}$. This yields the following pricing rule

$$\frac{\tilde{P}^M - \tilde{k}}{\tilde{P}^M} = \frac{1}{|\varepsilon|}, \tag{11}$$

where $|\varepsilon| = \left(\tilde{P}/Q\right)\left(\partial Q/\partial \tilde{P}\right) > 1$ denotes the (marshallian) price elasticity of Q with respect to $\tilde{P}$. Observe that while equation (11) provides a closed form solution only when ε is constant (which we do *not* assume), it is valid for any demand function.[12] For future reference, notice that the RHS of this expression does not depend on c nor t; the two copayment parameters only appear on the LHS, through $\tilde{k} = c + tk$. Consequently, we can solve (11) with respect to $\widetilde{k}$ for a given level of $\tilde{P}^M$ which yields

$$\tilde{k} = \tilde{P}^M \left(1 - \frac{1}{|\varepsilon|}\right). \tag{12}$$

This expression shows that the insurance provider can effectively bring about any level of $\tilde{P}^M$ by setting $\tilde{k}$ at the appropriate level.

Let $\Pi^* (t, c)$ denote the profit the monopolist achieves in the equilibrium of the health care market. The monopolist is entering the market only as long as its "participation constraint" is satisfied

$$\Pi^* (t, c) \geq 0, \tag{13}$$

so that its revenues are sufficient to cover *all* its cost, including the fixed cost. Profit is zero when the total mark-up (over marginal cost) is just sufficient to cover the fixed cost F.

Before studying the general problem we study the optimal insurance scheme when only one of the instruments is available. In these two subsections (and only there), we shall assume for simplicity that demand elasticity, ε, is constant. Interestingly enough, these solutions correspond to the level of insurance protection when the regulator wants to ration the provision of insurance as advocated by Feldstein (1973). As will be clear later, this rationing solution is far from being optimal.

[12] As long as there is an interior solution.

4.2 The optimal coinsurance reimbursement rule

Assume first that the insurance provider can only use the coinsurance rate t. In other words, a fraction $(1-t)$ of the cost of the medical product is reimbursed so that $\widetilde{P} = (1-t)P$. Setting $c = 0$, its problem is now:

$$\max_{\pi,t} \quad E_\theta u_\theta = \int_\theta u\left[w - \pi - \tilde{P}^M q_\theta^*\left(\tilde{P}^M, \tilde{w}\right), q_\theta^*\left(\tilde{P}^M, \tilde{w}\right), \theta\right] dG(\theta), \tag{14}$$

$$\text{s.t.} \quad B = \pi - \left(P^M - \tilde{P}^M\right) Q^*\left(\tilde{P}^M, \tilde{w}\right) dG(\theta) \geq 0. \tag{15}$$

where with $\tilde{k} = tk$, the monopoly pricing rule (11) implies

$$\tilde{P}^M = tk/\left(1 - 1/|\varepsilon|\right) \quad \text{and} \quad P^M = k/\left(1 - 1/|\varepsilon|\right). \tag{16}$$

When ε is constant, the producer price P^M is not affected by the insurance scheme. The problem of the insurance provider is thus the same as under perfect competition except that P, which was equal to k under perfect competition has to be replaced by P^M as defined by (16). From equation (8) the optimal reimbursement rate is then given by

$$\left.\frac{P^M - \tilde{P}^M}{\tilde{P}^M}\right|_{c=0} = -\frac{\operatorname{cov}\left(q_\theta^*, \Lambda^\theta\right)}{E_\theta q_\theta^* S_\theta^*} > 0. \tag{17}$$

In words, when the covariance term is positive so that the provision of health insurance is desirable, we get a positive (but lower than 1) coinsurance rate t, the level of which is determined by the trade-off between risk sharing and *ex-post* moral hazard described in Section 3.

4.3 The optimal specific reimbursement rule

Assume now that the insurance provider can only use a specific reimbursement rate. To provide health insurance, the copayment rate c must then be negative. Setting $t = 1$, the insurance provider's problem is now

$$\max_{\pi,c} \quad E_\theta u_\theta = \int_\theta u\left[w - \pi - \tilde{P}^M q_\theta^*\left(\tilde{P}^M, \tilde{w}\right), q_\theta^*\left(\tilde{P}^M, \tilde{w}\right), \theta\right] dG(\theta) \tag{18}$$

$$\text{s.t.} \quad B = \pi - \left(P^M - \tilde{P}^M\right) Q^*\left(\tilde{P}^M, \tilde{w}\right) dG(\theta) \geq 0, \tag{19}$$

where with $\tilde{k} = c$, the monopoly pricing rule (11) implies

$$\tilde{P}^M = (c+k) / (1 - 1/|\varepsilon|) \quad \text{and} \quad P^M = (c/|\varepsilon| + k) / (1 - 1/|\varepsilon|). \tag{20}$$

Unlike with a coinsurance rate, the monopoly price now depends positively on the copayment rate level, c. As a result, a higher specific *reimbursement,* that is a lower level of c, decreases the monopoly price.

Assuming again that $|\varepsilon|$ is a constant, we show in Appendix B that the optimal reimbursement rate is now given by

$$\left.\frac{P^M - \tilde{P}^M}{\tilde{P}^M}\right|_{t=1} = -\frac{\frac{1}{|\varepsilon|} Q\left(\tilde{P}^M, \tilde{w}\right) + \operatorname{cov}\left(q^*_\theta, \Lambda^\theta\right)}{E_\theta q^*_\theta S^*_\theta} > 0. \tag{21}$$

Compared to the coinsurance and perfect competition case, we now have an extra term in the numerator, namely $Q\left(\tilde{P}^M, \tilde{w}\right) / |\varepsilon|$ which is positive so that it tends to increase the reimbursement, *i.e.*, decrease c. Recall that a decrease in c reduces the producer price, which represents a second positive effect, on top of the provision of insurance. The price effect is stronger the less elastic is demand for medical products, which explains why $|\varepsilon|$ appears in the denominator of the extra term.[13]

4.4 The optimal health insurance contract with multiple instruments

We now return to the case where the insurance provider can use both instruments. Recall that we do not restrict c to be positive nor t to be less than one. Before determining the optimal level of the instruments we shall examine if and how the two instruments have to be combined.

To do this we start from an arbitrary reimbursement policy (t, c) and consider a variation $dc = -k dt$, with $dt > 0$. Since $\tilde{k} = c + tk$, this variation leaves $\tilde{k}$ unaffected. By equation (11), $\tilde{P}^M$ thus $Q\left(\tilde{P}^M, \tilde{w}\right)$ are not affected either, so that the indirect utility remains the same. Rewriting (11) in terms of the producer price P^M yields

$$\left(c + tP^M\right)\left(1 - \frac{1}{|\varepsilon|}\right) = \tilde{k},$$

[13] Differentiating (20) yields

$$\frac{\partial P^M}{\partial c} = \frac{1}{|\varepsilon| - 1} > 0.$$

where the demand elasticity is constant as long as the consumer price $\tilde{P}^M$ is given. The variation $dc = -kdt$ thus implies

$$\left.\frac{dP^M}{dt}\right|_{\tilde{k}=\text{constant}} = -\frac{P-k}{t} < 0.$$

Consequently, the considered variation lowers the producer price which, by the budget constraint in (4) decreases the insurance provider's expenses without affecting its objective function. In words, by reducing c and increasing t, according to $dc = -kdt$, the insurance provider can reduce the producer price and profit while leaving the consumer price unaffected. This in turn immediately implies that the optimal policy (t^*, c^*) must be such that the monopolist's participation constraint is binding *i.e.*, $\Pi^*(t^*, c^*) = 0$.

Consequently, the budget constraint of the insurance provider

$$B = \pi - \left(\tilde{P}^M - P^M\right) Q^*(\tilde{P}^M, \tilde{w}) = \pi - \left(\tilde{P}^M - k\right) Q^*(\tilde{P}^M, \tilde{w}) + \left(k - P^M\right) Q^*(\tilde{P}^M, \tilde{w}) \geq 0$$

can be written as $\pi + \left(\tilde{P}^M - k\right) Q^*(\tilde{P}^M, \tilde{w}) - F$, which is the same as under perfect competition except that we have to account for the fixed cost of the monopolist.

Therefore the problem of the insurance provider can be decomposed into two stages. First, we can solve the insurance provider's problem with $(\pi, \widetilde{P})$ as decision variables exactly as under perfect competition, except that F has to be included in the budget constraint. Recall that since the RHS of equation (11) and thus also of (12) is independent of $\tilde{k}$, the insurance provider can effectively "choose" the level of $\tilde{P}^M$. This stage of the insurance provider's problem is then given by

$$\max_{\pi, \tilde{P}} \quad E_\theta u_\theta = \int_\theta u\left[w_0 - \pi - \tilde{P}^M q^*_\theta(\tilde{P}^M, \tilde{w}), q^*_\theta(\tilde{P}^M, \tilde{w}), \theta\right] dG(\theta) \tag{22}$$

$$\text{s.t.} \quad \pi + \left(\tilde{P}^M - k\right) Q^*(\tilde{P}^M, \tilde{w}) - F \geq 0. \tag{23}$$

Since this is the same problem as under perfect competition (except that we require $B \geq F$ rather than $B \geq 0$), the optimum is described by equation (8) together with the budget constraint (23). The solution to this problem yields the optimal consumer price $\tilde{P}^{M*}$ and from equation (12) the optimal level of $\widetilde{k}^* = c + tk$.

To obtain the full solution to the insurance provider's problem it is then sufficient to set the levels of c^* and t^* such that

$$c^* + t^* k = \widetilde{k}^* \tag{24}$$

and

$$\Pi^*(t^*, c^*) = 0.$$

Figure 1 illustrates this reasoning. The dashed line with slope $-k$ represents equation (24); and shows the combination of c and t that yields $\widetilde{k}^*$. The downward sloping curves represent the iso-profit curves. Their slope is given by

$$-\frac{\frac{\partial \Pi^*(t^*,c^*)}{\partial t}}{\frac{\partial \Pi^*(t^*,c^*)}{\partial c}} = -\left(k + \left(\tilde{P}^{M*} - k\right)/t\right), \tag{25}$$

where we have differentiated equation (10) by making use of the envelope theorem. Observe that the iso-profit curves are steeper than the line representing equation (24); their slope exceeds k (in absolute value). The bold curve corresponds to a profit level of zero. On the south west of the bold curve, profits are negative while they are positive in the north east. The solution is given by the intersection of this iso-profit curve and the straight line. Intuitively one can think of this as moving along the line towards the point where profits are exactly zero.

5 Regulated monopoly: price bargaining

Assume now that when a new medical product is accepted for introduction in the market, its price is set by a bargaining process between the regulator and the producer. This process is represented by a Generalized Nash Bargaining Solution, with respective bargaining weights of γ for the regulator and of $(1-\gamma)$ for the producer. It represents the second stage of our three stage game. When the bargaining takes place, t and c are already set. Recall that in stage 1 the insurance provider anticipates the equilibrium of the subsequent stages induced by its strategy. The outside option (or threat point) corresponds to the situation where the medical product is not sold at all. The regulator objective is to maximize the consumer's expected utility while the firm maximizes profits. The outside option yields a given level of utility EU_0 to the consumer which corresponds to the expected utility when $q_\theta = 0$ for all θ, while the producer's profit is equal to $-F$. The difference between the producer's profit and its threat point level is

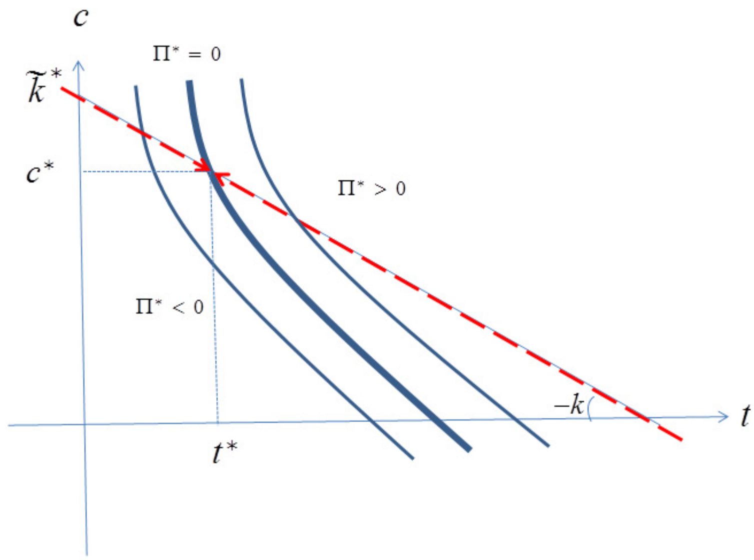

Figure 1: The optimal combination of co-insurance and co-payment.

thus equal to:

$$(P-k)\,Q\,(P) = [(P-k)\,Q\,(P) - F] - (-F).$$

The solution to the bargaining process is then given by

$$P = \arg\max_{P} \left(E_\theta u_\theta\,(P) - EU_0\right)^\gamma \left((P-k)\,Q\,(P)\right)^{1-\gamma}.$$

This problem is equivalent to determining $\tilde{P}$ such that

$$\tilde{P} = \arg\max_{\tilde{P}} \left(E_\theta u_\theta\left(\tilde{P}\right) - EU_0\right)^\gamma \left(\frac{1}{t}\left(\tilde{P}-\tilde{k}\right)Q\left(\tilde{P}\right)\right)^{1-\gamma}.$$

Differentiating this expression and rearranging the first-order condition yields

$$\frac{\left(\tilde{P}-\tilde{k}\right)}{\tilde{P}} = \frac{1}{\frac{\gamma}{(1-\gamma)}\frac{E_\theta q_\theta\left(\left(\tilde{P}\right)\right)}{E_\theta u_\theta\left(\tilde{P}\right)-EU_0} - \varepsilon}. \tag{26}$$

When solving the full game, this equation replaces its counterpart in the unregulated case, namely (11). Interestingly, the RHS of equation (26), like that of equation (11) does not depend on $\tilde{k}$. Consequently one can proceed in exactly the same way as in section 4.4, so that the program of the insurer continues to be the one stated by (22)–(23) and we are back to the second-best rule stated by (8).

To sum-up, the solution is exactly the same whether or not the monopoly is regulated; profit maximization and price bargaining give the same result. This is true, whatever the bargaining weight of the firm, which may appear surprising at first. The key to understanding this is that the insurance provider's reimbursement policy is a "very powerful" tool since it can bring the producer to a zero profit level in any case.

6 Differentiated oligopoly with N firms: discrete choice approach

To study this setting, we have to redefine not only Stage 2 but also Stage 3 of our game. Specifically we have to redefine the consumer's preferences to account for the possibility to choose between N variants of a horizontally differentiated medical product. Suppose now there are $N \leq I$ firms in the market indexed by $i = 1....N$. Firms are ranked by their fixed cost F^i, $F^1 \leq ... \leq F^i \leq F^{i+1} \leq ...F^I$. In order to make the problem tractable, we now assume that a patient $\theta \in \Theta$ consuming q^i units from producer i derives a utility:

$$u_\theta^i = u\left(x^i\right) + v\left(q^i, \theta\right) - s\left(\tilde{z}_\theta^i\right),$$

where $\tilde{z}_\theta^i \in Z$ measures a random cost of consuming product i over Θ^I. The v function represents the health status of the patient which is assumed to be separable from the utility of consumption[14]. The random variables $\tilde{z}_\theta^i$ are iid and their (marginal) distribution is exponential with a mean normalized to 1, so that their cdf is represented by $H(z) = 1 - \exp(-z)$. We further assume that the cost function is defined by $s(z) = \exp(\alpha z)$, where $\alpha > 0$ measures how substitutable the horizontally differentiated products q^i are, with a lower α implying a higher degree of substitution between the products.[15]

6.1 The individual problem reconsidered

We assume that the N variants of the medical product are consumed in a mutually exclusive way, *i.e.* each patient of type θ consumes a single product i. This property applies to many markets for medical products. For instance patients often use a single drug selected amongst different molecules that treat the same disease; z_θ^i measures in this case possible side effects caused by drug i on patient of type θ. Patients may also choose hospitals or specialists with respect to the geographical distance or the kind of treatment they offer, etc.

Individual demand conditional on consuming variant i, q_θ^i is defined by

$$q_\theta^i = \arg\max_{q^i} u\left(w - \pi - \tilde{P}^i q^i\right) + v\left(q^i, \theta\right),$$

where $\tilde{P}^i = c + tP^i$.

The first-order condition is given by

$$-\tilde{P}^i u'\left(w - \pi - \tilde{P}^i q^i\right) + \frac{\partial v\left(q^i, \theta\right)}{\partial \tilde{P}^i} = 0,$$

which implicitly defines $q_\theta^i \equiv q_\theta\left(\tilde{P}^i, \tilde{w}\right)$, the (conditional) demand of individual θ for q^i. An individual of type θ choosing variant i obtains an indirect utility $V_\theta^i \equiv V_\theta\left(\tilde{P}^i, \tilde{w}\right) =$

[14] This assumption is compatible with the findings of Finkelstein et al. (2013).

[15] For a survey of discrete choice product differentiation models, see Anderson et al. (1992).

$\omega^i_\theta\left(\tilde{P}^i,\tilde{w}\right) - s\left(\tilde{z}^i_\theta\right)$ where $\omega^i_\theta\left(\tilde{P}^i,\tilde{w}\right) \equiv u\left(\tilde{w} - \tilde{P}^i q^i_\theta\right) + v\left(q^i_\theta,\theta\right)$ denotes the indirect utility gross of the random cost of consuming variant i.

An individual of type θ chooses the variant i that solves

$$\max_{i=1\ldots N}\left\{V^i_\theta\right\}.$$

In a symmetric equilibrium where $\tilde{P}^i = \tilde{P}$ for every $i \in (1,N)$, each individual of type θ minimizes the loss so its indirect utility function writes:

$$V^i_\theta = \omega_\theta\left(\tilde{P},\tilde{w}\right) - \min_i s\left(\tilde{z}^i_\theta\right). \tag{27}$$

As a result, at a symmetric equilibrium, the expected indirect utility function writes:

$$\begin{aligned} E_\theta V_\theta &= \omega_\theta\left(\tilde{P},\tilde{w}\right) - E_\theta\left\{\min_i\left[\exp\alpha\tilde{z}^i_\theta\right]\right\} \\ &= \omega_\theta\left(\tilde{P},\tilde{w}\right) - \frac{N}{N-\alpha}. \end{aligned}$$

Observe that the number of variants N, has a *direct* positive impact on expected utility. This makes sense within the context of medical products. As a larger diversity of treatments becomes available, patients have more options when choosing the product which has the smallest adverse effect. In addition, the number of variants also has an *indirect* effect on utility because it affects the market price, the determination of which we shall now study.

6.2 Equilibrium price

To determine the symmetric Nash equilibrium price, we have to determine the best-response of producer i, $\tilde{P}^i$, to a price $\tilde{P}$ set by all other firms. Appendix C shows that the expected demand for product i of individual θ is given by

$$D^i_\theta\left(\tilde{P}^i,\tilde{P}\right) = q^i_\theta\int_0^\infty \exp(-z)\left(\alpha\left(\omega\left(\tilde{P},\theta\right) - \omega\left(\tilde{P}^i,\theta\right)\right) + \exp(\alpha z)\right)^{-\frac{N-1}{\alpha}} dz, \tag{28}$$

and that the producer's total demand is equal to

$$D^i\left(\tilde{P}^i,\tilde{P}\right) = E_\theta\left[D^i_\theta\left(\tilde{P}^i,\tilde{P}\right)\right].$$

Differentiating this expression with respect to $\tilde{P}^i$ yields

$$\frac{\partial D^i}{\partial \tilde{P}^i} = E_\theta \frac{\partial D^i_\theta \left(\tilde{P}^i, \tilde{P}\right)}{\partial \tilde{P}^i}. \tag{29}$$

Using (28) and rearranging shows that

$$\frac{\partial D^i_\theta \left(\tilde{P}^i, \tilde{P}\right)}{\partial \tilde{P}^i} = \frac{1}{N}\frac{\partial q^i_\theta}{\partial \tilde{P}^i} + \frac{(N-1)}{N+\alpha} q^i_\theta \frac{\partial \omega \left(\tilde{P}^i, \theta\right)}{\partial \tilde{P}^i}.$$

Using (29), Roy's identity, and evaluating at $\tilde{P}^i = \tilde{P}$ we obtain

$$\frac{\partial D^i \left(\tilde{P}, \tilde{P}\right)}{\partial \tilde{P}^i} = E_\theta \left(-\frac{(N-1)}{N+\alpha} \left(q_\theta \left(\tilde{P}\right)\right)^2 u'(x_\theta) + \frac{1}{N}\frac{\partial q^i_\theta}{\partial \tilde{P}} \right). \tag{30}$$

Each producer i maximizes its profit and solves

$$\max_{\tilde{P}^i} \Pi^i = \frac{1}{t}\left(\tilde{P}^i - \tilde{k}\right) D^i \left(\tilde{P}^i, \tilde{P}\right) - F^i.$$

The first-order condition is

$$D\left(\tilde{P}^i, \tilde{P}\right) + \left(\tilde{P}^i - \tilde{k}\right) \frac{\partial D^i \left(\tilde{P}^i, \tilde{P}\right)}{\tilde{P}^i} = 0. \tag{31}$$

Define $Q\left(\tilde{P}\right) = ND^i(\tilde{P}, \tilde{P})$. In a symmetric equilibrium each producer has demand $D^i(\tilde{P}, \tilde{P}) = Q\left(\tilde{P}\right)/N$. Using (30) and rearranging shows that (31) can be rewritten as

$$\frac{\left(\tilde{P} - \tilde{k}\right)}{\tilde{P}} = \frac{1}{\frac{N(N-1)}{N+\alpha}\tilde{P}\frac{E_\theta\left\{\left(q_\theta\left(\tilde{P}\right)\right)^2 u'(x_\theta)\right\}}{Q\left(\tilde{P}\right)} - \varepsilon}, \tag{32}$$

where ε is the price elasticity of Q. When solving the full game, this equation replaces its counterpart in the monopoly case, namely (11) and defines $\tilde{P}$ as a function of $\tilde{k}$ and N. Observe that the RHS of this expression, like that of equation (11) in the monopoly case, and equation (26) in the bargaining case, does not depend on $\tilde{k}$.

The equilibrium profit of firm i is then given by

$$\Pi^{i*}(t, c, N) = (P - k)\, Q\left(\tilde{P}\right)/N - F^i,$$

where $\tilde{P}$ is determined by (32), so that P follows from (1).

6.3 The insurance provider's program

There are I potential firms ranked and indexed in increasing order according to their increasing fixed costs $F^1 \leq ... \leq F^i \leq F^{i+1} \leq ...F^I$. We consider two different problems. First, we assume that the number of active firms N is given. This is the short-run problem. Second, in the long-run problem, N can also be chosen.

6.3.1 Short-run

In the short-run, the number of active firms and available variants of the medical products, N, is exogenously given. The reimbursement policy must be designed such that all active firms make positive profits. In a symmetric equilibrium, firms differ only in fixed costs so that it is the participation constraint of producer N which is binding. In words, the active firm with the highest fixed cost realizes zero profits.

Recall that the RHS of equation (32) does not depend on $\tilde{k}$ and therefore one can repeat the same reasoning as the one in Section 4.4. The participation constraint of producer N will be binding so that the insurance provider's program is

$$\begin{array}{ll} \max\limits_{\pi,\tilde{P}} & \int_\theta u(w_0 - \pi - \tilde{P} q^*_\theta(\tilde{P}, \tilde{w})) + v\left(q^*_\theta(\tilde{P}, \tilde{w}), \theta\right) dG(\theta) - \dfrac{N}{N-\alpha} \\ \text{s.t.} & \pi + \left(\tilde{P} - k\right) Q^*(\tilde{P}, \tilde{w}) - NF^N \geq 0. \end{array}$$

Again the second-best reimbursement rule (8) applies.

6.3.2 Long-run

For the sake of simplicity, we neglect the fact that N must be an integer and denote by $F(N)$ the sunk cost of firm $N \leq I$ with $F'(N) > 0$. There are two ways to think about this long-run solution: either there is free-entry so that firms enter until $\Pi^{N*}(t, c, N) = 0$, or the insurance provider selects the N firms to be on the list of the reimbursable medical products, which once again implies $\Pi^{N*}(t, c, N) = 0$. Since the insurance provider through its reimbursement policy effectively controls the prices, both of these scenarios are equivalent. Either way, the insurance provider then solves the following problem

$$\max_{\pi,\tilde{P},N} \quad \int_{\theta} u(w_0 - \pi - \tilde{P} q_{\theta}^{*}(\tilde{P},\tilde{w})) + v\left(q_{\theta}^{*}(\tilde{P},\tilde{w}),\theta\right) dG(\theta) - \frac{N}{N-\alpha}$$

$$\text{s.t.} \quad \pi + \left(\tilde{P} - k\right) Q^{*}(\tilde{P},\tilde{w}) - NF(N) \geq 0. \tag{33}$$

The optimal interior solution is again characterized by (8), (33) and the additional first order condition with respect to N

$$F(N) + NF'(N) = \frac{\alpha}{\lambda (N-\alpha)^2},$$

where λ is the Lagrange multiplier associated to the insurance provider's budget constraint. The left hand side represents the budgetary cost of an additional firm while the right hand side represents the (monetary equivalent) expected marginal utility gain brought about by the availability of an additional variant of the medical product.[16]

7 Conclusion

This paper has studied the design of health insurance under *ex post* moral hazard, when there is imperfect competition in the market for the medical product. We have considered various *scenarios,* such as a profit maximizing monopoly, price negotiation or a horizontally differentiated oligogopoly with or without free entry. The insurance contract specifies two types of copayments: an *ad valorem* coinsurance rate and a specific (per unit) copayment rate. By combining both copayment rates in an adequate way the insurer can effectively control the producer price, which is then set so that the producer's revenue just covers fixed costs. This combination involves the lowest possible level of copayment rate (possibly negative) associated with the highest possible level of coinsurance rate such that the participation constraint of health providers is binding. Consequently, a suitable regulation of the copayment instruments leads to the *same* reimbursement rule of individual expenditures as under perfect competition for medical products. Rationing coverage because of imperfect competition as advocated by Feldstein (1973) is thus *not* necessary. Interestingly the optimal policy closely resembles a

[16] An increase in N also affects $\widetilde{P}$ and π. However, by the envelope theorem these induced variations have no first-order effect on the insurer's objective.

reference price mechanism in which copayment rates are low (possibly negative) and coinsurance rates are high. The enforcement of reference pricing is part of the health care reform advocated by the Obama administration. Our results do provide support for this policy. However, for our policy to be fully effective, both instruments have to be regulated; controlling the *average* reimbursement rate as envisioned by the Affordable Care Act is *not* enough. Instead the type of contract and the form of the copayments would have to be regulated.

To keep our model tractable we had to neglect some important aspects. These include first and foremost the quality of health care products. As our results imply full extraction of the producers' surplus, one might object that the policy we characterize may undermine the overall quality of health care services. This objection is only justified in part because the fixed cost can be interpreted as including a "fair" rate of return on the research and development investments incurred by the producers. Still, this leaves open the issue of the "optimal" investment in medical research and more general the determination of the appropriate quality of medical products. When these aspects are accounted for, it might be the case that the optimal policy involves a tradeoff between the extraction of producers surplus (rents) and the level of quality of health care products driven by innovations. Introducing quality competition in a setting with *ex post* moral hazard is a challenging endeavour which is on our agenda for future research.

References

[1] Anderson, S, A. De Palma and J.-F. Thisse. 1992. *Discrete choice theory of product differentiation,* MIT Press, Cambridge, MA.

[2] Arrow, K.J. 1963. "Uncertainty and the welfare economics of medical care". *American Economic Review* 53: 941–973.

[3] Besley, T. 1988. "Optimal reimbursement health insurance and the theory of Ramsey taxation", *Journal of Health Economics* 7: 321–336.

[4] Delipalla, S. and M. Keen. 1992. "The comparison between ad valorem and specific taxation under imperfect competition". *Journal of Public Economics* 49: 351–367.

[5] Diamond, P. 1975. "A many-person Ramsey tax rule". *Journal of Public Economics* 4: 335–342.

[6] Feldman, R. and B. Dowd. 1991. "A new estimate of the welfare loss of excess health insurance". *American Economic Review* 81: 297–301.

[7] Feldstein, M.S. 1973. "The welfare loss of excess health insurance". *Journal of Political Economy* 81: 251–80.

[8] Finkelstein, A., E. Luttmer and M. Notowidigdo. 2013. "What good is wealth without health? The effect of health on the marginal utility of consumption", *Journal of the European Economic Association* 11: 221-258.

[9] Fullerton, D and G.E. Metcalf. 2002. "Tax incidence" in: Auerbach, A. and M. Feldstein eds, *Handbook of Public Economics,* Vol 4, 1788–1872.

[10] Gaynor, M., Haas-Wilson, D. and W.B. Vogt. 2000. "Are Invisible Hands Good Hands? Moral Hazard, Competition, and the Second-Best in Health Care Markets". *Journal of Political Economy* 108: 992-1005.

[11] Lakdawalla, D. and N. Sood. 2013. "Health Insurance as a Two-Part Pricing Contract". *Journal of Public Economics* 102: 1–12.

[12] Myles, G. 1995. *Public Economics*, Cambridge University Press, Cambridge UK.

[13] Pauly, M.V. 1968. "The Economics of Moral Hazard: Comment". *American Economic Review* 58: 531–37.

[14] Robinson, J. C. and T. T. Brown, 2013. "Increases In Consumer Cost Sharing Redirect Patient Volumes And Reduce Hospital Prices For Orthopedic Surgery", *Health Affairs* 32: 1392–1397.

[15] Vaithianathan, R. 2006. "Health insurance and imperfect competition in the health care market". *Journal of Health Economics* 25: 1193–1202.

[16] Zeckhauser, R. 1970. "Medical Insurance: A Case Study of the Trade-Off Between Risk Spreading and Appropriate Incentives", *Journal of Economic Theory* 2, 10–26.

Appendix

A Derivation of equation (8)

Denoting λ the Lagrange multiplier associated to the insurance provider's budget constraint, the FOCs with respect to $\tilde{P}$ and π are respectively given by

$$- \quad E_\theta q_\theta \frac{\partial u_\theta}{\partial x} + \lambda \left[Q(\tilde{P}) + \left(\tilde{P} - k \right) Q'(\tilde{P}) \right] = 0, \tag{34}$$

$$- \quad E_\theta \frac{\partial u_\theta}{\partial x} + \lambda \left[1 + \left(\tilde{P} - k \right) \frac{\partial Q}{\partial \pi} \right] = 0. \tag{35}$$

Multiplying (35) by $E_\theta q_\theta = Q(\tilde{P})$ and subtracting (34) yields:

$$E_\theta q_\theta \frac{\partial u_\theta}{\partial x} - E_\theta q_\theta E_\theta \frac{\partial u_\theta}{\partial x} - \lambda \left(\tilde{P} - k \right) E_\theta \frac{\partial q_\theta^*}{\partial \tilde{P}} + \lambda \left(\tilde{P} - k \right) E_\theta q_\theta E_\theta \frac{\partial q_\theta}{\partial \pi} = 0. \tag{36}$$

Denoting the compensated (Hicksian) demand by q^c and substituting the Slutsky equation

$$\frac{\partial q_\theta^c}{\partial \tilde{P}} = \frac{\partial q_\theta^*}{\partial \tilde{P}} + q_\theta^* \frac{\partial q_\theta^*}{\partial \pi} \tag{37}$$

into equation (36) yields:

$$E_\theta q_\theta \frac{\partial u_\theta}{\partial x} - E_\theta q_\theta E_\theta \frac{\partial u_\theta}{\partial x} - \lambda \left(\tilde{P} - k \right) E_\theta \frac{\partial q_\theta^c}{\partial \tilde{P}} -$$
$$\lambda \left(\tilde{P} - k \right) E_\theta q_\theta \frac{\partial q_\theta^*}{\partial \pi} + \lambda \left(\tilde{P} - k \right) E_\theta q_\theta E_\theta \frac{\partial q_\theta}{\partial \pi} = 0.$$

Using the definition of Λ_θ in expression (9), this can be rewritten as

$$\text{cov}\left(q_\theta, \Lambda^\theta \right) = \left(\tilde{P} - k \right) E_\theta \frac{\partial q_\theta^c}{\partial \tilde{P}},$$

or

$$\frac{\left(\tilde{P} - k \right)}{\tilde{P}} = \frac{\text{cov}\left(q_\theta, \Lambda^\theta \right)}{E_\theta q_\theta S_\theta}, \tag{38}$$

where $S_\theta^* = \left(\tilde{P} / q_\theta^c \right) \left(\partial q_\theta^c / \partial \tilde{P} \right)$. Multiplying by -1 then yields expression (8).

B Derivation of equation (21)

Using (20), problem (18)–(19) can be written as

$$\begin{aligned}\max_{\pi,t}\ E_\theta u_\theta &= \int_\theta u(w-\pi-\tilde{P}^M q_\theta^*\left(\tilde{P}^M,\tilde{w}\right), q_\theta^*\left(\tilde{P}^M,\tilde{w}\right))dG(\theta).\\ \text{s.t.}\quad &: \quad \pi + cQ^*\left(\tilde{P}^M,\tilde{w}\right)\geq 0.\end{aligned}$$

Assuming that $|\varepsilon|$ is a constant and denoting λ the Lagrange multiplier associated to the insurance provider's budget constraint, the first-order conditions with respect to c and π are respectively given by:

$$- \quad \frac{1}{1-\frac{1}{|\varepsilon|}}E_\theta q_\theta^*\frac{\partial u_\theta}{\partial x}+\lambda\left[Q^*(\tilde{P})-\left(P^M-\tilde{P}^M\right)\frac{1}{1-\frac{1}{|\varepsilon|}}\frac{\partial Q^*}{\partial \tilde{P}^M}\right]=0, \tag{39}$$

$$- \quad E_\theta\frac{\partial u_\theta}{\partial x}+\lambda\left[1-\left(P^M-\tilde{P}^M\right)\frac{\partial Q^*}{\partial \pi}\right]=0. \tag{40}$$

Multiplying (39) by $Q(\tilde{P})$ and subtracting (40) yields:

$$\frac{1}{1-\frac{1}{|\varepsilon|}}E_\theta q_\theta\frac{\partial u_\theta}{\partial x}-E_\theta q_\theta E_\theta\frac{\partial u_\theta}{\partial x}+\lambda\frac{1}{1-\frac{1}{|\varepsilon|}}\left(P^M-\tilde{P}^M\right)E_\theta\frac{\partial q_\theta^*}{\partial \tilde{P}}-\lambda\left(P^M-\tilde{P}^M\right)E_\theta q_\theta E_\theta\frac{\partial q_\theta}{\partial \pi}=0. \tag{41}$$

Using the Slutsky equation (37) and rearranging yields

$$\begin{aligned}&\frac{1}{1-\frac{1}{|\varepsilon|}}E_\theta q_\theta^*\frac{\partial u_\theta}{\partial x}-E_\theta q_\theta^* E_\theta\frac{\partial u_\theta}{\partial x}+\lambda\frac{1}{1-\frac{1}{|\varepsilon|}}\left(P^M-\tilde{P}^M\right)E_\theta\frac{\partial q_\theta^c}{\partial \tilde{P}}+\\ &\lambda\frac{1}{1-\frac{1}{|\varepsilon|}}\left(P^M-\tilde{P}^M\right)E_\theta q_\theta^*\frac{\partial q_\theta^*}{\partial \pi}-\lambda\left(P^M-\tilde{P}^M\right)E_\theta q_\theta^* E_\theta\frac{\partial q_\theta^*}{\partial \pi}=0.\end{aligned}$$

Multiplying by $1/\left(1-1/|\varepsilon|\right)$ and using the definition of Λ_θ in expression (9), this can be rewritten as:

$$\operatorname{cov}\left(q_\theta^*,\Lambda_\theta\right)+\frac{1}{|\varepsilon|}\left(\frac{E_\theta q_\theta^* E_\theta\frac{\partial u_\theta}{\partial x}+\lambda\left(P^M-\tilde{P}^M\right)E_\theta q_\theta^* E_\theta\frac{\partial q_\theta^*}{\partial \pi}}{\lambda}\right)=\left(\tilde{P}^M-P^M\right)E_\theta\frac{\partial q_\theta^c}{\partial \tilde{P}}.$$

Using (39) yields

$$\operatorname{cov}\left(q_\theta^*,\Lambda_\theta\right)+\frac{1}{|\varepsilon|}Q^*=\left(\tilde{P}^M-P^M\right)E_\theta\frac{\partial q_\theta^c}{\partial \tilde{P}},$$

so that

$$\left(\tilde{P}^M-P^M\right)=\frac{\frac{1}{|\varepsilon|}Q^*+\operatorname{cov}\left(q_\theta^*,\Lambda_\theta\right)}{E_\theta\frac{\partial q_\theta^c}{\partial \tilde{P}}}.$$

Multiplying by -1, dividing by $\tilde{P}^M$ and using the definition of S_θ^* then yields expression (21).

C Derivation of equation (28)

The probability that variant i is chosen by a patient θ is determined by

$$\begin{aligned}
& \Pr\left(\omega\left(\tilde{P}^i,\theta\right) - s\left(z_\theta^i\right) \geq \omega\left(\tilde{P},\theta\right) - \min_{l\neq i}\left\{s\left(\tilde{z}_\theta^l\right)\right\} \mid z_\theta^i\right) \\
= \ & \Pr\left(\min_{l\neq i}\left\{s\left(s_\theta^l\right)\right\} \geq \omega\left(\tilde{P},\theta\right) - \omega\left(\tilde{P}^i,\theta\right) + s\left(z_\theta^i\right) \mid z_\theta^i\right) \\
= \ & \exp\left(-\log\left(\alpha\left(\omega\left(\tilde{P},\theta\right) - \omega\left(\tilde{P}^i,\theta\right)\right) + \exp\left(\alpha z_\theta^i\right)\right)\right)^{\frac{N-1}{\alpha}} \\
= \ & \left(\alpha\left(\omega\left(\tilde{P},\theta\right) - \omega\left(\tilde{P}^i,\theta\right)\right) + \exp\left(\alpha z_\theta^i\right)\right)^{-\frac{N-1}{\alpha}}.
\end{aligned}$$

Integrating over all possible realizations of z_θ^i establishes (28).