Journal of Health Economics 30 (2011) 1137–1149

Contents lists available at SciVerse ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Competition and the Reference Pricing Scheme for pharmaceuticals

Simone Ghislandi*

ECONPUBBLICA, Department of Institutional Analysis and Public Management, Bocconi University, Via Roentgen 1, 20136, Milan, Italy

ARTICLE INFO

Article history:
Received 12 December 2009
Received in revised form 19 August 2011
Accepted 22 August 2011
Available online 3 September 2011

JEL classification:
I18
L10
L51
L41

Keywords:
Rerence pricing
Pharmaceuticals
Generic drugs
Competition

ABSTRACT

By introducing n (>1) firms with infinite cross-price elasticity (i.e. generic drugs), we explore the effects of competition on the optimal pricing strategies under a Reference Pricing Scheme (RPS). A two-stage model repeated infinite number of times is presented. When stage 1 is competitive, the equilibrium in pure strategies exists and is efficient only if the reference price (R) does not depend on the price of the branded product. When generics collude, the way R is designed is crucial for both the stability of the cartel among generics and the collusive prices in equilibrium. An optimally designed RPS must set R as a function only of the infinitely elastic side of the market and should provide the right incentives for competition.

1. Introduction

In a pharmaceutical Reference Pricing Scheme (RPS), companies are free to set their own prices, but the (insured) consumer pays the difference between the reference price (R) and the actual price of the drug, if this is higher than R. If the product's price is lower than R, the insurance will reimburse the firm only for the price set, and the drug will be available free-of-charge to the patient.

There are two types of products in the market: the branded drug (i.e. patent holder) and the generics (drugs marketed with the name of the Chemical Entity, allowed into the market only after patent expiration and with lower prices than the branded).[1] Since patients pay more for more expensive drugs, generics producers in a RPS should have an incentive to lower their prices in order to undercut their competitors and appear as the "reference" for the reimbursement level. In linking the reimbursement level to the existing prices, the RPS could represent a tool to promote competition among products when private costs can not be observed. Ideally, in a well functioning RPS the reimbursement level should thus be set by the competitive forces.[2]

Despite this promising framework, it has often been commented that RPSs do not respond well to the competitive pressure from generic firms. As a consequence, pharmaceutical authorities normally have to intervene either by imposing discretional generalized price drops or by conditioning the authorization for new entries to significant price reductions (Danzon and Ketcham, 2003). Even more importantly, it has been noted (e.g. Puig-Junoy, 2004) that a price-insensitive reimbursement level might result in generics competing for pharmacists' discounts (i.e. generics sell underpriced to pharmacists, leaving R untouched), thus reallocating competition related welfare gains from society to retailers. Understanding the relation between the RPS and competition is thus of primary importance for the design of an optimal regulatory strategy.

Motivated by these concerns, the main purpose of this paper is to explore the links between generics' pricing strategies and the RPS. Technically, the new features of our model can be summarized in two points. First, we allow a number n (>1) of generics to compete and to collude among themselves in order to manipulate the level

* Tel.: +39 0258365531.
E-mail address: simone.ghislandi@unibocconi.it

[1] In this paper we refer to the form of RPS commonly known as "generic referencing", i.e. drugs subject to the same R have the same active compound and differ only in the brand name. On the opposite, most of the recent theoretical literature analyzed the "therapeutic referencing", where the same R is applied to groups of drugs with different (but equally effective and safe) active compounds. Generic referencing is the most common form of RPS around the world.

[2] There are of course other approaches that link competition with reimbursement. In particular, see Merino (2003), Mestre-Ferrandiz (2001), Brekke et al. (2007), Miraldo (2009) for the relation between RPS and co-payments.

0167-6296/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2011.08.010

of R. Generics' demand is assumed to exhibit infinite cross-price elasticity and is defined residually with respect to the demand of the (only) branded product. Second, we model the endogenous RPS as a two-stage game with free pricing. In the first period R is determined endogenously by firm's behavior, while in the second firms compete taking the level of R as exogenous.

The relation between generics' competition and reimbursement levels has surprisingly received little attention in the economic literature. Since the first theoretical analyzes R has been modeled as *exogenous* to firms' pricing decisions (e.g. Zweifel and Crivelli, 1996; Danzon and Liu, 1997; Mestre-Ferrandiz, 2001). In this setting firms have no incentive to undercut their competitors, since any price lower than R reduces their per-unit revenues without increasing the associated demand (because for consumers the drug is free anyway). This results in an unambiguous profit loss. Therefore, for fixed R, generics competition can not represent a reasonable assumption. The main consequence of the introduction of the RPS is thus on the branded drug only, which now faces a kinked demand, i.e. inelastic below R and price elastic above (Danzon and Liu, 1997).

In this paper these important insights are accounted for by separating the competition *for R* (stage 1) from the competition *given R* (stage 2). The kinked demand and the lack of generics' competition characterize the second stage of the game.

The endogenous nature of R has been recognized in recent literature (Merino, 2003; Brekke et al., 2007; Miraldo, 2009). In Merino (2003) the level of R depends on the price of one branded and one generic products. The level of R is then determined simultaneously by both firms' optimal pricing decisions. In Brekke et al. (2007) the reference price is set to the lowest price in the market. Their objective is to compare the performances of different reimbursement mechanisms, using a variety of outcome measures. In the model from Miraldo (2009) the focus is on collusion between branded and generic firms in both horizontally and vertically differentiated settings. The approach sets on two firms in a two-period game similar to the two-stage approach taken in our model.

For the purposes of this paper, the key issue is that none of the existing studies allow intra-generics competition. This basically reduces to the assumption that generics can perfectly coordinate in a cartel, i.e. competition and coordination problems *among generics* are ruled out *a priori*. A well designed RPS, however, should be able to exploit the high level of competitiveness among generics in order to obtain lower reimbursement levels. In fact, R should be *mainly* determined by the level of competitiveness among generic firms. It is thus important to understand *whether and why* competition works, pointing out, where possible, the role of the RPS in facilitating eventual collusive behaviors.

Building on this framework, this paper has a few important policy and theoretical implications. First of all, a RPS *per se* does not introduce incentives for collusion among generics. However, collusion is very sensitive to the way the RPS is designed. Specifically, if prices can react quickly to an adjustment in R, collusion among generics is the *only* possible outcome, given there is no gain from cheating. This is relevant in all those countries (for example Italy) where the RPS is not accompanied by any other price related policy (e.g. price-percentage co-payments), prices are public knowledge and can drop quickly.

Second, under collusion, if firms coordinate on previous reimbursement prices, R is sticky. Even when the RPS does not introduce in itself serious incentives to collude, R still represents a good coordination price for generics. This explains, under a more correct framework of endogenous RPS, the lack of sensitivity to generics' competition repeatedly pointed out in the empirical literature.

Finally, the model shows that an equilibrium in pure strategies exists, and is efficient, only if R is a function of the perfectly elastic segment of the market. In other words, an optimal RPS should be designed as a function of generics' prices only.[3]

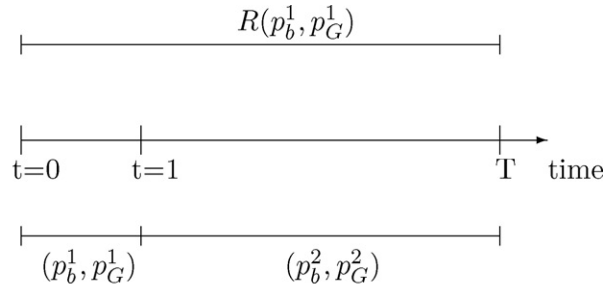

Fig. 1. Timing in the two-stage RPS.

The paper is organized as follows. After presenting the general setting of the model (Section 2), stage 2 of the model is analyzed in Section 3, while Section 4 focuses on the first stage, under the two alternative assumptions of perfect competition (benchmark scenario) and collusive agreement among generics firms. Section 5 defines the conditions for the existence of collusion under a RPS, while Section 6 provides an example of analysis based on linear RPS. The policy implications of the model are discussed in Section 7. Proofs are in Appendix A at the end of the paper.

2. General setting

There are $z = n + 1$ firms competing in the market. One firm produces a branded drug (b) and n firms produce one generic each. The branded drug has been on the market longer than any generic. Consumers have experienced its quality, while they are still uncertain about either the effectiveness or the safeness of the generic version. Hence, their *ex-ante* (or expected) health gains from buying a generic or a branded are different. On the other hand, there is no difference between one generic and another, which implies infinite own and cross-price elasticity among generics. In the following, p_b is the price for the branded product, p_G the vector of the n generic prices and p_{ig} the price of a generic firm i. A superscript will also be added to indicate the stage of the game (either 1 or 2).

2.1. The timing

The RPS is intrinsically dynamic. In our two-stage setting, the level of R is defined in one ($t = 0$) period and *is not changed for the next T periods*.[4] Firms must then take into account the effects of their prices on both actual and, through R, future profits.

The dynamic of the game is the following (Fig. 1). Prices can vary in each period. In $t = 0$ the prices for the first period (p_b^1 and p_G^1) are set. At the same time, R is introduced as a function of the existing prices. R is then kept fixed until the end of the game, which is set in T. Since firms are perfectly informed and R does not change, every stage 2 node of the game is strategically equivalent. In equilibrium, prices set in $t = 1$ (p_b^2 and p_G^2) will not be changed again before T. We will refer to the period from $t = 1$ to T as the second stage and to the period before as the first stage of the game. The game will then be solved backwards, starting from stage 2.

[3] This is not always the case in real world RPS. For example, countries like Germany and, partially, Belgium and Hungary directly or indirectly included the branded price into the definition of R (Galizzi et al., 2011).

[4] Although in many countries it would be technically feasible to update R automatically and simultaneously when prices change, most countries update R formally on a more or less regular basis and with different frequency.

Note that the timing of the game is different and somehow less intuitive than in Miraldo (2009), where R is introduced only in the second stage, after authorities have observed the existing prices. Consequently, in Miraldo the first stage is not subject to the RPS. However, by allowing the game to be repeated without periods of non-application of the RPS, the setting in Fig. 1 suites better the purposes of the present analysis. In addition, it will be shown later that this higher level of initial abstraction is worthwhile, since the model encompasses many existing approaches and can provide useful practical and theoretical insights for the optimal design of a RPS.

Finally note that formally T is the length of the second stage of the game measured in number of "periods". In this model, however, one "period" is not a fixed time length, but should more correctly be seen as the time needed by any firm to react to a price change. The quantity $1/T$ represents the length of the first period relative to the total length of the updating process. In this sense, then, T can increase because both the updating is postponed by authorities and firms can react more quickly to price changes. This last interpretation will be particularly useful in the discussion of the results.

2.2. Competition and equilibria

Firms compete on prices in a simultaneous-move game. At any stage of the game the branded is always competing with the generic producers. In stage 1 generics can either compete or collude among themselves. In stage 2, on the other hand, we simplify the setting by assuming that generics always compete. This assumption does not change the structure of the game and is motivated by three main reasons.

First, if generics could collude on a price higher than R (once R has been set), the game loses interest from a theoretical point of view. The normal framework of collusion under a perfectly elastic demand could be used, with no new insights from the model. A second, more practical, reason to exclude collusion at this stage arises from noting that RPSs are means for public authorities to introduce competition and lower prices in markets for perfectly homogeneous goods. As both R and p_{ig} are known to the authorities, stage 2 collusion is particularly easy to detect: every time all or the majority of the generics set a price higher than R, a collusive agreement must be at work.[5] Of course generics producers would be unlikely to challenge public authorities in such a detectable way. Third, for finite values of Ts, if the game is played only once, it is well-known that a backward induction argument does not allow for collusion. From a theoretical point of view, thus, in order to justify collusion in this general setting we have to make the assumption that either T is infinite or that the game is repeated an infinite number of times. This last scenario will be analyzed later in the paper, but the assumption of stage 2 competition will be maintained.

We focus on the supergame-theoretic approach that addresses the issue of enforcement of collusive behavior. Due to the illegality of explicit collusion, we focus on tacit or implicit collusion. Optimal pricing strategies are found by backward induction. Furthermore, in line with the literature on collusion (e.g. Tirole, 1988), we will refer to a collusive agreement as being "sustainable" if firms have no incentive to cheat, i.e. if no firm can do better by undercutting the collusive price.[6]

[5] This point will become clearer in the next section.

[6] Note that this approach is different from, for example, d'Aspremont et al. (1983) and Escrihuela-Villar (2009), where the concept of cartel stability is introduced in a context where firms' entry and exit are considered.

2.3. The reference price function

R is set in stage 1. It is a function $R(p_b^1, p_G^1)$ of stage 1 market prices, described by the following:

$$\min(p_G^1) \leq R(\cdot) \leq p_b^1 \tag{1}$$

which requires R to be somewhere between the maximum and the minimum price of the market. This is an *ad hoc* general assumption which is not problematic, since it is consistent with the majority of the existing RPS in the world.

In analyzing the impact of R on the price equilibrium of firms, it is useful to introduce the distinction between an R which is a function of all the existing products (e.g. weighted average of all the prices in the market) or an R that depends only on generics (e.g. the minimum price).

Definition. The RPS exhibits **Brand Independence (BI)** if the level of R depends on generics' prices only. Formally, BI requires that $R(p_b^1 + \varepsilon, p_G^1) - R(p_b^1, p_G^1) = 0, \forall \varepsilon$.

On the other hand, when BI does not hold, it is assumed that R is always differentiable with respect to p_b^1. In addition, it is reasonable to impose that $0 < (\partial R(\cdot)/\partial p_b^1) < 1$.

Generics' prices always affect the level of R. For simplicity we assume that R is always differentiable in p_{ig} and that:

$$\frac{\partial R(\cdot)}{\partial \min(p_G^1)} > 0 \tag{2}$$

This excludes the cases in which R is not sensitive to a change in the minimum price.[7] Finally, R is weakly increasing in p_{ig}^1.

2.4. The demand functions

Here we want to keep the analysis as general as possible. To simplify the notation, we also introduce the quantities $\Delta_b \equiv p_b - R$ and $\Delta_{ig} \equiv p_{ig} - R$. We use $D_b(\cdot)$, $D_G(\cdot)$ and $D_{ig}(\cdot)$ to indicate the demand for the branded, the demand for all the generics and the demand for generic firm i, respectively, when the relevant prices are set at $(\cdot)$.

The system of demand relies on three main assumptions. First, generics' demand is own and cross-price infinitely elastic. This is equivalent to assuming that patients perceive generics product as perfectly homogeneous. Furthermore, this assumption also implies that pharmacies always provide the complete set of generic products among which patients can choose from.[8] Second, it is reasonable to assume that when generics and branded prices are the same, patients always choose the branded. More generally, we assume that the demand for the generic market is defined residually with respect to the one for the branded.[9] This also implies that $p_b \geq p_{ig}$. Finally, we make the simplifying assumption that the residual demand is equally shared among all the generic products with the lowest prices. This follows the literature on competition when goods are homogeneous. In this context, however, it should

[7] This condition imposes an oversimplification since in some countries R is defined as, for example, the nth lowest price in the market (e.g. South Africa). In general, results from the following model should not change. However, we prefer to keep the analysis as simple as possible.

[8] Although sometimes this assumption might be unrealistic, it is in line with the most common literature on homogeneous goods.

[9] This assumption is equivalent to seeing branded and generics as vertically differentiated in patients' perceptions, as in Brekke et al. (2007) and Miraldo (2009). A micro-funded example of this is given in Section 6.

also be noted that this assumption is possible because we are ruling out co-payments in the form of price-percentages.[10]

A simple way to proceed is then to imagine that the market size for the drug is given by M consumers (i.e. patients who were prescribed the drug). Starting from the former, the demand of the branded product can be formally defined as:

$$0 < D_b(p_b - \max\{\min(p_G); R\}) \leq M \quad (3)$$

with $D_b(0) = M$ and $\partial D_b(\cdot)/\partial(\cdot) \leq 0$.

In words, when the branded costs as much as the cheapest generic ($p_b = \min(p_G)$) or is free for consumers ($p_b \leq R$) the branded firm captures all the demand for the drug. On the other hand, when it is more expensive, demand is lower the higher the difference between the branded price and the highest between $\min(p_G)$ and R. Note that we consider only the value $\min(p_G)$ as a reference for the demand of the branded. This is a consequence of the assumption of infinite cross-price elasticity among generics, which implies that the only relevant price for consumers' choice is the minimum.

The aggregated demand for generics is then defined residually as $D_G = M - D_b$. When $D_G > 0$, the demand for a single generic firm i can then be positive only in two cases. First, when the drug is always free for consumers ($\Delta_{ig} \leq 0$). Second, when $\min(p_G) = p_{ig} > R$, i.e. drug i, although not free, is the cheapest option available. Formally:

$$D_{ig} = \begin{cases} \dfrac{D_G}{j} & \text{if } \ (\min(p_G) \leq p_{ig} \leq R \text{ or } \min(p_G) = p_{ig} > R) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $j = \#\{i : (p_{ig} \leq R) \wedge (p_{ig} = \min(p_G)\}$.

2.5. The profit function

Firms face a constant marginal cost c. At any point in time, profits are given by:

$$\pi_z = (p_z - c)D_z \quad (5)$$

where z can be a branded or a generic firm. The profit function is well-behaved in the sense that it is twice differentiable with respect to p_z and with $\partial^2\pi_z/\partial^2 p_z < 0$.

Finally, we define δ as the discount factor of future profits, common to all firms.

3. Stage 2: optimal prices for a given *R*

At this stage firms take R as exogenous. They then price-maximize their profits. Defining generally $\varepsilon_{zj} = (D_z/p_j)^{-1} \cdot (\partial D_z/\partial p_j)$ as the price elasticity of z with respect to price j, we can use standard maximization to find stage 2 equilibrium prices.

Lemma 1. *The stage 2 Pure Strategy Nash Equilibrium (PSNE) with free prices, perfect competition among generics and $n \geq 2$ is:*

$$p_b^{**} = \begin{cases} \dfrac{c}{1 + (\varepsilon_{bb})^{-1}} & \text{if } \ \dfrac{c}{1 + (\varepsilon_{bb})^{-1}} \geq R \\ R & \text{otherwise} \end{cases} \quad (6)$$

and

$$p_{ig}^{**} = R \quad (7)$$

Proof. See Appendix A. □

[10] Co-payments in the form of fixed-fees are however consistent with our design. The relation between RPS and co-payments will be discussed in Section 7.

Generics will only be priced *at R*. Concerning the branded, firm prices according to the classical Lerner index. However, R represents a low threshold below which it is not profitable to price (see proof). This result is another way to express the kinked demand approach of Danzon and Liu (1997).

Finally, notice that stage 2 pricing is basically independent from either the shape or the updating process of the RPS. As a consequence, the result of Lemma 2 is general and does not make use of any assumption on R.

4. Stage 1: setting *R*

At this stage firms maximize their profits and indirectly set R through their pricing strategies. In this section we provide an analysis of the equilibrium prices under the two opposite scenarios of non-cooperative and cooperative generic firms.

4.1. Perfect competition among generics

The case of perfect competition can be modeled as a one-shot game. Firms choose their prices in order to maximize their profits and R is simultaneously derived as a function of the existing prices. This scenario should set the benchmark for a RPS, since generics are competing *a la Bertrand*.[11]

While examining the PSNE of this game, it turns out that the shape of the R function can make a difference. In particular, the existence of a PSNE depends on whether the RPS exhibits Brand-Independence or not.

Lemma 2. *In a simultaneous RPS with free pricing and $n \geq 2$ a PSNE exists if and only if BI holds. In this case, the equilibrium prices are given by (ignoring apex for stage 1):*

$$\begin{cases} p_b^* = \dfrac{c}{1 + (\varepsilon_{bb})^{-1}} \\ p_{ig}^* = c \quad \forall i \end{cases} \quad (8)$$

Consequently, $R = c$.

If R does not exhibit BI, $R = c$ can never be an equilibrium.

Proof. See Appendix A. □

This raises the following points. First, under BI, generics end up pricing at the marginal cost because no other price can prevent rivals from undercutting the reimbursement level. Second, the price of the branded drug is related to the elasticity of the demand and is influenced by Δ_b and by the level of R. Finally, even under this ideal situation of perfect knowledge and infinite price elasticity, R will be equal to the marginal cost *only* if BI holds.

4.2. Collusion among generics

In the case of stage 1 collusion the equilibrium of the game depends on the type of competition between the generics and the branded firm.[12]

This section provides some results for the case of price competition between the branded and the generics' cartel. We focus on this setting for three main reasons. First of all, it is simpler to analyze for

[11] Because generics' demand is defined residually, a Stacklberg type of equilibrium could also suit this setting. Results, however, do not change because generics reaction functions are determined by their internal competition, not by the branded pricing strategy.

[12] For example, a Stackleberg competition where the branded firm moves first can be considered, as in Merino (2003). Alternatively, it is possible to introduce further collusion between the generic cartel and the branded producer, like in Miraldo (2009).

general functional forms. Second, the Nash Equilibrium (NE) usually represents the most natural base case scenario. Third, the focus of the analysis is on generics' collusion only. Further complications are thus not needed.

In choosing the right price in stage 1, firms must take into account its effects on stage 2 profits through R. For notational purposes, it is useful to introduce the quantity:

$$\bar{\pi}_z = \frac{\partial \pi_z^2}{\partial p_z^1} \tag{9}$$

which is the partial derivative of the stage 2 profits with respect to own stage 1 prices. For interpreting the results, it also useful to know the following lemma.

Lemma 3. *When BI holds, $\bar{\pi}_b = 0$. When BI does not hold, $\bar{\pi}_b > 0$. The sign of $\bar{\pi}_G$ is ambiguous.*

Proof. See Appendix A. □

We then have:

Lemma 4. *A stage 1 finite NE exists only if $(\partial \Delta_b^{**}/\partial R) < 0$, or $((\partial p_b^{**}/\partial R) - 1) < 0$. In this case the mark-ups are given by:*

$$\begin{cases} p_b^* - c = -\left[\left(1 - \dfrac{\partial R}{\partial p_b^1}\right) \dfrac{\partial D_b}{\partial \Delta_b}\right]^{-1} \left(D_b^1 + \bar{\pi}_b \displaystyle\sum_1^T \delta^t\right) \\ p_G^* - c = \left[\dfrac{\partial R}{\partial p_G^1} \dfrac{\partial D_G}{\partial \Delta_b}\right]^{-1} \left(D_G^1 + \bar{\pi}_G \displaystyle\sum_1^T \delta^t\right) \end{cases} \tag{10}$$

Proof. See Appendix A. □

The first relevant result from considering Lemmas 3 and 4 together is that when BI holds stage 1 prices are at their minimum. This proves that under BI prices (hence, reimbursement) are always lower than when BI does not hold.

In case stage 2 did not exist (i.e. the only relevant period is between $t = 0$ and $t = 1$), Lemma 4 gives the result of a one-shot NE when b and G are competing on prices and R is endogenous. Introducing the second stage forces firm to behave more strategically by taking the effects on future profits into account. For the branded, this implies that when BI does not hold, since $\bar{\pi}_b > 0$, stage 1 price increases with respect to the one-shot game. Note that, all else equal, prices are also higher for higher values of δ and for lower elasticities.

Regarding generics, these relations are less clearcut. As shown in Lemma 3, $\bar{\pi}_G$ has no clear sign. On the one hand, an increase in p_G^1, by increasing R, increases stage 2 per-unit profits. This can be referred to as the "price effect". On the other hand, a higher R will reduce Δ_b^{**}, hence lowering stage 2 demand for generics. This can be called "demand effect". When the latter prevails on the former, stage 1 prices for generics are lower when stage 2 is considered than in a situation where stage 2 does not exist.

The condition $\partial \Delta_b^{**}/\partial R < 0$ is what prevents R from "exploding". It can be explained as follows. First, note that profits of the branded firm are always monotonically increasing with R. Consequently, the branded has an interest in an infinitely high reimbursement level. The generics, however, must trade off price and demand effects on stage 2 profits. If an increase in R is associated to a lower Δ_b^{**}, i.e. if the distance between the optimal branded price and the reimbursement level decreases when R increases, the residual demand for generics is reduced. The incentives for an infinitely high level of R are thus counterbalanced and R is prevented from taking infinite values even when generics can collude.

5. Maintaining the agreements

Since we have shown that competitive and collusive equilibria can both exist in a two-stage RPS with n generic firms, this raises the question about what conditions must hold for a cartel among generic firms to exist and be stable. In this section, we try to provide some reasonable answers. In other words, we are looking inside the black box of the single generic firm usually assumed in the existing theoretical literature, searching for the founding conditions of the implicit assumption of a single representative generic firm.

Imagine that a generic agreement exists. In the following we will identify the conditions for it to be broken by a single cheating firm. The analysis is divided in three steps. First, we identify the best cheating strategy for the betraying firm. In the classical theory of collusion with perfectly elastic demand, firms undercut the common price by small amount. In this setting, however, it is shown that the price needed to undercut the rivals is different and depends crucially on the shape of R. Second, explicit conditions for collusion among generics are provided. Finally, the simple setting is extended to take into account that firms do not meet only for T periods. If the two-stage game is repeated, firms might foresee the future impact of their strategies and change the conditions for their cooperation. Results are then provided for what we call the infinitely repeated two-stage game.

Before proceeding some further notation is required.[13] First of all, we start from perfect collusion and define θ as the collusive price, i.e. $p_G = \theta$. The related level of R is $\bar{R} = R(p_b^*, \theta)$ (hereafter $\bar{R}$). Under this scenario, each generic firm's profit is $\pi(\theta)/n = (\theta - c) \cdot (D_G(p_b^* - \bar{R})/n)$ in the first stage, followed by $\pi(\bar{R})/n = \bar{R} \cdot (D_G(p_b^{**} - \bar{R})/n)$ for the next T periods (Lemma 1).

A firm deciding to break the agreement in stage 1 sets a price $\widetilde{p_i}$. This represents the maximum cheating price a firm can set in order to break the agreement and capture all the generics' demand. A more precise definition of this quantity is provided in the next subsection. If one firm cheats, the new level of R is $\tilde{R} \equiv R(p_b^*, p_{-ig} = \theta, \widetilde{p_i})$ (hereafter $\tilde{R}$). The cheating price then defines profits $\pi(\widetilde{p_i})$ for the first period. In the second stage, since the new R is $\tilde{R}$, each firm's profit will be $\pi(\tilde{R})/n = \tilde{R} \cdot (D_G(p_b^{**} - \tilde{R})/n)$.[14]

5.1. Breaking the cartel: the cheating price

For cheating to be profitable, a generic firm must capture all the (residual) demand. Hence, it must set a price in such a way that its product is cheaper than other generic drugs for consumers. This implies that price must drop to a level that R becomes lower than the collusive price. Depending on the shape of R, this cheating price can vary substantially.

To formalize these ideas, given the notation presented above, the maximum cheating price can be defined as:

$$\widetilde{p_i} \equiv \lim_{\varepsilon \to 0} [p_{ig} : \tilde{R} = \theta - \varepsilon] \tag{11}$$

That is, $\widetilde{p_i}$ is the maximum price that generic firm i should set in order to drop R below θ.

Then, it is useful to better define the relations between all the involved variables making the usual distinction between RPS with and without BI:

Lemma 5. *If BI does not hold, $\widetilde{p_i} < \theta < \bar{R}$. Also, $\tilde{R} \cong \theta < \bar{R}$ and $\pi(\bar{R}) > \pi(\theta)$. If BI holds, $\widetilde{p_i} \cong \theta = \bar{R}$. Also, $\tilde{R} \cong \theta = \bar{R}$ and $\pi(\bar{R}) = \pi(\theta)$.*

[13] In the following, we will generally ignore the superscript for the stages, with no loss of generality.

[14] Note that when the game is played only once, there is no room for punishment. The level of R in the second stage is R and firms cannot do anything to change it.

Proof. See Appendix A. □

Lemma 5 defines also the relations between different levels of R and profits. This information will be used in the next subsections. Here it is important to underline that the cheating price in this setting depends on the shape of R. Intuitively, if R depends on the branded price, all else equal, it will be higher than θ. In order to decrease R enough to undercut θ, firm i must set a particularly low price. This will obviously have serious consequences on the sustainability of an agreement.

In order to better clarify the definition of cheating price and the relation among all the defined variables, a simple example is given here. Imagine that R is linear with a weight d on the branded price and $(1-d)/n$ on each of the n generic prices. Then:

$$R = dp_b + \frac{(n-1)(1-d)}{n}\theta + \frac{(1-d)}{n}p_{ig} \tag{12}$$

From the definition, $\widetilde{p_i}$ is the level of p_{ig} at which $R=\theta$. After some straightforward manipulations we have:

$$\tilde{p}_i = \frac{dn(\theta - p_b) + \theta}{(1-d)} \tag{13}$$

The cheating price is thus lower the higher the price of the branded, the higher the weight given to the branded and the more the generics in the market.

5.2. Conditions for collusion

Noting that the two-stage game ends at T, it is then possible to prove the following lemma.

Lemma 6. *Stage 1 collusion among generics requires that*

$$\frac{\delta(1-\delta^T)}{(1-\delta)} > \frac{n\pi(\tilde{p}) - \pi(\theta)}{\pi(\bar{R}) - \pi(\tilde{R})} \tag{14}$$

Proof. See Appendix A. □

Eq. (14) can be written to directly compare the gains (lhs) and losses (rhs) from betrayal:

$$\pi(\tilde{p}) - \frac{\pi(\theta)}{n} < \frac{\delta(1-\delta^T)}{(1-\delta)}\frac{\pi(\bar{R}) - \pi(\tilde{R})}{n} \tag{15}$$

As in the standard approach to collusion, gains come from excluding competitors in the first stage of the game. The profit made in this case is $\pi(\tilde{p})$, which is then compared with the profit made by following the agreement, i.e. $\pi(\theta)/n$. The associated costs from cheating are due to the reduction in the reference level from $\bar{R}$ to $\tilde{R}$. Note however that when BI holds Lemma 5 states that $\widetilde{p_i} \simeq \theta$ and $\bar{R} \simeq \tilde{R}$. Therefore, the gains from betrayal are high ($\widetilde{p_i}$ is at the maximum possible level), while the costs are null (the reduction in the reimbursement level is close to 0). In other words, in this two-stage game, collusion can *never* be sustained when BI holds. Although this specific result is not robust to changes in the structure of the game, the main lesson to be drawn here is that BI helps competition by facilitating the breaking of any agreement among generic firms.

Other two characteristics of Eq. (15) are worth mentioning. First of all collusion is less likely for larger n and smaller δ. This finding is in line with the classical conclusions from the general literature on collusion. Second, collusion is more easily sustainable for larger Ts. This result follows from $\bar{R} \geq \tilde{R}$ (Lemma 5). By extending T, the loss from cheating as opposed to collusion is also extended, making it a less likely strategy to be pursued. This finding, nevertheless, although straightforward, is intrinsically related to the nature of the game considered and it is not robust to a change in its structure (as will be shown in the next subsection).[15]

5.3. The infinitely repeated two-stage game

In this section the previous setting is extended by allowing the RPS two-stage game to be repeated an infinite number of times. This framework is suitable to a situation where firms have time horizons that go beyond the updating period.

Here the punishment works similarly to the classical theory of collusion: if one firm breaks the agreement, the others will play a grim trigger strategy where prices are dropped at marginal costs forever (Friedman, 1971).[16] The first point to note is that, even if firms price at c, the resulting R under punishment (R_p), can be greater than c. Indeed, when BI holds $R_P = R(p_{-ig} = c, p_{ig}) \geq c$. $R_P = c$ in two cases. First, when $p_{ig} = c$. Second, if R does not depend on *all* generics' prices.[17] On the contrary, if R_p is a function of *all* prices, then $R_p > c$ if and only if $p_{ig} > c$. Theoretically, if p_{ig} can increase without limits, the cheater can always set a price such that $R_P = \bar{R}$. More generally, if p_{ig} can influence R_p with no constraints, the profits after T for the cheating firm are equal to 0 in each updating period (because in stage 1 $p_{ig} > R_P$) and $\pi(R_P)/n$ in each stage 2 period. In order to facilitate the analysis, however, in the following we make the useful assumption that $R_P = c$.[18]

The general condition for collusion is given by the following lemma.

Lemma 7. *In a two-stage infinitely repeated game, generics collude if:*

$$n < \frac{(1/(1-\delta^{T+1}))(\pi(\theta) - \pi(\bar{R})) + 1/(1-\delta)\pi(\bar{R}) - ((\delta(1-\delta^T)/(1-\delta))\pi(\tilde{R})}{\pi(\tilde{p})} \tag{16}$$

Proof. See Appendix A. □

In order to understand this result, it is useful to consider first the case of a RPS with BI (where $\theta = \bar{R} = \tilde{R} = \tilde{p}$). In this case, Eq. (16) reduces to

$$n < \frac{1 - \delta(1-\delta^T)}{1-\delta} \tag{17}$$

which implies that, when BI holds, collusion is possible and is more likely for high δ, low n and short T. Hence, if the game is repeated infinitely collusion is possible even under BI. In addition, there is an inverse relation between the length of the updating period (T) and the likelihood of collusion. This is because a higher T is associated with a longer "protection" from punishment. Postponing the zero profits period increases the discounted flow of profits for a cheater, thus increasing the incentives to break the cartel.

When BI does not hold, the relation between collusion and either n or δ does not change. The role of T, on the other hand, is ambiguous. The following corollary helps.

Corollary 8. *In a two-stage infinitely repeated game a marginal increase in T reduces the sustainability of the cartel among generics if:*

$$(1-\delta)(\pi(\bar{R}) - \pi(\theta)) < (1-\delta^{T+1})^2\pi(\tilde{R}) \tag{18}$$

Proof. See Appendix A. □

---

[15] Since higher Ts are *de facto* associated to longer time horizons by firms, in the one-shot game T actually serves two purposes. Its meaning in this context is thus more ambiguous than in the cases of an infinitely repeated game.

[16] In this section we will concentrate on the incentives to collude only, as the equilibrium analysis of prices does not change when the game is repeated.

[17] For example, when R is defined as the minimum price in the market.

[18] It is possible to show that the basic results shown for $R_P = c$ hold for $R_P > c$.

Intuitively, when T is increased by 1 unit, a colluding firm gains from postponing the "price setting" moment (because $\theta < \bar{R}$). A cheating firm, on the other hand, will gain from the delayed punishment. In addition, because $\bar{R} > \tilde{R}$, postponing the updating moment implies a loss for the cheating firm with respect to the profits under collusion. Note, however, that under the reasonable assumption that $\pi(\bar{R}) - \pi(\theta) < \pi(\tilde{R})$ the inequality in (18) is *always* satisfied. In this case, by increasing T authorities can reduce the incentives to collude for generics.

6. An example: linear demand and linear RPS

Assume collusion among generics. This is like assuming that there are just one originator and one generic representative good. The choice of a patient is between the branded and its generic version. All the M consumers are the same in terms of health benefits from the drug, which we denote with H. The utility associated to this health gain is $V(H)$ and is identical across patients. Although drugs are homogeneous, not all patients believe that generics can be effective and safe enough to substitute the branded. Here we introduce the quantity P_m, which is the subjective probability that patient m assigns to a generic drug giving the expected health benefit. The utility functions are quasi-linear. Denoting income with I, the utility function of consumer m from buying a branded product can be written as:

$$U_b(H, I_m, \Delta_b = p_b - R) = V(H) + I_m - \Delta_b \tag{19}$$

which must be compared with the utility from buying a generic drug. In equilibrium we know that $p_G \leq R$, hence:

$$U_G(H, I_m, \Delta_b = p_b - R, P_m) = P_m V(H) + I_m \tag{20}$$

Remembering that M is the number of people who got the drug prescribed, market demand will be defined by (here F is the distribution function):

$$D_b = MF\left(1 - \frac{\Delta_b}{V(H)}\right) \tag{21}$$

for the branded and

$$D_G = M\left(1 - F\left(1 - \frac{\Delta_b}{V(H)}\right)\right) \tag{22}$$

for the generic.

Different distributions will give rise to different demands. One important case is the uniform distribution.[19] In this situation F is linear with respect to Δ_b and $D_b = M(1 - (\Delta_b/V(H)))$. That is, D_b is a linear demand function with intercept (α) equal to M and price coefficient (γ) equal to $M/V(H)$. This implies that the demand is less price sensitive for drugs with higher health impact. In other words, if they expect important benefits from the drugs, consumers are less prone to take risks and buy products they do not trust entirely.

Then:

$$D_b = \alpha - \gamma(p_b - R) \tag{23}$$

and

$$D_G = \gamma(p_b - R) \tag{24}$$

With this demand functions,[20] Lemma 1 can be directly applied. Stage 2 optimal prices are thus easily found. For generics, price is always equal to R. For the branded, Lemma 1 gives:

$$p_b^{**} = \begin{cases} \dfrac{1}{2}\left(\dfrac{\alpha}{\gamma} + R + c\right) & \text{if } \left(\dfrac{\alpha}{\gamma} + c\right) \geq R \\ R & \text{if } \left(\dfrac{\alpha}{\gamma} + c\right) < R \end{cases} \tag{25}$$

In stage 1, firms can influence the level of R. Here we assume a general RPS of the form:

$$R(p_b, p_g) = \sum_{i=1}^{n} w_i p_{ig} + dp_b \tag{26}$$

where $0 \leq d \leq 1$ is the weight given to the branded firm (see Section 5.1). BI then requires that $d = 0$. Applying the previous framework to this specific example, it is possible to find both the equilibrium prices and the conditions for collusion.

6.1. Equilibrium prices

If generics compete, Lemma 2 can be directly applied, implying that $R^* = c$. In case of collusion, the equilibrium price is given by the solution to (34) and (35) with $R \equiv wp_G + dp_b$ and $w + d = 1$. The algebraic solution is given in Appendix B. In Figs. 2 and 3 we show the behavior of the equilibrium prices for different levels of w (i.e. for different weights of the branded drug in the definition of R) and for given numerical values of α and γ. In the figures patient firms refer to a scenario with $\delta = 0.8$, as opposed to the $\delta = 0.2$ of impatient (no forward looking) firms. The relevant demand parameters for the equilibrium are fixed at $\alpha = 10{,}000$ and γ either 50 (inelastic demand) or 200 (elastic demand).[21] Finally, without loss of generality assume that $c = 0$.

From Fig. 2, all the quantities behave as expected. As predicted, the higher the role of generics in setting R (higher w), the lower the value of R. Nonetheless, it is clear that the cost of collusion is always significant (remember that under competition, $R = 0$), and it grows significantly for a less elastic demand (first versus second row). The discount factor, on the other hand, does not play a particularly important role, mainly because the second stage of the game is only one period long.

Moving to Fig. 3, T increases to 10. Demand elasticity and the value of w have the same impact as before. Price levels, however, are on average higher than in the case of $T = 1$. This is expected since firms now are more willing to trade off actual profits for future (longer) gains. In doing so, they increase the reimbursement level. The role of the discount factor this time is more evident. Specifically, when $\delta = 0.8$ the optimal stage 1 price for the generic cartel is increasing, rather than decreasing, in w. Once again, this is the consequence of the trade off between today and tomorrow profits in setting R.

This example shows that strategic pricing under RPS is important. The generic agreement is constantly under the pressure of two opposite forces. On the one hand, the firms have an incentive to obtain a higher R; on the other hand, they do not want the branded to be too close to the reimbursement level. The balance between these two forces determines the equilibrium. Importantly, both

[19] It is useful to note that the uniform distribution is a special case of a Beta distribution. The Beta distribution suites well this example because its values are bounded between 0 and 1, exactly like the random element of the utility function, P_i.

[20] Note that the intercept of (24) is set to zero because the residual demand for generics is null when the branded product is free.

[21] This scenario implies that when the branded firm prices at R, 10,000 people will buy its product. Note that these numbers are just used for the simulations. They are not a tentative to calibrate the model. Indeed, we know nothing about the actual price sensitivity of consumers when they face the choice between branded and generic in an RPS system.

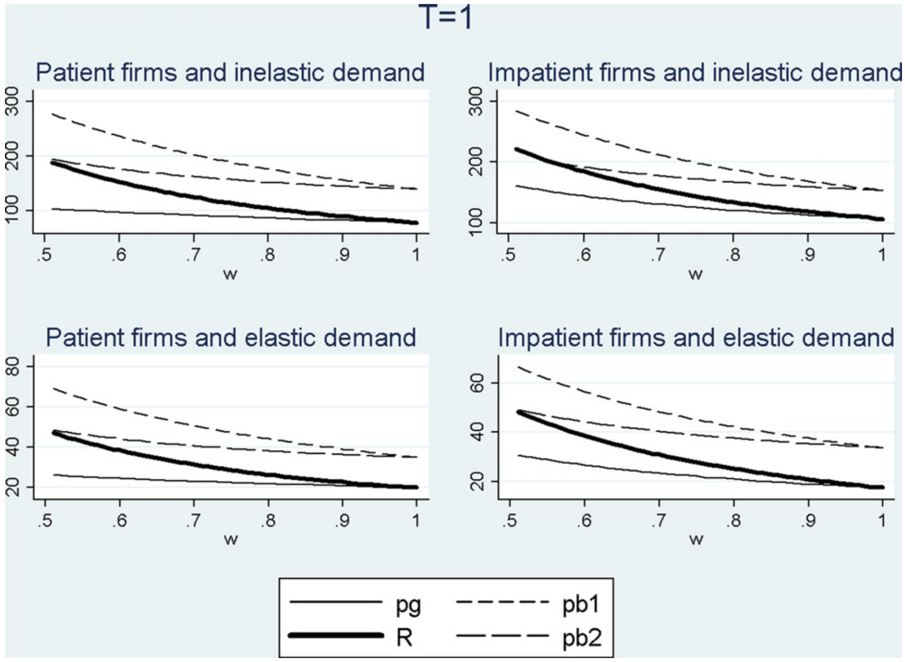

Fig. 2. Equilibrium prices with $T = 1$. Patient firm: $\delta = 0.8$, impatient firm $\delta = 0.2$. Elastic demand: $\gamma = 200$, inelastic demand: $\gamma = 50$.

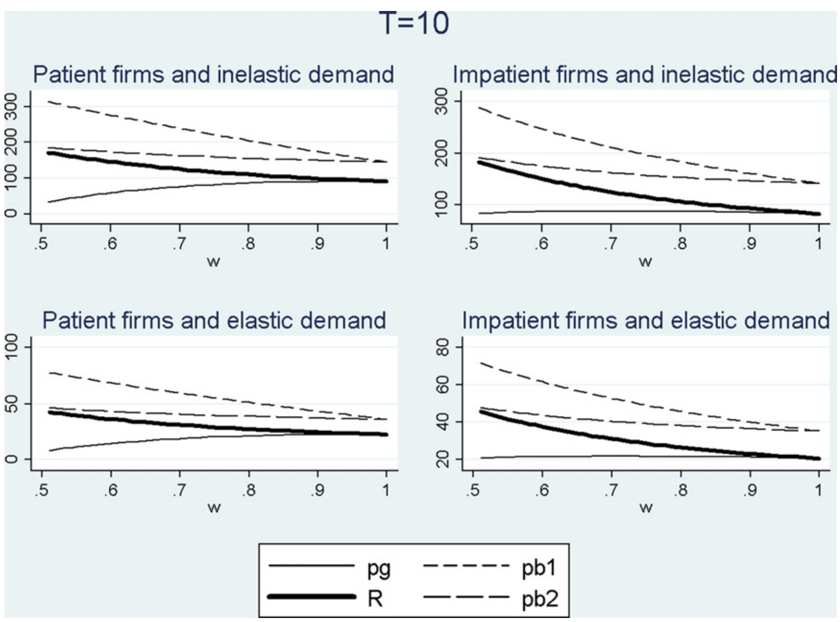

Fig. 3. Equilibrium prices with $T = 10$. Patient firm: $\delta = 0.8$, impatient firm $\delta = 0.2$. Elastic demand: $\gamma = 200$, inelastic demand: $\gamma = 50$.

the length of updating and relative weight of the branded affect significantly the optimal pricing strategies.

6.2. Maintaining the agreements

Defection requires that the present value of the profits from breaking the agreement is greater than the present value from maintaining it. In the classical economic theory collusion is easier for fewer firms and for higher discount rates. When prices can change freely (without RPS) and punishment is in the form of zero profits forever, it is well-known that $n < 1/(1 - \delta)$ is the condition for collusion (e.g. Tirole, 1988).

When an RPS is introduced, relations are more complex. If BI holds, collusion is possible only when the two-stage game is infinitely repeated. In that case, the condition for the existence of a cartel among generics is given by (17). When BI does not hold, the condition can only be seen through simulation. Here it is worth reducing the number of variables by assuming that R is the arithmetic unweighted average of all the existing prices. That is, $w_i = d = 1/(n+1)$. Formulas for the main variables are then given in Appendix B.

The relation between the number of firms and the discount rate is depicted in Fig. 4. The three cases of no RPS (simple case of free pricing, n firms and zero profits punishment), infinitely repeated RPS with BI, and two-stage RPS with no BI (average of

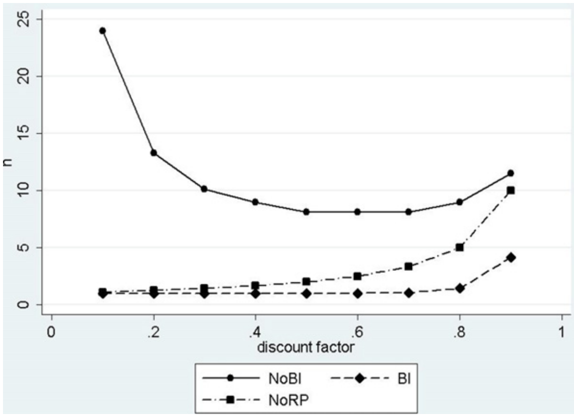

Fig. 4. Threshold number of firms for collusion.

all existing prices) are reported. The curves represent the relation between the discount factor and the threshold number of generic firms for collusion to hold. For a given discount rate, if the actual value of n is lower than the threshold the agreement can be maintained.

The figure shows two important points. First, for any given level of δ, the threshold is lower in the case of RPS with BI than in the classic case of free pricing. In other words, when BI holds collusion among firms is more difficult under an RPS. This is a general result that does not depend on the shape of either the demand or the profit functions.

Second, the threshold number of firms does not increase monotonically with the discount factor. Specifically, for a low discount factor, collusion seems to be facilitated. This happens because when firms do not consider the future, the first period optimal branded price is higher. This implies that $\widetilde{p_i}$ must be very low, to the point that it is often not worth cheating. An increase in δ reduces the "pulling" effect of the branded price. For sufficiently high discount factors, thus, the increasing shape of the curve is recovered.

7. Discussion

7.1. Reference price and co-payment

The present analysis focuses on the case of RPS only. The purpose of this theoretical approach is exactly to isolate the impact of one specific regulatory tool on market behaviors. In the real world, however, the RPS is applied together with other regulatory measures, of which co-payment is probably the most common. Since the model is sufficiently flexible to accommodate this new scenario, it is interesting to understand what would change if consumers had to pay an extra-charge for their pharmaceutical product.

When co-payments are in the form of fixed-fee results in general should not change. Since consumers are not sensitive to the actual price of the drug, the relations among firms are not affected.[22] When patients have to pay a percentage of the price of the drug (p_C), then the relevant pricing function becomes (e.g. Brekke et al., 2007; Miraldo, 2009):

$$p_C = \begin{cases} ap & \text{if } \ p < R \\ ap + (p - R) & \text{if } \ p \geq R \end{cases} \tag{27}$$

[22] It should also be noted that in some countries, e.g. Italy, price-percentage co-payments represent a very unpopular measure that governments typically try to avoid.

where a represents the co-payment rate and p the price of the drug.

Under the two opposite scenarios of perfect competition and collusion, the analysis of equilibria would not change much. Lemma 2 directly applies when generics compete with each other. Equilibrium (under the *assumption* that generics can maintain a collusive agreement throughout both stages) could be found by substituting the new price structure in the equations of the game. No substantial modifications would need to be introduced in order to find the new solution.

More interestingly, when co-payments are introduced in a setting with n firms and infinite cross-price elasticities, the conditions for an agreement change substantially. Two main differences must be underlined:

- With co-payments, the incentive to undercut exists even when R is exogenously given (stage 2). With no further agreements, generics would then compete *a' la Bertrand*. If generics compete in stage 2, equilibrium implies $p_{ig} = c$ for all i. If the game is played only once, no stage 2 collusive agreement can be maintained. Hence, the competitive equilibrium is the only one possible.
- In an infinitely repeated two-stage game, collusion is still possible. Actually, since firms can drop prices at marginal cost from $t = 1$ onwards, punishment can be more effective than without co-payments. Intuitively, conditions for both stage 2 and stage 1 collusion would not differ from the classical textbook case (i.e. $\delta > (1 - 1/n)$).

Hence, the RPS does not introduce any particular incentive to collude once price-percentage co-payments are introduced. This conclusion shows two important points. First, the framework provided in the paper is particularly easy to extend to different regulatory scenarios. Second, the assumptions of an endogenous R higher than c with price-related co-payments and of one representative generic firm should both be more carefully justified by the theoretical literature. Indeed, this outcome is only possible if collusion is at work among generics. In these cases R is probably better understood as a focal point for possible collusive agreements.

7.2. Policy implications

The most immediate policy implication refers to the shape of R. The reimbursement price should be set to depend only on the perfectly elastic side of the market, for two reasons. First, R is higher when it depends on the branded price, even when generics actually compete with each other. Second, collusion is facilitated by a level of R higher than the cartel price. An optimal policy in this sense could then be to set R as the average of the prices associated to all the perfectly substitutable generic products. Alternatively, if one is worried about the assumption of perfect substitutability among generic products (i.e. some "branded" generics might rely on a certain degree of brand loyalty), R could be set as the minimum existing price in the market.

Another implication of the model is related to the responsiveness of the RPS to a change in the market structure for generics. It has often been commented that RPSs do not respond well to the competitive pressure from generic firms (e.g. Lopez-Casanovas and Puig-Junoy, 2000; Danzon and Ketcham, 2003). This is probably the main reason why the theoretical literature models R as exogenous to pricing decisions.

Indeed, the lack of sensitivity to generics' competition represents a serious concern and might jeopardize the long-run effectiveness of a RPS. For example, the RPS needs to be flexible in order to adjust to the initial post-expiry variability in pricing behavior. Typically the number of generic firms in a market increase with time and competition is appropriately exploited only after the first adjustment period. Often, the level of R does not respond to these natural changes in the market structure so the benefits from increased competition can never be properly captured. In addition, when more competition does not lead to lower reimbursement prices, firms must look for other ways to gain market shares. For example, in some countries generics end up competing on pharmacists' discounts (i.e. selling underpriced to pharmacists and leaving the retail price untouched) rather than on prices.[23] The rigidity of the RPS thus *de facto* translates into a redistribution of the competition-related welfare gains from the public sector to the retailers.

Our approach can easily capture the lack of sensitivity of the RPS to the pressure from generics' competition. In a RPS, a lower price, causing a lower R, would immediately be detected. If prices can react flexibly, then $\pi(\tilde{p}) - \pi(\theta) \simeq 0$ and collusion is immensely facilitated. Hence, no generic firm will ever undercut the existing reimbursement price.

Furthermore, note that past values of reimbursement could represent the most natural focal points for coordination.[24] If collusion is possible and p_G^* is realistically unknown, the existing reimbursement level represents the most natural value generics would agree on. This natural strong attraction to the *status quo ante* explains the long-run stickiness in R levels noted in the literature (Galizzi et al., 2011).

According to the model presented, however, this rigidity is not intrinsic to the RPS. Rather, it represents a consequence of a sub-optimal design which sets the wrong incentives to incumbent firms. The structure of the RPS could be changed accordingly. As the level of R must remain public knowledge (this is the nature of the RPS), one important policy suggestion would be to increase the time needed for price reaction. Formally, this would reduce $1/T$ and would then be similar to an increase in T. However, the ways for boosting competition here are substantially different, since increasing stage 1 would essentially make $\pi(\tilde{p}) - \pi(\theta)$ positive, introducing some serious incentive to break collusive agreements. In this sense authorities can, for example, keep the same RPS structure, freezing prices for a period after R has been updated.

There are other gains from viewing R as the result of a dynamic game among generic firms. Standard results from the literature on dynamic competition could be easily applied. For example, one could observe that in most countries the ratio between the number of generics firms and the number of available markets is usually small. Multimarket contacts, representing an important source of collusion (see Bernheim and Whinston, 1990; Evans and Kessides, 1994), might be a relevant topic also for pharmaceutical markets under RPS. More in general, our model suggests that authorities should focus their attention, among other things, on the existence of potential sources of generics' collusive agreements.

8. Conclusion

Generic drugs are homogeneous products competing with infinite elastic demands. There are thus good reasons to think that pharmaceutical regulators should exploit these characteristics in order to obtain reimbursement prices that could reflect marginal production costs. The presented article examines the extent to which these expectations can be fulfilled by a regulation based on the RPS.

Results show that a well-working RPS can be effective only if the market for generics is competitive. On the opposite, collusion represents a serious problem and can be facilitated by a wrongly designed RPS. More specifically, it is shown that competition and collusion coexist: firms can coordinate on a level of R in the first stage and then compete taking R as given in the second stage. The equilibrium under this game is always greater than the marginal costs (although never infinitely large), implying a welfare loss for society.

If collusion represents an issue, the RPS can be modified accordingly. Rather than looking for other price control mechanisms not based on competition forces (e.g. unilateral reductions in the level of reimbursement), regulators could indeed break the described collusive patterns and make competition more profitable for incumbents and potential entrants.

Given the novelty of the approach, technical complications have been oversimplified. The main limits to the model presented come from these oversimplifications. Among them, probably the most relevant one is that the number of generics is not endogenously determined in the model. Clearly, however, generic entry might influence and in turn be influenced by the RPS. For example, our study shows that even if generics compete, their demand is determined by the branded price elasticity. If elasticity is high, Δ_b is small and the residual market size for generics is low. Even without providing a formal analysis of entry, it is clear that a combination of price-sensitive patients and RPS can suffocate the market for generics, raising concerns over the long-run sustainability of an RPS. On the other hand, our analysis points to a substantially positive relation between T and competition. When BI holds, for example, authorities do better by not updating R, because for $T \to \infty$ firms never collude. This is clearly unrealistic. The issue here is that firms' behavior and regulatory outcomes might depend on changing factors considered as static or exogenous in the model. In this sense, entry explains why authorities might do better by updating R more frequently. Further analysis is thus needed on this issue.

Acknowledgements

I would like to thank Howard Smith, Simon Cowan, Alistair McGuire, Andrea Patacconi and two anonymous referees for the very useful comments. I am also indebted to Livio Garattini for the many policy insights. All the usual disclaimers apply.

Appendix A. Proofs

Lemma 1

All quantities here refer to the second stage.

First, we prove that an exogenous R introduces a kink in the demand of both types of products. First we note that, for any type of product z, it is always true that $D_z(p_z < R) = D_z(p_z = R)$, since the drug is free to patients anyway. Hence, it is straightforward to understand that $\pi_b(p_b = R) > \pi_b(p_b < R)$ and $\pi_{ig}(p_{ig} = R) > \pi_{ig}(p_{ig} < R)$. For the generics, infinite elasticity with $n > 1$ generic firms competing among themselves implies that in equilibrium $\pi_{ig}(p_{ig} > R) = 0$. Hence, $\pi_{ig}(p_{ig} < R) \geq \pi_{ig}(p_{ig} > R)$. This means that generic firms can never do better than setting price *at R*, which proves that $p_{ig}^{**} = R$.

[23] This is the case, for example, of pre-2004 Spain, as reported by Puig-Junoy (2004). Another example, Italy, has institutionalized extra-discounts by imposing an 8% increase on pharmacists margins when generics are dispensed (Law 79, 2009) This approach of course might help the competition between generics and branded drugs, but does not affect intra-generics competition.

[24] "More impressive, perhaps, is the remarkable frequency with which long negotiations over complicated quantitative formulas or *ad hoc* shares in some costs or benefits converge ultimately on something as crudely simple as [...] the shares agreed on in some previous but logically irrelevant negotiation. Precedent seems to exercise an influence that greatly exceeds its logical importance or legal force" (Schelling, 1960).

When the branded is priced above R, the optimal price is given by solving the following:

$$\max_{p_b} \ \pi_b = (p_b - c)D_b(p_b - R) \tag{28}$$

Given the definition of own price elasticity, this represents the classical monopolistic maximization problem, whose well known solution is the Lerner index for mark-ups. Rearranging so to have the optimal second stage in the left hand side gives the result.

Lemma 2

The proof is for $n = 2$ and is organized in steps. All quantities are for the first stage.

1. In any equilibrium it must be that

$$p_{ig} = R \tag{29}$$

If $p_{1g} > R$, then $D_{1g} = 0$. If $p_{1g} > R$ the firm can always do better by increasing its price. Hence $p_{1g} \neq R$ for $i = 1, 2$ is not an equilibrium.

2. If BI holds, then a PSNE exists and is defined as in Lemma 2. If both generics set the same price, define this price as $\overline{p_g}$. In equilibrium under BI:

$$R(p_b, \overline{p_g}) = \overline{p_g} \tag{30}$$

However, not any level of R is an equilibrium. By BI, any decrease in p_{ig} reduces R, cutting firm j out of the market. So, $p_{ig} < p_{jg}$ is not an equilibrium. $p_{ig} = p_{jg} = R > c$ is not an equilibrium either, because firms have room for undercutting each other. Thus in equilibrium

$$p_{ig} = p_{jg} = R^* = c$$

Finally, consider the behavior of the originator. It takes $p_{ig}^* = c$ as given and solves

$$\max_{p_b} \ \pi_b = (p_b - c)D_b(p_b - R) \tag{31}$$

with $R = c$ from BI. This is equal to:

$$(p_b^* - c)\frac{\partial D_b(p_b^* - c)}{\partial p_b} + D_b(p_b^* - c) = 0$$

which gives the solution.

3. If a PSNE exists, then BI must hold. Assume BI does not hold. If the generic prices are equal to $\overline{p_g}$, it must be that

$$p_b > \overline{p_g} \Rightarrow R(p_b, p_{ig} = \overline{p_g} \forall i) > \overline{p_g}$$

which contradicts point 1. Thus, $R(p_b, \overline{p_g}) = \overline{p_g}$ only if $p_b = \overline{p_g}$. But this is not an equilibrium for $\varepsilon_{bb} \neq 0$.

4. When BI does not hold, $R^* = c$ is not an outcome in mixed strategies either. It is enough to prove that if the mixing strategies probability distribution of the brand product, $F(p_b)$, is defined in the interval $[\widetilde{p_b}, \widehat{p_b}]$, the R associated to $\widetilde{p_b}$ is everywhere greater than c. Given the problem, $\widetilde{p_b}$ is found by maximizing the branded drug profit function with generics prices fixed at c – i.e. by maximizing (31) with $R(p_b, c) > c$. This leads to:

$$\widetilde{p_b} \Rightarrow \frac{\widetilde{p_b} - c}{\widetilde{p_b}} = -\frac{1}{\varepsilon_b(c)}\left(1 - \frac{\partial R(c)}{\partial p_b}\right)^{-1}$$

where $R(c) \equiv R(p_b, c)$. By assumption

$$\frac{\partial R(c)}{\partial p_b} < 1$$

so

$$\frac{\widetilde{p_b} - c}{\widetilde{p_b}} > 0$$

and

$$\widetilde{p_b} > c$$

Finally, given that $\partial R(\cdot)/\partial p_{ig} > 0$, and considering that generics will mix in the interval $[c, \bar{R}]$, $R(\widetilde{p_b}, c)$ represents the lowest possible value of R that can be chosen in a mixed strategy equilibrium. But it is always true that

$$R(\widetilde{p_b}, c) > c \tag{32}$$

which completes the proof.

Lemma 3

When BI holds stage 1 p_b (p_b^1) does not influence future profits, hence $\bar{\pi}_b = 0$. When BI does not hold, an increase in p_b^1 increases R. Hence, we need to know what happens to stage 2 profits when R increases. Take $R_1 < R_2$ which define the optimal stage 2 deltas, Δ_1^{**} and Δ_2^{**} (pedix for b are omitted). Profits can then be written as $\pi_1 = (\Delta_1^{**} + R_1 - c)D_b(\Delta_1^{**})$ and $\pi_2 = (\Delta_2^{**} + R_2 - c)D_b(\Delta_2^{**})$. It is clear that, for a given R, it is always possible to choose a price such that $\Delta_1^{**} = \Delta_2^{**}$, implying $\pi_2 > \pi_1$. This implies that for higher values of R, the branded product can always increase profits. Hence, $\bar{\pi}_b > 0$.

Regarding generics:

$$\bar{\pi}_G = \underbrace{\frac{\partial R}{\partial p_G^1} D_G^2}_{\text{price effect}} + \underbrace{(R - c)\frac{\partial D_G^2}{\partial \Delta_b^{**}}\frac{\partial R}{\partial p_G^1}\left(\frac{\partial p_b^{**}}{\partial R} - 1\right)}_{\text{demand effect}} \tag{33}$$

While the price effect (i.e. the change in R due to a change in stage 1 prices) is always positive, the demand effect (i.e. the change in stage 2 demand due to the change in stage 1 prices) is always negative. If the former is greater, in absolute value, than the latter, the generic cartel will strategically decrease the optimal price in order to maintain a higher demand for the next T periods.

Lemma 4

The branded product must choose an optimal stage 1 price, p_b^*, so that to maximize the flow of profits:

$$\Pi_b = (p_b^1 - c)D_b(p_b^1 - R(p_b^1, p_G^1)) + \sum_1^T \delta^t (p_b^{**} - c)D_b(p_b^{**} - R(p_b^1, p_G^1)) \tag{34}$$

Similarly, p_G^* must maximize the joint profit:

$$\Pi_G = (p_G^1 - c)D_G(p_b^1 - R(p_b^1, p_G^1)) + \sum_1^T \delta^t (p_G^{**} - c)D_G(p_b^{**} - R(p_b^1, p_G^1)) \tag{35}$$

Note that none of the demand functions depend on generics prices. For stage 2, this is a consequence of Lemma 1. For stage 1, $p_G^1 \leq R$ by (1). If $\partial\Delta_b^{**}/\partial R < 0$ the equilibrium in (10) follows from maximizing Eqs. (34) and (35).

In order to prove the 'only if' condition, imagine that $\partial\Delta_b^{**}/\partial R \geq 0$. Then, for each stage 2 period:

$$\frac{\partial\pi_G^{**}}{\partial R} \equiv (R-c)\frac{\partial D_G(\Delta_b^{**})}{\partial\Delta_b}\frac{\partial\Delta_b^{**}}{\partial R} + D_G(\Delta_b^{**}) > 0$$

This would imply that in the optimum both $\partial\pi_b^{**}/\partial R > 0$ and $\partial\pi_G^{**}/\partial R > 0$. Knowing this, however, the generic cartel would increase R in stage 1 (also increasing its stage 1 profits), which contradicts the fact that Δ_b^{**} is an equilibrium.

Lemma 5

If $p_G = \theta$ is the stage 1 collusive price and we assume that $p_b > \theta$ (this is the only way to have a positive D_G for generics), then, by (1) and the assumption that $\partial R/\partial p_b > 0$, it is also true that $\theta \leq \bar{R}$.

Then, it must be shown that in order to gain from the betrayal, it must be that $\tilde{R} \cong \theta$. Start from the case of $\tilde{R} > \theta$. Then all generic products are free to consumers and demand for each firm joining the collusive agreement does not change. Hence, cheating does not provide any benefit to firm i. On the other hand, if $\tilde{R} < \theta$, firm i captures all the residual demand left for generic products. However, profits are not maximized since firm i can always do better by setting a price such that $\tilde{R} + \varepsilon < \theta$ with $\varepsilon > 0$. In equilibrium, $\tilde{R}$ must then be infinitely close to θ.

When BI does not hold, if $\bar{R}$ is the level of R under collusion, $\theta < \bar{R} < p_b$, by (1). Since after cheating it must be that $\tilde{R} \cong \theta$, clearly $\widetilde{p_i} < \theta$ and $\tilde{R} < \bar{R}$. Finally, since $\theta < \bar{R}$, it is easy to understand that $\pi(\bar{R}) > \pi(\theta)$.

When BI holds, $\theta = \bar{R} < p_b$. Hence a minimal reduction in price will be sufficient to reduce R just below θ. That is, $\widetilde{p_i} \cong \theta$. Since by definition of BI $\theta = \bar{R} < p_b$, it is also true that $\tilde{R} \cong \theta = \bar{R}$ and $\pi(\bar{R}) = \pi(\theta)$.

Lemma 6

The gains from betraying the cartel are

$$\pi(\widetilde{p_i}) - \frac{\pi(\theta)}{n}$$

in the first period, then followed by a loss of

$$\frac{1}{n}\sum_1^T(\delta)^t(\pi(\bar{R}) - \pi(\tilde{R})) = \frac{1}{n}\frac{\delta(1-\delta^T)}{1-\delta}(\pi(\bar{R}) - \pi(\tilde{R}))$$

for all the remaining ones. Thus, for firm i collusion is profitable if and only if

$$n\pi(\widetilde{p_i}) - \pi(\theta) < \frac{\delta(1-\delta^T)}{1-\delta}(\pi(\bar{R}) - \pi(\tilde{R})) \quad (36)$$

which gives the result.

Lemma 7 and corollary

The stream of profits from collusion is

$$\frac{1}{n}\left(\sum_{s=0}^{\infty}(\delta^{T+1})^s\pi(\theta) + \sum_{s=0}^{\infty}\delta^s\pi(\bar{R}) - \sum_{s=0}^{\infty}(\delta^{T+1})^s\pi(\bar{R})\right)$$
$$= \frac{1}{n}\left(\frac{1}{1-\delta^{T+1}}(\pi(\theta) - \pi(\bar{R})) + \frac{1}{1-\delta}\pi(\bar{R})\right)$$

In case of cheating followed by the classical "marginal cost" punishment, the flow of profits is

$$\pi(\tilde{p}) + \frac{1}{n}\left(\sum_{t=1}^{T}\delta^t\pi(\tilde{R})\right) = \pi(\tilde{p}) + \frac{1}{n}\frac{\delta(1-\delta^T)}{1-\delta}\pi(\tilde{R})$$

So, collusion is possible if:

$$\Pi_C \equiv \frac{1}{n}\left(\frac{1}{1-\delta^{T+1}}(\pi(\theta) - \pi(\bar{R})) + \frac{1}{1-\delta}\pi(\bar{R}) - n\pi(\tilde{p}) - \frac{\delta(1-\delta^T)}{1-\delta}\pi(\tilde{R})\right) \geq 0 \quad (37)$$

which gives the result in (16).

Taking the derivative with respect to T:

$$\frac{\partial\Pi_C}{\partial T} = \frac{\delta^{T+1}\log(\delta)(\pi(\theta) - \pi(\bar{R}))}{(1-\delta^{T+1})^2} + \frac{\delta^{T+1}\log(\delta)\pi(\tilde{R})}{1-\delta}$$

Collusion is less sustainable when T increases if the derivative is (weakly) negatively. Remembering that $\log(\delta) \leq 0$:

$$\frac{\partial\Pi_C}{\partial T} \leq 0 \Rightarrow \frac{(\pi(\theta) - \pi(\bar{R}))}{(1-\delta^{T+1})^2} + \frac{\pi(\tilde{R})}{1-\delta} \geq 0$$

which gives the result in (18).

Appendix B. Simulation

For the branded the problem is:

$$\max_{p_b} p_b[\alpha - \gamma(p_b - (wp_G + dp_b))] + \sum_1^T\delta^t\frac{1}{2}\left(\frac{\alpha}{\gamma} + (wp_G + dp_b)\right) \times \left[\alpha - \gamma\left(\frac{1}{2}\left(\frac{\alpha}{\gamma} - (wp_G + dp_b)\right)\right)\right]$$

The brand reaction function is then:

$$p_b^*(p_G) = \frac{\alpha\left(1 + \left(\left(\sum_1^T\delta^t d\right)/4\right)\right) + p_G\gamma\left(w + \left(\left(\sum_1^T\delta^t wd\right)/2\right)\right)}{\left(2\gamma w - \left(\left(\sum_1^T\delta^t d^2\right)/2\right)\right)} \quad (38)$$

For the generics:

$$\max_{p_G} p_G\gamma[(p_b - (wp_G + dp_b)] + \sum_1^T\delta^t(wp_G + dp_b) + \gamma\left[\left(\frac{1}{2}\left(\frac{\alpha}{\gamma} + (wp_G + dp_b)\right) - (wp_G + dp_b)\right)\right]$$

so that the generics' reaction function is:

$$p_G^*(p_b) = \frac{\left(\left(\sum_1^T\delta^t w\alpha\right)/2\gamma\right) + p_b\left(w - \sum_1^T\delta^t wd\right)}{\left(2w + \sum_1^T\delta^t w^2\right)} \quad (39)$$

Solving the system is tedious but easy.

With p_b^* and p_G^* all the other quantities follow. For the simulation of the condition for cartel's agreement, $w = n/(n+1)$ so that $\bar{R} = (n/(n+1))p_G^* + (1/(n+1))p_b^*$. The cheating price is found by solving:

$$p_G^* = \frac{n-1}{n+1}p_G^* + \frac{1}{n+1}\tilde{p} + \frac{1}{n+1}p_b^*$$

which gives $\tilde{p} = 2p_G^* - p_b^*$.

In order to apply Eq. (14), the following quantities are calculated:

1. stage 2 branded price under defection: $\widetilde{p_b^{**}} = (1/2)((\alpha/\gamma) + \tilde{R})$.
2. stage 1 defection profit: $\pi(\tilde{p}) = \gamma(p_b^* - \tilde{R})\tilde{p}$.
3. stage 1 profit under collusion: $\pi(p_G^*) = \gamma(p_b^* - \bar{R})p_G^*$.
4. stage 2 profit under defection: $\pi(\tilde{R}) = \gamma(\widetilde{p_b^{**}} - \tilde{R})\tilde{R}$.
5. stage 2 profit under collusion: $\pi(\bar{R}) = \gamma(p_b^{**} - \bar{R})\bar{R}$.

The n for collusion is defined as in (16). The threshold number of firms in Fig. 4 is then the first integer greater that the threshold n. Note also that γ cancels out, so that results do not depend on the price sensitivity parameter.

References

Bernheim, B.D., Whinston, M., 1990. Multimarket contact and collusive behaviour. Rand Journal of Economics 21, 1–26.

Brekke, K.R., Koningbauer, I., Straune, O.R., 2007. Reference pricing of pharmaceuticals. Journal of Health Economics 26 (3), 613–642.

Danzon, P., Ketcham, J.D., 2003. Reference Pricing of Pharmaceuticals for Medicare: Evidence from Germany, the Netherlands and New Zealand. NBER Working Paper w10007.

Danzon, P., Liu, H., 1997. Reference Pricing and Physician Drug Budgets: The German Experience in Controlling Pharmaceutical Expenditures. Working Paper Wharton.

d'Aspremont, C., Jacquemin, A., Gabszewick, J.J., Weymark, J.A., 1983. On the stability of collusive price leadership. Canadian Journal of Economics 16 (1), 17–25.

Escrihuela-Villar, M., 2009. A note on cartel stability and endogenous sequencing with tacit collusion. Journal of Economics 96 (2), 137–147.

Evans, W.N., Kessides, I.N., 1994. Living by the 'golden rule': multimarket contact in the US airline industry. Quarterly Journal of Economics 109, 341–366.

Friedman, J., 1971. A non-cooperative equilibrium for supergames. Review of Economic Studies 38, 1–12.

Galizzi, M.M., Ghislandi, S., Miraldo, M., 2011. Effects of reference pricing in pharmaceutical markets: a review. PharmacoEconomics 29 (1), 17–33.

Lopez-Casanovas, G., Puig-Junoy, J., 2000. Review of the literature on reference pricing. Health Policy 54, 87–123.

Merino, A., 2003. Impact of Reference Price System on the Pharmaceutical Market: A Theoretical Approach. Working Paper Pompeu Fabra.

Mestre-Ferrandiz, J., 2001. Reference price: the Spanish way. Investigaciones Economicas XXVII (1), 125–149.

Miraldo, M., 2009. Reference pricing and firms' pricing strategies. Journal of Health Economics 28, 176–197.

Puig-Junoy, J., 2004. Incentives and pharmaceutical reimbursement reforms in Spain. Health Policy 67, 149–165.

Schelling, T., 1960. The Strategy of Conflict. Harvard University Press.

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press, New York.

Zweifel, P., Crivelli, L., 1996. Price regulation of drugs: lessons from Germany. Journal of Regulatory Economics 10, 257–273.