WORKING PAPERS

N° 1223

May 2021

Coinsurance vs. copayments: reimbursement rules for a monopolistic medical product with competitive health insurers

Helmuth Cremer and Jean-Marie Lozachmeur

Toulouse School of Economics

Coinsurance vs. copayments: reimbursement rules for a monopolistic medical product with competitive health insurers[1]

Helmuth Cremer[2]
and Jean-Marie Lozachmeur[3]

May 2021

[1]Financial support from the Chaire "*Marché des risques et creation de valeur*" of the FdR/SCOR is gratefully acknowledged. We also gratefully acknowledge the funding received by TSE from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).
[2]Toulouse School of Economics, University of Toulouse Capitole, France.
[3]Toulouse School of Economics, CNRS.

Abstract

This paper studies a market for a medical product in which there is perfect competition among health insurers, while the good is sold by a monopolist. Individuals differ in their severity of illness and there is *ex post* moral hazard. We consider two regimes: one in which insurers use coinsurance rates (*ad valorem* reimbursements) and one in which insurers use copayments (specific reimbursements). We show that the induced equilibrium with copayments involves a lower producer price and a higher level of welfare for consumers. This results provides strong support for a reference price based reimbursement policy.

JEL Codes: I11, I13, I18.

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

1 Introduction

Following Feldstein (1970, 1973) and later on Feldman and Dowd (1991), there is a common view among economists that insurance coverage is too large when *(i)* the market for health insurance is perfectly competitive and *(ii)* there is imperfect competition among health care providers.

The reason underlying this result is that competitive insurers do not internalize the effect of their insurance policy on the pricing of medical goods. As a result, prices set under monopoly or oligopoly are too high in equilibrium, and there is room for limiting insurance coverage. In an environment with *ex post* moral hazard and linear reimbursement rates, this point has been proved first by Chiu (1997) for the case where demand for health care is perfectly (or sufficiently) inelastic, and later by Vaithinianathan (2006) who allows for elastic demand (but considers Cournot competition). In both papers insurers cover a share of expenditures so that the reimbursement is proportional to the price.

In this paper, we re-consider the same game between insurers and a monopolistic health provider, but allow for two types of copayments used by insurers. In a first case, denoted A, insurers use *ad valorem* copayment rates (also called coinsurance rates), that is copayment rates that are proportional to the value of health care consumed by the patient. This is the type of reimbursement considered in the papers mentioned in the previous paragraph. In a second case, referred to as S, insurers use specific copayments, that is copayments that are proportional to the quantity consumed by the patient, but do not depend on the price. In both scenarios, the equilibrium concept is the same. As in the aforementioned papers, insurers compete by choosing a level of premium and a level of copayment for a given price of the medical product while the monopolist chooses its price taking as given the insurance policy chosen by insurers and both markets equilibrate simultaneously.

The medical product cum insurance markets we consider can be interpreted in two ways. The monopolistic producer is viewed as private and profit maximizing either way. As to the insurers, one can assume first that they are private and profit maximizing and that there is free entry and perfect competition in the insurance market. In equilibrium and absent of adverse selection the equilibrium contract then maximize consumers' expected utilities. The role of regulation is confined to the specification of the *type* of reimbursement (A or S) rule that is used. This is the main interpretation adopted for the writing of this paper.

Alternatively on could assume that there is a single public and welfare maximizing insurer but that it can only commit to the *type* of rule that is used and not to the specific *levels* of reimbursement rates. Formally, both interpretations yield exactly the same game and thus the same results.

The main result of this paper is that under very general specification of individual preferences, the level of individuals' welfare is strictly higher when insurers compete on copayment rates as opposed to coinsurance rates. Our result stems from the fact that for a given copayment level paid by the insuree, the producer price chosen by a monopolist will be lower under specific reimbursement.

The following three papers are the most closely related. In the first of these papers Gaynor et al. (2000) show that a lower producer price of medical good is always welfare superior, even if there is *ex post moral* hazard. Second Lakdawalla and Sood (2013), show that if the monopolist moves first and uses general non linear prices, the outcome of the market can lead to the same level of insurance coverage as under perfect competition, but in which the entire surplus is absorbed by the monopolist. Their result however rests on the strong assumption that the monopolist can choose both quantity and price. Finally, Cremer et. al. (2016) consider a setting in which a public insurer moves first. They show that a suitable regulation of the copayment instruments leads to the same reimbursement rule as under perfect competition for medical products. In contrast to these two latter contributions, we study the equilibrium of a market in which neither the insurance companies nor the provider of the medical good can commit to their policy, so that the equilibrium of the market is a Nash equilibrium in which insurers choose their insurance policy for a given price of the medical good and the monopolist chooses its price for a given level of insurance policy.

By analogy with the industrial economics literature one can think of Cremer et. al.

(2016) as a setting where the regulator or national insurer is a Stackelberg leader with full commitment power. In the current model, commitment is limited and when it comes to the level of reimbursement the insurer is simply a Nash competitor. While Cremer et. al. (2016) and Lakdawalla and Sood (2013) are useful as normative benchmark it is known from the regulation literature that commitment to specific levels such as caps on prices is difficult to enforce in a credible way (because the outcome is not subgame perfect and the both parties would benefit from renegotiation). Consequently, this paper and our earlier one are complementary and interestingly, while the specification of the game differs both yield similar results in that they point to the superiority of copayments (per unit reimbursement) over *ad valorem* rules.

Our results are consistent with the empirical literature. Specific copayments can indeed be interpreted as reference pricing. In its extreme form, the reference pricing mechanism sets a fixed level of reimbursement per unit of product. Above this level, the patient pays the difference between the producer price and the fixed level of reimbursement. Recent evidence shows that the introduction of this policy yields a significant decrease in producer price (e.g., see Aouada et al (2019) or Whaley and Brown (2018)).

2 The Model

Consider a mass 1 of *ex ante* identical individuals whose exogenous income is w. Their *ex post* type θ,representing the state of health, is distributed over $\Theta \equiv \left[\underline{\theta}, \overline{\theta}\right]$ according to $G(\theta)$ with a density $g(\theta)$. The health state is iid. across individuals. Preferences are represented by a Von-Neumann-Morgenstern utility function $u_\theta \equiv u\left(y\left(\theta\right)\right)$, where $y = x + h\left(q, \theta\right)$. The variable x denotes consumption of a numeraire good, while q represents the consumption of a medical product, which is sold by a monopolist at a (producer) price of P. The function u is strictly increasing and strictly concave while the function h is increasing and strictly concave in q but decreasing in θ. In words, a larger θ indicates a larger adverse health shock. We assume that $\partial^2 h(q, \theta)/\partial q \partial \theta > 0$, so that marginal benefit of the medical product is increasing in the health risk.

The consumption of q is chosen *ex post.* Its cost may be covered by a health insurance scheme, which can be seen as a partial subsidy on P. Concerning the reimbursement scheme, we consider two regimes, A and S. In regime A, the reimbursement rate chosen by insurers is *ad valorem. The insurance then covers* a given *share* of Pq, the total expenses on the medical product. Consequently, the out of pocket costs which represent the remaining share are determined by an *ad valorem* copayment *rate*, denoted by t, which in the health insurance literature is often referred to as “coinsurance” rate. The consumer price is then given by $\tilde{P}^A = tP$. In the second regime, referred to as “specific”, the insurers reimburse a fixed amount per unit. The reimbursement is simply proportional to the quantity q but independent of the producer price P. In practice this corresponds to the case where the reimbursement is based on an exogenous “reference price”. Let c denote the per unit insurance coverage, so that the net (or consumer) price of the medical product is given by $\tilde{P}^S = P - c$. In the literature this is often referred as copayment. Of course, as long as there is some insurance, $t < 1$ and $c > 0$.

To buy an insurance contract offering a consumer price $\tilde{P}$ for the medical product, policyholders pay a premium π.[1] In state θ, the consumption of the numeraire good is thus given by

$$x_\theta = w - \pi - \tilde{P} q_\theta, \tag{1}$$

where q_θ denotes the consumption of the medical product in state θ. Expected utility can now be written as

$$E_\theta u_\theta = \int_\Theta u\left(y\left(\theta\right)\right) dG(\theta) = \int_\Theta u(w - \pi - \tilde{P} q_\theta + h\left(q_\theta, \theta\right)) dG(\theta),$$

where E_θ denotes the expectation operator over Θ.

2.1 Timing

The timing of the game is as follows. In Stage 0 the regulator decides which regime A or S insurers have to adopt. In Stage 1 insurers simultaneously choose the contract they offer and which is bought by consumers, while the producer sets the price P. In Stage 2, the state of health is realized for each individual who choose the consumption of the medical product q given the consumer price implied the insurance contract bought in Stage 1.

[1] The superscript $i = A, S$ is omitted when not necessary.

To prepare the grounds for studying the Nash equilibrium of Stage 1, we first study the individuals' *ex post* problem. Then we specify the strategies and payoffs of the insurers and the producer under each of the two regimes.

2.2 The individual problem

Once the state of nature θ is revealed policyholders choose their consumption of q to solve

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

taking $\tilde{P}$ and π as given. The first-order condition with respect to q is given by

$$-\tilde{P} + \frac{\partial h(q,\theta)}{\partial q} = 0. \tag{2}$$

Assuming an interior solution the following condition holds for all $\theta \in \Theta$

$$\tilde{P} = \frac{\partial h(q,\theta)}{\partial q}. \tag{3}$$

Equation (3) implicitly defines the standard Marshallian demand in state θ, $q_\theta^* \equiv q_\theta^*\left(\tilde{P}\right)$, as a function of the (consumer) price. Expected demand is denoted by $Q^* \equiv Q^*\left(\tilde{P}\right) = E_\theta q_\theta^*\left(\tilde{P}\right)$, while $u_\theta^* = u\left(y_\theta^*\right) = u\left(w - \pi - \tilde{P}q_\theta^* + h\left(q_\theta^*, \theta\right)\right)$ stands for the level of indirect utility in state θ. Demand elasticity is denoted $\varepsilon\left(\tilde{P}\right) = \left(\tilde{P}/Q^*\left(\tilde{P}\right)\right)\partial Q^*\left(\tilde{P}\right)/\partial\tilde{P} < 0$.

For future reference, the following lemma characterizes the consumption and utility levels as a function of the health state.

Lemma 1 *y_θ^* is decreasing in θ while q_θ^* is increasing.*

Proof. Differentiation of $y_\theta^* = w - \pi - \tilde{P}q_\theta^* + h\left(q_\theta^*, \theta\right)$ with respect to θ, while using the envelope theorem yields

$$\frac{\partial y_\theta^*}{\partial\theta} = \frac{\partial h(q_\theta^*, \theta)}{\partial\theta} < 0,$$

which is negative by definition. Moreover, total differentiation of (2) establishes

$$\frac{\partial q_\theta^*}{\partial\theta} = -\frac{\frac{\partial^2 h(q,\theta)}{\partial q\partial\theta}}{\frac{\partial^2 h(q,\theta)}{\partial q^2}} > 0,$$

where the sign of the RHS follows from our assumptions on $h(q, \theta)$. ■

An individual's choices depend on the state of nature and on the consumer price. They do not *directly* depend on the reimbursement regime that is adopted by insurers. This reimbursement regime along with the producer price, P, determine the consumer price, $\tilde{P}$, which is the only cost variable relevant to the consumer.

2.3 Insurers

Given that there is perfect competition in the insurance market, insurers will enter the market as long as profits are positive. Consequently, the objective of a representative insurer is to maximize the expected utility of the representative individual subject to the break-even constraint. This condition requires that the premium equals the expected cost of the insurance coverage so that the insurer's expected profit is zero. We consider a simultaneous game played by a representative insurers on the one hand and the monopolistic producer of the medical product on the other hand. The insurer's strategy is the vector $\{\pi, t\}$ in regime A and a vector $\{\pi, c\}$ in regime S. The monopolist chooses the producer price P in either scenario. In regime A, a Nash equilibrium is triplet $\left\{P^{*A}, \pi^{*A}, t^{*A}\right\}$ so that each players strategy is the best- response to the other player's strategy. In regime S the equilibrium is described by $\left\{P^{*S}, \pi^{*S}, t^{*S}\right\}$.

In regime A the best-response levels of premium and coinsurance rates are given determined by

$$\left\{\pi^{*A}(P), t^{*A}(P)\right\} \equiv \left\{ \begin{array}{c} \arg\max_{\pi,t} E_\theta u_\theta^* = \int_\theta u(w - \pi - tPq_\theta^*, q_\theta^*, \theta) dG(\theta), \\ \text{s.t.} \quad \pi - (P - tP) E_\theta q_\theta^* = 0. \end{array} \right\} \tag{4}$$

In regime S we have

$$\left\{\pi^{*S}(P), c^{*S}(P)\right\} \equiv \left\{ \begin{array}{c} \arg\max_{\pi,c} E_\theta u_\theta^* = \int_\theta u(w - \pi - (P - c)\, q_\theta^*, q_\theta^*, \theta) dG(\theta), \\ \text{s.t.} \quad \pi - cE_\theta q_\theta^* \geq 0, \end{array} \right\} \tag{5}$$

2.4 Producer

The medical product is supplied by a profit maximizing monopoly. Denoting by k and F respectively the marginal and the fixed cost, which includes all the sunk costs incurred during the development of the medical product, the best-response producer prices in the two regimes are determined by

$$P^{*A}(t) \equiv \left\{ \begin{array}{c} \arg\max_P \;\; \Pi = (P-k)\,Q^* - F, \\ \text{s.t.} \qquad Q^* = E_\theta q_\theta^*(tP) \end{array} \right\} \tag{6}$$

and

$$P^{*S}(c) \equiv \left\{ \begin{array}{c} \arg\max_P \;\; \Pi = (P-k)\,Q^* - F, \\ \text{s.t.} \qquad Q^* = E_\theta q_\theta^*(P-c) \end{array} \right\} \tag{7}$$

3 The two regimes

Our main point is to show that policy holders have higher expected utility under regime S than under A. Roughly speaking this is because it implies a lower equilibrium producer price. We now turn to the study of the best-response functions. To compare expected utility levels it is convenient to take a "detour" and characterize the combinations of producer and consumer prices $(P, \widetilde{P})$ that can be generated by the best responses both on the insurer's and on the producers sides. We represent these by the function $\widetilde{P}(P)$for the insurer and $P(\widetilde{P})$ for the producer. While these are not strictly speaking best-response functions, it is plain that their intersection corresponds to the Nash equilibrium levels.

3.1 The insurer's best-response

Substituting t by $\tilde{P}/P$ and c by $P-\tilde{P}$, the problem of the insurer as described by (4) and (5) can be rewritten as the choice of an insurance package $\left\{\pi(P), \tilde{P}(P)\right\}$ that solves

$$\left\{\pi(P), \tilde{P}(P)\right\} = \left\{ \begin{array}{c} \arg\max_{\pi,\tilde{P}} E_\theta u_\theta = \int_\theta u(w - \pi - \tilde{P}q_\theta^* + h(q_\theta^*, \theta))dG(\theta), \\ \text{s.t.} \qquad \pi - \left(P-\tilde{P}\right) E_\theta q_\theta^* \left(\tilde{P}\right) \geq 0. \end{array} \right\} \tag{8}$$

This statement of the problem is valid in both regimes. Note that we do *not* change the underlying strategic variables, which remain t or c. Substituting the break-even constraint into the objective function, the problem reduces to choosing $\tilde{P}(P)$ so that

$$\tilde{P}(P) = \arg\max_{\tilde{P}} E_\theta u_\theta = \int_\theta u(w - \left(P-\tilde{P}\right) E_\theta q_\theta^* - \tilde{P}q_\theta^* + h(q_\theta^*, \theta))dG(\theta), \tag{9}$$

Differentiating (9) and using the envelope theorem yields the following first-order condition

$$\begin{aligned}
&E_\theta\left[q_\theta^*\right] E_\theta\left[u'\left(w-\left(P-\tilde{P}\right) E_\theta q_\theta^*-\tilde{P} q_\theta^*+h\left(q_\theta^*, \theta\right)\right)\right]- \\
&\left(P-\tilde{P}\right) E_\theta \frac{\partial q_\theta^*\left(\tilde{P}\right)}{\partial \tilde{P}} E_\theta\left[u'\left(w-\left(P-\tilde{P}\right) E_\theta q_\theta^*-\tilde{P} q_\theta^*+h\left(q_\theta^*, \theta\right)\right)\right]- \\
&E_\theta\left[q_\theta^* u'\left(w-\left(P-\tilde{P}\right) E_\theta q_\theta^*-\tilde{P} q_\theta^*+h\left(q_\theta^*, \theta\right)\right)\right]=0,
\end{aligned} \tag{10}$$

which can be rewritten as

$$\left(P-\tilde{P}\right) E_\theta \frac{\partial q_\theta^*\left(\tilde{P}\right)}{\partial \tilde{P}}=-\frac{\operatorname{cov}\left(q_\theta^*, u'\left(w-\left(P-\tilde{P}\right) E_\theta q_\theta^*-\tilde{P} q_\theta^*+h\left(q_\theta^*, \theta\right)\right)\right)}{E_\theta\left[u'\left(w-\pi\left(P, \tilde{P}\right)-\tilde{P} q_\theta^*+h\left(q_\theta^*, \theta\right)\right)\right]}$$

or

$$\frac{\left(P-\tilde{P}\right)}{\tilde{P}}=\frac{1}{|\varepsilon|} \frac{\operatorname{cov}\left(q_\theta^*, u'\left(y_\theta^*\right)\right)}{E_\theta q_\theta^* E_\theta u'\left(y_\theta^*\right)}. \tag{11}$$

Equation (11) implicitly defines the function $\tilde{P}(P)$ describing the set of pairs $(P, \widetilde{P})$ which are compatible with a best-response strategy of the insurer. Recall that this function is the same in the two regimes.

Lemma 2 *We have $\tilde{P}(P)<P$ for any P.*

Proof. The proof of follows immediately from Lemma 1. Since q_θ^* is increasing in θ while y_θ^* is decreasing in θ, $\operatorname{cov}\left(q_\theta^*, u'\left(y_\theta^*\right)\right)>0$ because $u" < 0$. Equation (11) thus yields $\tilde{P}<P$ for any P. ■

Intuitively, Lemma 2 simply states that for any producer price, as long as consumers are risk averse, the insurance contract always effectively provides some insurance. To see this note that in Regime S, $\tilde{P}<P$ implies $c=P-\tilde{P}>0$ while it yields $t=\tilde{P}/P<1$ under Regime A. In other words, in both cases consumers do not pay the full cost of their consumption of the medical product. While this result needs to be established to ensure that our argument is consistent and complete, it does of course not come as a surprise.

For the remainder of our arguments no monotonicity properties of $\tilde{P}(P)$ are needed. And indeed without further restriction on u, $\tilde{P}(P)$ may or may not be monotonically increasing. This depends on the properties of the utility function. An increase in P is effectively like a decrease in income (via π) which has an ambiguous effects the "demand for insurance". Under CARA for instance $\tilde{P}(P)$ is increasing.

3.2 The producers best response

While the regime has no impact on the insurer's best-reply, it does affect the pricing behavior of the producer. Under regime A the FOC associated with (6) is given by

$$\frac{\partial \Pi^A}{\partial P} = Q^*(tP) + t(P-k)\frac{\partial Q^*(tP)}{\partial \widetilde{P}} = 0, \tag{12}$$

while that under regime S is given by

$$\frac{\partial \Pi^S}{\partial P} = Q^*(P-c) + (P-k)\frac{\partial Q^*(P-c)}{\partial \widetilde{P}} = 0. \tag{13}$$

But this implies that when $tP = P - c$, so that $\widetilde{P}^A = \widetilde{P}^S$, it follows that when

$$\frac{\partial \Pi^S}{\partial P} = 0 \quad \text{we have} \quad \frac{\partial \Pi^A}{\partial P} > 0. \tag{14}$$

as long as $t < 1$ (and profits are concave).[2] So that defining $P^A(\widetilde{P}) = P^{*A}\left(\tilde{P}/P\right)$ and $P^S(\widetilde{P}) = P^{*S}\left(\tilde{P} - P\right)$ we can state the following lemma

Lemma 3 *For any given consumer price associated with $t < 1$ or $c > 0$ we have $P^A(\widetilde{P}) > P^S(\widetilde{P})$.*

To illustrate this property consider the case where the elasticity of the demand for the medical product, ε is constant. In that case problems (6) and (7) yield closed-form solutions which are given by

$$P^A\left(\tilde{P}\right) = \frac{k}{1 - \frac{1}{|\varepsilon|}}, \tag{16}$$

[2] Concavity requires

$$\begin{aligned} \frac{\partial^2 \Pi^A}{\partial P^2} &= t\left(2\frac{\partial Q^*(\widetilde{P})}{\partial \widetilde{P}} + t(P-k)\frac{\partial^2 Q^*(\widetilde{P})}{\partial \widetilde{P}^2}\right) < 0 \\ \text{and } \frac{\partial^2 \Pi^S}{\partial P^2} &= 2\frac{\partial Q^*(\widetilde{P})}{\partial \widetilde{P}} + (P-k)\frac{\partial^2 Q^*(\widetilde{P})}{\partial \widetilde{P}^2} < 0 \end{aligned}$$

so that assuming

$$2\frac{\partial Q^*(\widetilde{P})}{\partial \widetilde{P}} + (P-k)\frac{\partial^2 Q^*(\widetilde{P})}{\partial \widetilde{P}^2} < 0 \tag{15}$$

is enough for concavity of profits in the two regimes (over the relevant range $t < 1$). Note that it is simply the traditional condition for the concavity of a monopolist's problem.

and

$$P^S\left(\tilde{P}\right) = k + \frac{\tilde{P}}{|\varepsilon|}. \tag{17}$$

Combining these expressions and rearranging yields

$$P^S\left(\tilde{P}\right) - P^A\left(\tilde{P}\right) = \frac{\tilde{P} - P^A\left(\tilde{P}\right)}{|\varepsilon|} < 0, \tag{18}$$

as long as $\tilde{P} - P^A\left(\tilde{P}\right)$, which from Lemma 2 must hold in equilibrium (which is on the insurers best-response function). Observe that P^S and P^A intersect only at the 45 degree line, that is when $t = 1$ and $c = 0$. When there is no insurance both regimes are equivalent. This is most obvious for the constant elasticity case, but it also follows from (14) for the general case.

Intuitively Lemma 3 can be understood from the traditional *ad valorem* v. specific tax incidence under monopoly. The traditional result is that for a given consumer price the *ad valorem* tax yields the larger tax revenue and thus a smaller consumer price. In our setting this comparison is reversed because "taxes" are negative; the insurance payment is formally like a subsidy. Consequently, for a given consumer price the *ad valorem* subsidy yields a larger producer price, which is exactly what is stated in Lemma 3.

4 Nash equilibria and welfare

The two Nash equilibria are illustrated in Figure 1 for the special case where $\tilde{P}(P)$ is monotonically increasing, and where demand elasticity is constant. In this case, Regime A yields both a larger producer price and a larger consumer price. However, the comparison of consumer prices relies on the monotonicity of $\tilde{P}(P)$; a simple inspection of the figure shows that when $\tilde{P}(P)$ is first increasing and then decreasing the ranking of consumer prices *may* be reversed.

The comparison of producer prices, on the other hand is robust. This follows from a simple graphical argument, elaborating on Figure 1. The insurer's best-response function $\tilde{P}(P)$ is everywhere below the 45 degree line and in that region we have $P^S\left(\tilde{P}\right) < P^A\left(\tilde{P}\right)$. Consequently the intersection of $\tilde{P}(P)$ with $P^S\left(\tilde{P}\right)$ must lie to the left of its intersection with $P^S\left(\tilde{P}\right)$. This evidently remains true when $\tilde{P}(P)$ is not monotonic.

This property has interesting implications for the welfare properties of the two regimes. To see this let us return to the insurer's problem as stated in 8. Recall that due to competition in the insurance market the equilibrium contract maximizes expected utility of the *ex ante* identical consumers subject to the constraint that the premium covers expected reimbursements. Now it follows from the envelope theorem that the maximum level of expected utility that can be achieved decreases with the producer price P. This in turn implies that Regime S yields a higher expected utility level than Regime A. This establishes our main result summarized in the following proposition.

Proposition 1 *The specific reimbursement regime S yields a lower producer price and a larger expected utility than the ad valorem regime.*

This result provides strong support for the use of reference pricing. Specific copayments can indeed be interpreted as a form of reference pricing with a fixed level of reimbursement per unit of the product.[3] Above this level, the patient pays the difference between the producer price and the fixed level of reimbursement.

5 Conclusion

To summarize, we have shown that when insurers compete using copayment rates rather than coinsurance rates, the market equilibrium, in which the medical product is sold by a monopolist, involves a higher level of welfare for consumers. Intuitively our results can be understood from the traditional *ad valorem* v. specific tax incidence under monopoly. The traditional result is that for a given consumer price the *ad valorem* tax yields the larger tax revenue and thus a smaller consumer price. In our setting this comparison is reversed because "taxes" are negative; the insurance payment is formally like a subsidy.

[3]This refers to case of an *exogenous* references price. Alternatively, and particularly in the case of off-patent product for which there is competition (which is not the case in our paper) the reference price can be endogenous and given for instance by the lowest price within a therapeutic class, see for instance Danzon (2011).

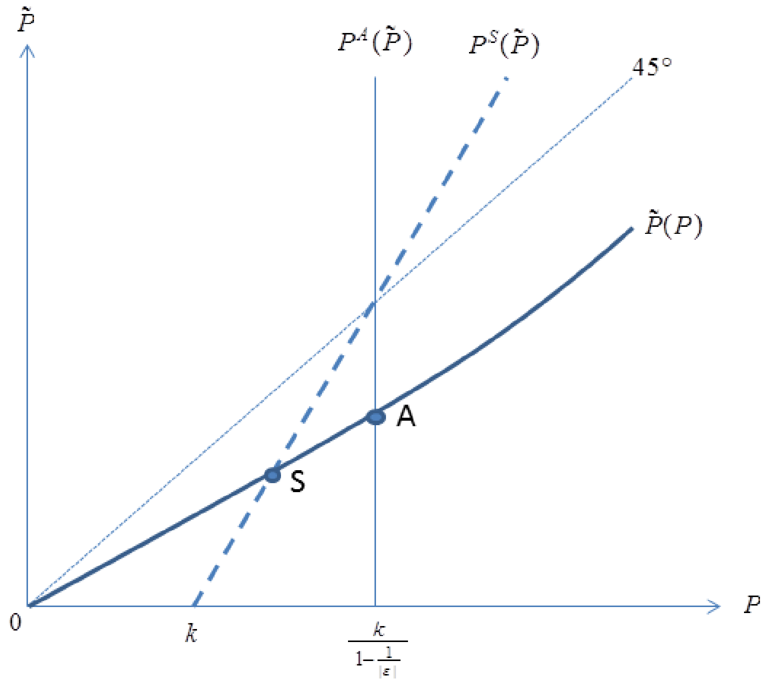

Figure 1: The Nash equilibria

Consequently, for a given consumer price the *ad valorem* subsidy yields a larger producer price.

Because a copayment rule is equivalent to a reimbursement based on a fixed reference price this results provides support for the use of reference pricing even in the context of a monopoly. In practice reference pricing has been used typically for off-patent products for instance in Germany, the Netherlands and New Zealand. The reference price is then often endogenous and for instance given by the lowest price within a the relevant therapeutic class. By contrast we have considered a product produced by a monopoly thus assuming that patent protection has not expired so that to be relevant the reference price has to be exogenous (or obtained by "external benchmarking").

References

[1] Aouada M., T. T. Brown and C.M. Whaley. 2019, Reference pricing: The case of screening colonoscopies, *Journal of Health Economics* 65: 246–259.

[2] Chiu, H.W. 1997, Health insurance and the welfare of health consumers. *Journal of Public Economics* 64: 125-133.

[3] Cremer, H., D. Bardey, and J.M. Lozachmeur. 2016, The design of insurance coverage for medical products under imperfect competition. *Journal of Public Economics* 137: 28–37.

[4] Danzon, P. M., 2011, The economics of the biopharmaceutical industry, in: Gield, S. and P. Smith (eds.), *The Oxford Handbook of Health Economics,* Oxford University Press, Oxford, UK, 520–554.

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

[6] Feldstein, M.S. 1970, The rising price of physician services. *Review of Economics and Statistics* 52: 121-133.

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

[8] 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.

[9] Lakdawalla, D. and N. Sood. 2013, Health insurance as a two-part pricing contract. *Journal of Public Economics* 102: 1–12.

[10] Laffont, J-J. and J. Tirole. 2000, *Competition in Telecommunications,* MIT Press, Cambridge, MA.

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

[12] Whaley, C. M. and T. T. Brown. 2018. Firm responses to targeted consumer incentives: Evidence from reference pricing for surgical services, *Journal of Health Economics* 61: 111–133.