Reference pricing of pharmaceuticals

Kurt R. Brekke [a], Ingrid Königbauer [b], Odd Rune Straume [c,d,*]

[a] *Department of Economics, Norwegian School of Economics and Business Administration, Norway*
[b] *Department of Economics, University of Munich, Ludwigstr. 28 VG, 80539 Munich, Germany*
[c] *Department of Economics and NIPE, University of Minho, Portugal*
[d] *Health Economics Bergen (HEB), Norway*

Abstract

We consider a therapeutic market with potentially three pharmaceutical firms. Two of the firms offer horizontally differentiated brand-name drugs. One of the brand-name drugs is a new treatment under patent protection that will be introduced if the profits are sufficient to cover the entry costs. The other brand-name drug has already lost its patent and faces competition from a third firm offering a generic version perceived to be of lower quality. This model allows us to compare generic reference pricing (GRP), therapeutic reference pricing (TRP), and no reference pricing (NRP). We show that competition is strongest under TRP, resulting in the lowest drug prices (and medical expenditures). However, TRP also provides the lowest profits to the patent-holding firm, making entry of the new drug treatment least likely. Surprisingly, we find that GRP distorts drug choices most, exposing patients to higher health risks.

1. Introduction

Pharmaceutical markets are characterised by price inelastic demand mainly due to extensive medical insurance. Since individuals – once they are ill – only pay a small fraction of the medical cost, prices are likely to have a limited effect, not only on the choice of whether or not to consume

---

* Corresponding author. Tel.: +351 253604539.

E-mail addresses: kurt.brekke@nhh.no (K.R. Brekke), ingrid.koenigbauer@lrz.uni-muenchen.de (I. Königbauer), o.r.straume@eeg.uminho.pt (O.R. Straume).

a drug, but also on the choice between alternative drug treatments. On the supply-side, there are large, sunk R&D costs associated with discovery of new drug treatments. To stimulate innovation, pharmaceutical firms are granted market power (for a given period) by patent protection.

The combination of supply-side market power and price inelastic demand has induced purchasers to employ various means to control medical expenditures.[1] We can distinguish between two price control mechanisms: (i) regulation of drug prices (price caps); and (ii) regulation of the reimbursement level, frequently referred to as *reference pricing* (RP). While price caps limit pharmaceutical firms' ability to exploit market power by charging high prices, RP aims at stimulating competition by making demand more price elastic. In this paper, we analyse in detail the effects of RP on the price-setting strategies of the pharmaceutical firms. On the basis of this analysis, we discuss implications for market entry of new drug treatments, patient health risks, and optimal drug reimbursement policies. While these issues have received some empirical attention, theoretical contributions are very limited.[2]

RP of prescription drugs is quite novel, but has rapidly become a widely used price control mechanism in the pharmaceutical market. Germany's Statutory Health Insurance System, generally viewed as the pioneer in this regard, introduced RP for prescription drugs in 1989, which was followed in Europe by the Netherlands in 1991, Denmark and Sweden in 1993, Spain in 2000, and Belgium and Italy in 2001. Norway adopted RP in 1993, but abandoned it in 2001, because the expected cost savings did not materialise. Outside Europe, RP has been adopted by Australia, the Canadian province of British Columbia, and New Zealand.[3]

The reference price is constructed as follows: drugs are classified into clusters based on similar therapeutic effects. The regulator sets a reference price based on a relatively low-priced drug (e.g., the minimum or median price) in the cluster. The reference price is the maximum reimbursement for all products in the group. Pharmaceutical firms can set prices above the RP, but in this case the patient must pay the surcharge.[4]

The construction of therapeutic clusters for RP is by far the most controversial task in the development of such systems. These clusters may be narrowly or broadly defined: (i) products with the same active chemical ingredients, (ii) products with chemically related active ingredients that are pharmacologically equivalent, and (iii) products that may be neither chemically identical nor pharmacologically equivalent but have comparable therapeutic effects. By its nature, the first type of cluster includes only off-patent brand-name drugs and their generic substitutes. The second and third may include on-patent drugs. They differ in breadth, but are qualitatively similar. As commonly done, we refer to the first type as *generic reference pricing* (GRP), and the second and third as *therapeutic reference pricing* (TRP).

We construct a theoretical model that allows us to analyse the effects of the two RP systems, as well as the benchmark case of no reference pricing (NRP), where patients pay a fixed share

---

[1] Danzon (1997) provides an excellent overview and discussion of various regulatory mechanisms in the pharmaceutical industry.

[2] According to the extensive literature survey by Lopez-Casasnovas and Puig-Junoy (2001), the bulk of the RP literature is mainly descriptive, and there is a pronounced lack of theoretical studies analysing the effects of RP systems. See also Danzon (2001).

[3] In the US, RP has been proposed as a possible approach to drug reimbursement for a comprehensive Medicare drug benefit (Huskamp et al., 2000). Kanavos and Reinhardt (2003) argue that RP for drugs is compatible with US health care. Notably, generic reference pricing is well-established in the US through "maximum allowable charge" programs used by, e.g., Medicaid.

[4] On the other hand, if a firm's price is below the RP, the savings may be shared between the payer and the dispensing pharmacist.

(given by a coinsurance rate) of the drug price.[5] The basic set-up is a therapeutic market with potentially three pharmaceutical firms, where two of the firms offer original brand-name drugs with different chemical ingredients. One of the brand-name drugs is an old treatment (e.g., the breakthrough drug) that has lost its patent protection and faces competition from a third firm offering a generic version, perceived to be of lower quality than the off-patent brand-name drug.[6] The other brand-name drug is a new, horizontally differentiated treatment under patent protection that will be introduced in the market, if the profits are sufficient to cover the entry costs.[7] This modelling approach enables us to discuss the arguments for and against RP systems in general, and between TRP and GRP in particular.

The main argument in favour of RP is that it stimulates price competition by making demand more elastic, resulting in lower medical expenditures.[8] Intuitively, the effect on price competition should be stronger the wider the cluster is defined. Our model confirms this line of argument. We show that the price of every drug in the therapeutic market is highest under NRP and lowest under TRP.[9] It is worth noting that GRP not only reduces prices of the drugs in the reference cluster, but also puts a downward pressure on the price of the non-included, but therapeutically equivalent, drug.[10] This is due to prices being strategic complements.

The inclusion of on-patent drugs is perhaps the main source of controversy over RP-systems. It is argued that TRP per se effectively eliminates patent protection and will stifle innovation in drug therapy, while GRP, on the other hand, is considered to have a minimal effect on incentives for R&D since it applies only to off-patent drugs (see e.g., Danzon, 2001; Lopez-Casasnovas and Puig-Junoy, 2000). Our model confirms the first line of the argument, but not the second. We show that TRP provides the lowest profits to the patent-holding firm, making market entry (and innovation) of the new drug treatment least likely.[11] However, we also find that a patent-holding firm can be negatively affected by RP, even if on-patent drugs are exempted from this particular reimbursement system. Stronger price competition induced by GRP forces the patent-holding firm to lower the price of its drug in order to reduce the loss of market shares.

Another important concern about TRP is that this system forces a large number of patients to opt for a less suitable drug simply to avoid the extra copayment. The broader the therapeutic cluster, the

---

[5] The NRP regime is often referred to as "free pricing", but we find this somewhat imprecise, since RP in itself does not restrict price-setting of drugs by pharmaceutical firms. Only the reimbursement level is regulated, not drug prices.

[6] Empirical evidence strongly suggests that generic drugs are not perceived to be perfect substitutes to the original brand-name drug, despite being chemically identical. After generic entry, the original brand-name firm typically charges a higher price than its generic version and still has positive market shares (e.g., Grabowski and Vernon, 1992; Frank and Salkever, 1997; Scott Morton, 2000). These findings fit well with predictions of vertical differentiation models. Two recent papers applied to branded-generic competition are Cabrales (2003) and Königbauer (2006).

[7] One can think of the entry costs as a marketing cost associated with entering a new country-specific market. Alternatively, the entry costs can be thought of as (expected) R&D costs, which must be recouped for the discovery of a new drug treatment to take place.

[8] To be precise, demand becomes more elastic above the reference price. Unless the reference price is set equal to the lowest priced drug in the cluster, the price elasticity of demand remains unchanged below the reference price.

[9] Pavcnik (2002) provides strong evidence from Germany that the introduction of RP has induced pharmaceutical prices to drop, the effect being stronger for branded drugs facing generic competition. Aronsson et al. (2001) and Bergman and Rudholm (2003) provide similar evidence from Sweden.

[10] A recent paper by Brekke et al. (2006b) provides empirical evidence on a cross-price effect of (generic) RP on non-included therapeutic substitutes.

[11] This result has empirical support from Danzon and Ketcham (2004) who analyse the effect of RP on the availability of drugs in Germany, the Netherlands and New Zealand.

more severe is the trade-off between surcharges and increased health risks to patients.[12] GRP, on the other hand, is said to conserve third party funds without exposing patients to significant risks, because it applies to substitution only among generically equivalent drugs that have demonstrated bioequivalence to the original brand-name drug. For *given prices*, this is, of course, trivially true. However, the intention of the RP systems is to induce price responses from the pharmaceutical firms. Taking this into account, we show that, in fact, GRP distorts drug choices most, exposing patients to higher health risks. Since the on-patent drug is exempted from reference pricing under GRP, the patent-holding firm faces a less price-elastic demand than its competitors, and can thus charge a considerably higher price. This induces a larger fraction of patients to choose the drugs that are included in the reference cluster, which are less suitable, but has a lower copayment.

In terms of policy implications, our results suggest no clear-cut conclusions about the optimal choice of reimbursement system. We can, however, make distinctions among the following general cases. If the costs of launching a new drug in a specific country are low, with a corresponding low risk of no market entry for new drugs, then TRP is clearly socially favourable.[13, 14] However, if this is not the case, then either NRP or GRP might be necessary to stimulate market entry. The choice between NRP and GRP implies a trade-off, since the former yields higher drug expenditures but lower health risks to patients. A social planner's evaluation of this particular trade-off is determined by the importance of drug expenditures in the planner's objective function. GRP might thus be the favoured reimbursement system in countries where the pharmaceutical industry is insignificant or non-existent, while NRP might be preferred otherwise.

The theoretical literature on RP is, as mentioned above, very limited, with only a couple of notable exceptions. Zweifel and Crivelly (1996) analyse the pricing responses to the introduction of a RP system using a Bertrand duopoly model. They frame their analysis in the context of the introduction of the TRP system in Germany in 1989. Danzon and Lui (1996) use a monopolistic competition model with kinked demand and imperfect physician agency to predict price responses to RP. The modelling approaches are distinctly different from ours. The combination of horizontal and vertical differentiation allows us to analyse and compare GRP and TRP closely. Moreover, our model also enables the analysis of market entry and health risks to patients, which are lacking in the above mentioned studies.[15]

Our paper contributes also to the more general literature on horizontal and vertical product differentiation. Most papers within this field allow firms to invest in quality, but assume consumers

---

[12] Lopez-Casasnovas and Puig-Junoy (2000, p. 111)formulate this problem as follows:

> "First, if there is no interchangeability at the level of the individual patient [. . .] then the copayment may become not avoidable and the RP system may discriminate against some patients. Second, selection of a drug under a RP category may result in a lower level of effectiveness and potentially harmful side effects for the patient because the drug is chosen simply with a view to avoiding the copayment".

The same argument is presented by Danzon (2001).

[13] Country-specific launching costs include typically marketing activities like providing information about the drug to the government in order to obtain sales approval and, potentially, listing in the reimbursement plan, promotion of the drug to physicians (detailing), etc.

[14] In a broader perspective, the introduction of TRP may influence the global launch decision. If large (and rich) countries – like the US and the UK – implement TRP, this may have a significant effect on the global returns on a new drug, which in turn may induce lower R&D investments. On the other hand, it has been argued that TRP induces more R&D investments in drastic relative to "me-too" innovations. The net effect of TRP on global welfare is thus not clear-cut, and definitely outside the scope of the current paper.

[15] These important aspects of RP-systems are also absent in Merino-Castelló (2003), who studies the price effects of generic reference pricing in a vertical differentiation model.

to differ only in terms of the horizontal space (taste).[16] The present paper explicitly combines the horizontal differentiation framework of Hotelling (1929) with the vertical differentiation framework introduced by Gabszewicz and Thisse (1979, 1980) and Shaked and Sutton (1982, 1983). While these two approaches typically are applied separately, the pharmaceutical market – with both inter-brand (branded versus branded) and intra-brand (branded versus generic) competition – serves as a natural example for combining these frameworks.

The paper is structured as follows: in Section 2, the model is presented. In Section 3, the equilibrium prices are derived and characterised for all three regimes. Section 4 analyses the market entry decision of the firm with the new drug treatment. Section 5 analyses the welfare properties of the three different regimes, and presents some policy implications. In Section 6, we extend our model to the special case of no coinsurance. Finally, Section 7 provides a discussion and some concluding remarks.

2. The model

Consider a particular therapeutic market for prescription drugs with the following characteristics. There are two patient types, indexed by $j = H, L$, differing with respect to their gross valuation of drug treatment, due to, e.g., different degrees of illness. A fraction λ of the patients are H-types, with a gross valuation v; the remaining patients – the L-types – have a gross valuation γv, where $\gamma \in (0, 1)$. Both patient types are uniformly distributed on the line segment $S = [0, 1]$, with a total mass of 1, where the location of an arbitrary patient, $x \in S$, is associated with the patient's susceptibility towards specific drug characteristics. A "mismatch cost" parameter t measures the utility loss per unit of distance between a patent's ideal treatment – given by his location on S – and the drug actually consumed. We can think of such mismatch costs as reflecting various side-effects or contraindications that reduce the gross valuation of drug treatment.

There are potentially three pharmaceutical single-product firms, indexed by $i = 0, 1, G$, operating in the market. Firms 0 and 1 offer original brand-name drugs at prices p_0 and p_1, respectively. These drugs, which differ with respect to chemical compounds, are located at either end of the unit interval S, reflecting their horizontally differentiated treatment effects. We assume that drug 1 is a new treatment version – still under patent protection – that will be introduced in this particular market, if variable profits are sufficient to cover entry costs. Drug 0, on the other hand, has already lost its patent protection and faces generic competition from a third pharmaceutical firm G, offering a generic drug version at a price p_G. In terms of horizontal differentiation, the generic drug is (naturally) also positioned at 0. However, in the eyes of the patients, 0 and G are *vertically differentiated*. This is captured by assuming that patients' gross valuation of the generic drug is deflated by a factor $\theta \in (0, 1)$. Thus, the perceived quality difference between the two versions of drug treatment 0 is given by $(1 - \theta)$. This vertical differentiation might be due to differences in advertising intensity that creates perceived quality differences, or simply that the brand-name drug is perceived to be safer due to a longer life in the market.

---

[16] Several papers have added quality competition to a standard Hotelling-framework, see e.g., Ma and Burgess (1993) for the case of fixed locations under both price competition and price regulation, Economides (1989) for the case of endogenous locations and price competition, and Brekke et al. (2006a) for the case of endogenous locations and price regulation. However, none of these papers allow consumers to differ with respect to their willingness-to-pay for quality, which means that the vertical differentiation framework is not explicitly dealt with.

Each patient needs one unit of either drug version. A patient of type j who is located at x and consumes a unit of drug i obtains utility

$$U_j(x, i) = \begin{cases} u_j - t|x - i| - c_i & \text{if } i = 0, 1 \\ \theta u_j - tx - c_i & \text{if } i = G \end{cases},$$ (1)

where

$$u_j = \begin{cases} v & \text{if } j = H \\ \gamma v & \text{if } j = L \end{cases},$$ (2)

and c_i is the patient copayment for drug i. In absence of a reference price system, the patient copayment for drug i is given by $c_i = \alpha p_i + f$, where $\alpha \in (0, 1)$ is the coinsurance rate and f is the deductible (or a flat fee).[17] Since patients pay the same deductible irrespective of which drug they decide to buy, f does not affect drug choices and can therefore be normalised to zero, without any loss of generality. On the other hand, in the presence of a reference price system, the copayment is based on a reference price $\bar{p}$, and the patients must additionally pay the full price difference if choosing a drug in the reference group which is priced in excess of the reference price. Thus, if drug i is included in a reference price system, the copayment is given by

$$c_i = \begin{cases} \alpha p_i & \text{if } p_i \leq \bar{p} \\ \alpha \bar{p} + (p_i - \bar{p}) & \text{if } p_i > \bar{p} \end{cases}.$$ (3)

We analyse a three-stage game with the following sequence of events:

1. A benevolent regulator decides on the socially optimal drug reimbursement policy to implement. She chooses among the following policies: (i) no reference pricing, (ii) therapeutic reference pricing, or (iii) generic reference pricing.
2. Firm 1 decides whether to enter the market and thus to offer a new treatment, given that treatment 0 already exists and is offered in the form of both an original version (drug 0) and a generic substitute (drug G).
3. All pharmaceutical firms in the market play a simultaneous pricing game.

As usual, the game is solved by backward induction.

3. Drug pricing

In this section we derive the optimal pricing strategies of the pharmaceutical firms for each of the three possible reimbursement regimes. We look for an equilibrium where all firms are active and compete in terms of prices. This requires some restrictions on the parameters. More specifically, we assume that the mismatch cost parameter t is bounded from both below and above, i.e., $t \in (\underline{t}, \bar{t})$, where the lower and upper bounds are functions of the other parameters. In the appendix we show that, when $t \in (\underline{t}, \bar{t})$, there exists a vertically separating equilibrium,

---

[17] This copayment system is the most common one (see e.g., Kanavos, 2001). However, some countries (e.g., the Netherlands) do not have any coinsurance element in the patient copayment. In the US, tiered formularies are typically used instead of coinsurance, except for the new Medicare prescription drug plan, which may include coinsurance. In Section 6 we extend the model to capture also the special case of no coinsurance (i.e., $\alpha = 0$), while Section 7 includes a brief discussion of tiered formularies.

where the brand-name drug 0 is priced 'high' and consumed by the H-types only, while the generic substitute G is priced 'low' and consumed by the L-types only.[18] On the other hand, the horizontally differentiated brand-name drug 1 is consumed by both types in equilibrium. This is the only possible type of equilibrium where the generic drug can survive in the market, since all patients prefer drug 0 over drug G if $c_0 = c_G$, implying that either all or no patients of type j prefer 0 over G if $c_0 \neq c_G$.

It is worth noting that, in this context, it makes considerable intuitive sense to focus on *intermediate* values of the mismatch cost parameter t. On the one hand, a very low t is not compatible with patent protection, since a new drug must be sufficiently differentiated to obtain a patent. On the other hand, a very high t is not compatible with the notion of a 'therapeutic market'. In particular, the idea of therapeutic reference pricing requires that the drugs included in a reference group are not too differentiated.

3.1. Demand and profits

Let us first derive drug demand for each firm, under the assumption of vertical market segmentation. This requires the identification of two indifferent patients; one for each of the two patient types.

The H-types choose between the two brand-name drugs, and the location of the indifferent H-type patient, denoted $\tilde{x}_H$, is given by the solution to

$$U_H(\tilde{x}_H, 0) = U_H(\tilde{x}_H, 1),$$

yielding

$$\tilde{x}_H = \frac{1}{2} + \frac{c_1 - c_0}{2t}. \tag{4}$$

The L-types, on the other hand, choose between the generic drug G and the horizontally differentiated brand-name drug 1. The location of the indifferent L-type patient, denoted $\tilde{x}_L$, is given by the solution to

$$U_L(\tilde{x}_L, G) = U_L(\tilde{x}_L, 1),$$

yielding

$$\tilde{x}_L = \frac{1}{2} + \frac{c_1 - c_G - \gamma v(1 - \theta)}{2t}. \tag{5}$$

Under the additional assumption of full market coverage, so that all patients obtain non-negative utility from the consumption of their most preferred drug, the demand facing firm i is given by[19]

$$D_i = \begin{cases} \lambda \tilde{x}_H & \text{if } i = 0 \\ \lambda(1 - \tilde{x}_H) + (1 - \lambda)(1 - \tilde{x}_L) & \text{if } i = 1 \\ (1 - \lambda)\tilde{x}_L & \text{if } i = G \end{cases}. \tag{6}$$

---

[18] To be more precise, we show that an equilibrium exists when $t \in (\underline{t}, \overline{t}^k)$, $k =$ NRP, TRP, GRP. In other words, there is a common lower bound on t in all three regimes, whereas the upper bound generally differs between the regimes.

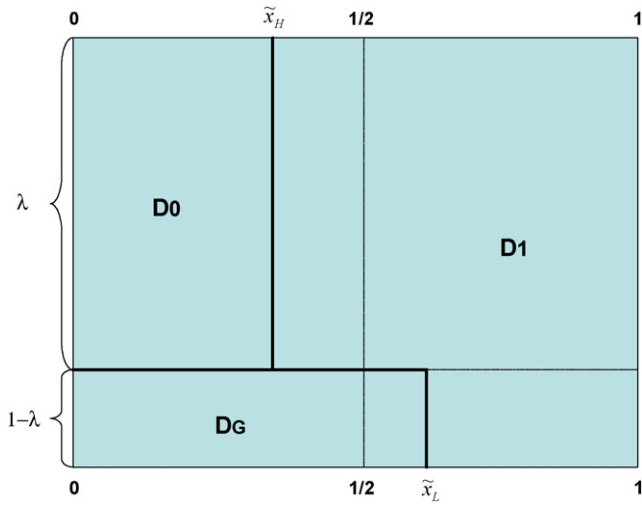[19] Fig. 1 illustrates the demand system.

Fig. 1. Illustration of the demand system.

Finally, assuming zero production costs, (variable) profits for firm i are simply given by[20]

$$\pi_i = p_i D_i. \tag{7}$$

3.2. No reference pricing

In the absence of any reference price system, the patient copayment for drug consumption is simply given by

$$c_i^{\text{NRP}} = \alpha p_i^{\text{NRP}}. \tag{8}$$

Explicit expressions for the profit functions under the NRP-system are easily found by using (8) in (4)–(7). In equilibrium, the two-brand name producers choose prices p_0^{NRP} and p_1^{NRP} that maximise π_0 and π_1, respectively, as defined by (7). The optimal strategy for the generic producer, on the other hand, is to choose a price p_G^{NRP} that is just low enough to make it unprofitable for firm 0 to deviate from p_0^{NRP} by setting a 'low' price that also captures the L-types. The equilibrium drug prices are given by[21]

$$p_0^{\text{NRP}} = \frac{3t}{\alpha} \Delta_0, \tag{9}$$

$$p_1^{\text{NRP}} = \frac{t}{\alpha} \Delta_1, \tag{10}$$

$$p_G^{\text{NRP}} = \frac{1}{\alpha} [3t\Delta_G - \gamma v(1 - \theta)], \tag{11}$$

---

[20] At this stage, market entry costs (R&D costs and/or marketing costs) are sunk and thus play no role for the analysis.

[21] A full derivation of the equilibrium is given in the appendix.

where

$$\Delta_0 := \frac{3 - (1 - \lambda)\sqrt{1 - \lambda}}{8 + \lambda(\lambda^2 + 3(1 - \lambda))} > 0, \tag{12}$$

$$\Delta_1 := \frac{10 - \lambda(\lambda^2 + 3(1 - \lambda)) - 6(1 - \lambda)\sqrt{1 - \lambda}}{8 + \lambda(\lambda^2 + 3(1 - \lambda))} > 0, \tag{13}$$

$$\Delta_G := \frac{4 - \lambda(2 - \lambda) - (4 - \lambda)\sqrt{1 - \lambda}}{8 + \lambda(\lambda^2 + 3(1 - \lambda))} > 0. \tag{14}$$

We see that all prices are increasing in t and decreasing in α. Higher mismatch costs reduce the substitutability, and thus the degree of competition, between the brand-name drugs, leading to higher prices. A higher coinsurance rate, on the other hand, increases the price elasticity of drug demand, leading to lower prices in equilibrium. It is also straightforward to show that $\partial \Delta_i / \partial \lambda > 0$, implying $\partial p_i / \partial \lambda > 0$, for all $i = 0, 1, G$. A higher fraction of H-types implies an increase in the overall willingness to pay, with a corresponding price increase, for the original drugs. This price increase also enables the generic producer to charge a higher price in equilibrium.[22] Note also that a reduction of the perceived quality difference between the two versions of treatment 0 (i.e., an increase in θ) leads to a higher price for the generic drug version, as expected.

On the other hand, a higher gross valuation of drug treatment for the L-types – i.e., an increase in γ – leads to a *lower* generic price in equilibrium. The reason is that a higher gross valuation for the L-types, implying a higher willingness-to-pay for drugs, makes it more profitable for firm 0 to lower its price in order to capture the L-segment of the market. Consequently, the generic firm must reduce its price in order to prevent this price-undercutting strategy from the brand-name firm. If the difference in gross valuations between the two patient types becomes sufficiently small – i.e., if γ becomes sufficiently close to 1 – it is not possible for the generic firm, with a (perceived) lower-quality product, to prevent that the brand-name firm serves both patient types in equilibrium. In this case, the generic drug is driven out of the market.

From (9)–(11) we can easily establish the following ranking of equilibrium drug prices:

$$p_0^{NRP} > p_1^{NRP} > p_G^{NRP}. \tag{15}$$

These price differences are reflected in the allocation of equilibrium market shares:

$$\tilde{x}_H^{NRP} = \frac{3[3 - (1 - \lambda)\sqrt{1 - \lambda}]}{2[8 + \lambda(\lambda^2 + 3(1 - \lambda))]} \in \left(\frac{3}{8}, \frac{1}{2}\right), \tag{16}$$

$$\tilde{x}_L^{NRP} = \frac{3[2 + (2 - \lambda)\lambda + (2 + \lambda)\sqrt{1 - \lambda}]}{2[8 + \lambda(\lambda^2 + 3(1 - \lambda))]} \in \left(\frac{1}{2}, 0.77\right). \tag{17}$$

Proposition 1. *Under NRP, the brand-name drug with a generic substitute always charges the highest price in equilibrium. Both patient groups are distorted; H-type patients consume more of the new, patent-protected, brand-name drug, while L-type patients consume more of the generic drug.*

It might seem counterintuitive that the price level is higher for the brand-name drug with a generic substitute, since, normally, we would expect prices to be lower for products that face

---

[22] From (11) and (14) we see that λ must be sufficiently high to secure a non-negative generic drug price, and thus equilibrium existence. See the appendix for exact conditions.

stronger competition. The reason for this result is that, due to generic competition, the optimal strategy of firm 0 is to concentrate exclusively on serving the H-type patients and leave the L-types to the generic competitor. Since firm 0 competes only for H-patients, with less price-elastic demand, while firm 1 competes for both patient types, firm 0 sets a higher price than firm 1 in equilibrium.[23] This theoretical result is reminiscent of several empirical findings of price increases for brand-name drugs after the entry of generic substitutes in the market.[24,25]

Inserting the equilibrium prices into (7), we derive equilibrium profits:

$$\pi_0^{NRP} = \frac{3t\lambda\Delta_0}{2\alpha}(1 + \Delta_1 - 3\Delta_0), \tag{18}$$

$$\pi_1^{NRP} = \frac{t\Delta_1}{2\alpha}(1 + 3(\Delta_G(1-\lambda) + \lambda\Delta_0) - \Delta_1), \tag{19}$$

$$\pi_G^{NRP} = \frac{(1-\lambda)}{2\alpha}(1 + \Delta_1 - 3\Delta_G)(3t\Delta_G - \gamma v(1-\theta)). \tag{20}$$

3.3. Reference pricing

Consider now the implementation of a reference pricing system. This implies that some drugs are aggregated into a cluster and are subject to the same reference price $\bar{p}$. The introduction of a reference pricing system involves the following decision-making.

First, the regulator must decide which drugs to include in a cluster, or reference group. In our model, this choice boils down to whether or not the new brand-name drug should be included. Inclusion of the horizontally differentiated new drug implies *therapeutic reference pricing*. On the other hand, if the reference group consists only of the old brand-name drug and its generic substitute, the reimbursement system is characterised as *generic reference pricing*.

Second, the regulator must decide on the reference price level. In most countries, this level is set at, or close to, the lowest drug price in the cluster. In the present analysis, we follow this practice by assuming that the lowest price in the reference group – i.e., the generic price – is chosen as the reference price level: $\bar{p} = p_G$.

3.3.1. Therapeutic reference pricing

Under TRP, the reference group consists of all three drugs in the therapeutic market, also the horizontally differentiated drug 1. By the assumption of $\bar{p} = p_G$, the copayments faced by patients under TRP are given by

$$c_i^{TRP} = \begin{cases} p_i^{TRP} - (1-\alpha)p_G^{TRP} & \text{if } i = 0, 1 \\ \alpha p_G^{TRP} & \text{if } i = G \end{cases}. \tag{21}$$

---

[23] Obviously, this result depends also on our assumption that the new horizontally differentiated product is not of higher quality than the old off-patent product.

[24] The empirical study by Grabowski and Vernon (1992) shows that generic entry was followed by price increases by the branded producer, a result later confirmed by Frank and Salkever (1997). This finding was called the "generic competition paradox" by Scherer (1993).

[25] In our model, it does not necessarily follow that generic entry leads to a higher price for the off-patent drug. Under the assumption of full market coverage in both patient segments, it is easily shown that equilibrium prices in the duopoly case, without generic competition, is given by $p_0 = p_1 = t/\alpha$, implying that generic entry leads to lower prices, although the price reduction is smallest for the off-patent drug 0. However, generic entry might lead to increased prices for the off-patent drug, if the L-segment is not fully covered in the duopoly equilibrium.

The copayments differ as compared to NRP, since the patients that are prescribed one of the original drugs are now also fully liable for the price difference with respect to the reference price.

As before, explicit expressions for the profit functions under the NRP-system are found by using (21) in (4)–(7), and the derivation of the equilibrium is similar to that under the NRP-system. We find equilibrium prices under TRP to be given by

$$p_i^{\text{TRP}} = \alpha p_i^{\text{NRP}}, \quad i = 0, 1, G. \tag{22}$$

Thus, compared with NRP, TRP implies that prices are set *as if* $\alpha = 1$. The reason is that, with TRP, the patients are fully liable for any price increase above the reference level. This also implies that equilibrium prices are independent of the coinsurance rate α. Furthermore, since equilibrium market shares are independent of α, both patient types are equally distorted under the two regimes.

Proposition 2. *In equilibrium, relative price differences and market shares are equal under NRP and TRP.*

Compared with the NRP-case, the (uniform) downward pressure on drug prices under TRP is also reflected in lower equilibrium profits, now given by

$$\pi_i^{\text{TRP}} = \alpha \pi_i^{\text{NRP}}, \quad i = 0, 1, G. \tag{23}$$

3.3.2. Generic reference pricing

Under GRP, only generic substitutes are grouped into the same cluster as the original, off-patent drugs. Horizontally differentiated, but therapeutically equivalent, drug versions are not included. In our model, copayments faced by consumers under GRP are thus given by

$$c_i^{\text{GRP}} = \begin{cases} p_i^{\text{GRP}} - (1-\alpha)p_G^{\text{GRP}} & \text{if } i = 0 \\ \alpha p_i^{\text{GRP}} & \text{if } i = 1, G \end{cases}. \tag{24}$$

While only a fraction α of the drug price needs to be paid on drugs G and 1, patients that are prescribed the brand-name drug 0 must additionally pay the full price difference between the original drug and the generic substitute.

Equilibrium prices, derived in the same way as previously, are given by

$$p_0^{\text{GRP}} = \frac{\left(2 + \alpha - \left(\sqrt{1-\lambda}\right)(2 - \lambda - \alpha)\right)\Gamma}{\tilde{\Delta}}, \tag{25}$$

$$p_1^{\text{GRP}} = \frac{t\bar{\Delta} + (1-\alpha)(1-\theta)\gamma v(\hat{\Delta} - 2(2+\alpha)) - \left(\sqrt{1-\lambda}\right)(2\alpha - \lambda(\alpha+1))\Gamma}{\alpha\tilde{\Delta}}, \tag{26}$$

$$p_G^{\text{GRP}} = \frac{3t(\alpha\lambda - 3\lambda + \lambda^2 + 4) - \gamma v(1-\theta)\hat{\Delta} - (4-\lambda)\Gamma\sqrt{(1-\lambda)}}{\tilde{\Delta}}, \tag{27}$$

where

$$\hat{\Delta} := 4\alpha + 5\lambda - 2\alpha\lambda - 4\lambda^2 + \lambda^3 + \alpha\lambda^2 + 4 > 0, \tag{28}$$

$$\tilde{\Delta} := 8\alpha + 8\lambda - 6\alpha\lambda - 5\lambda^2 + \lambda^3 + 2\alpha\lambda^2 + \alpha^2\lambda > 0, \tag{29}$$

$$\bar{\Delta} := 10\alpha + \lambda - 6\alpha\lambda + 2\lambda^2 - \lambda^3 + \alpha\lambda^2 + 2\alpha^2\lambda > 0, \tag{30}$$

$$\Gamma := 3t - \gamma v(1-\theta)(1-\alpha) > 0. \tag{31}$$

Using the equilibrium prices derived above, we can find the equilibrium market shares under GRP, characterised by the location of the indifferent patient in each patient-group:

$$\tilde{x}_{\mathrm{H}}^{\mathrm{GRP}} = \frac{\Gamma\left[(2+\alpha) - (2-\lambda-\alpha)\sqrt{1-\lambda}\right]}{2t\tilde{\Delta}}, \tag{32}$$

$$\tilde{x}_{\mathrm{L}}^{\mathrm{GRP}} = \frac{\Gamma\left[\alpha(2-\lambda) + \lambda(3-\lambda) + (2\alpha+\lambda)\sqrt{1-\lambda}\right]}{2t\tilde{\Delta}}. \tag{33}$$

Comparing with (16) and (17), it is also relatively straightforward to verify that

$$\tilde{x}_{j}^{\mathrm{GRP}} > \tilde{x}_{j}^{\mathrm{TRP}} = \tilde{x}_{j}^{\mathrm{NRP}}, \quad j = H, L, \tag{34}$$

implying that more patients choose one of the drugs included in the reference cluster under GRP—drug 0 and G.

In order to evaluate the ranking of equilibrium prices under GRP, we now make a rather weak assumption on the coinsurance rate, namely that $\alpha < (2/3)$. We are then able to make the following characterisation of the pricing equilibrium under generic reference pricing[26]:

Proposition 3. *Assume that $\alpha < (2/3)$. Then, under GRP, the brand-name firm without a generic substitute always charges the highest price in equilibrium. Both patient groups are generally distorted; the L-types always consume more of the generic drug, while the H-types consume more of the new patent-protected brand-name drug, if λ and/or t are sufficiently low, and more of the old off-patent product otherwise.*

A proof is given in Appendix D.

We see that the ranking of equilibrium prices changes under a generic reference price system; the price is now higher for the brand-name drug without a generic substitute. The reason is simply that drug 1 is not included in the reference cluster. If a consumer chooses this drug, her copayment is given by a share α on the *total* drug price. In contrast, if she chooses the off-patent drug 0, which is included in the reference cluster, she must pay the full price difference between the generic substitute and the brand-name drug. Thus, by not having its product included in the reference group, firm 1 faces a less elastic demand and will consequently charge a higher price in equilibrium.

In contrast to the NRP or TRP systems, equilibrium price differences do not automatically translate into equivalent differences in equilibrium market shares. The reason is the asymmetry introduced by different copayments for patients, depending on whether or not the demanded drug is subject to reference pricing. Consequently, even if firm 1 sets the highest drug price, it may not be the most expensive alternative for consumers, and consequently, this firm may have a higher market share in the H-segment. From Proposition 3 we see that this is will be the case if λ and/or t are sufficiently low. In this case, the price of the on-patent drug is kept relatively low by incentives to capture a larger share of the L-segment (which is more important the lower the level of λ) and/or fierce competition due to a relatively low degree of horizontal differentiation.

On the other hand, the location of the indifferent L-type patient is always distorted towards drug 1, as before. In other words, due to the price difference between generic and brand-name drugs, a larger share of L-patients consume the generic drug G. Finally, it should be noted that,

---

[26] We have tried, with no success, to find numerical examples where the ranking of equilibrium prices change for $\alpha > (2/3)$ (keeping in mind that equilibrium prices must satisfy the condition $t \in (\underline{t}, \overline{t})$). Thus, we believe that the results apply for all $\alpha \in (0, 1)$. Unfortunately, we are not able to prove this analytically.

even though the H-segment may be distorted 'both ways' under GRP, the L-segment is always more distorted towards drug 1. This can easily be verified from (32) and (33) by confirming that $\tilde{x}_L^{GRP} > \tilde{x}_H^{GRP}$.

Using the equilibrium prices reported in (25)–(27), we can derive equilibrium profits under GRP. These profit expressions are rather detailed, and are therefore relegated to the appendix.

3.4. Price comparison

As a next step, in order to evaluate how the reimbursement system affects drug prices, let us compare equilibrium price levels for the same drugs across different regimes. Using the equilibrium prices reported for the different cases above, it is relatively straightforward to verify that

$$p_i^{NRP} > p_i^{GRP} > p_i^{TRP}, \quad i = 0, 1, G, \tag{35}$$

for all $t > \underline{t}$. In other words:

Proposition 4. *The price of every drug in the therapeutic market is highest under NRP and lowest under TRP.*

This result reflects and confirms the main rationale behind reference pricing. By introducing a reference pricing system, price competition is generally increased since the price elasticity of drug demand increases for prices above the reference price level. Furthermore, this effect is stronger if more drugs are included in the reference cluster, implying that drug prices are lower under TRP than under GRP. Since prices are strategic complements, the introduction of a reference price system of either kind puts a downward pressure on the prices of *all* drugs in the market. Compared with the NRP case, the introduction of generic reference pricing has a direct negative effect on the price level of drug 0, which, in turn, leads to a reduction also in the price of drug 1, even though this drug is not included in the reference cluster under GRP. Furthermore, by going from GRP to TRP, firm 1 gets a direct incentive to cut its drug prices, which then indirectly leads to a further price reduction also for drug 0. Finally, lower prices for brand-name drugs imply that the generic producer must also lower its price in order to stay in the market.

4. Market entry

Let us now turn to the question of market entry. When interpreting the market in question as *country-specific* therapeutic market, demarcated by national regulation, we can realistically assume that firm 1 will enter this particular market (i.e., offer its newly developed product in this country) only if expected profits from sales in this market cover the market entry costs. When considering the costs and benefits of entry, the firm must take into account how the reimbursement policy in a given country is likely to affect profits from drug sales in this country.

In our model, there is a clear-cut ranking of equilibrium profits for the potential entrant (firm 1) across the different reimbursement regimes:

Proposition 5. *Equilibrium profits of the patent-holding entrant are always highest under NRP and lowest under TRP.*

A proof is given in Appendix D.

The profit comparison between NRP and TRP is straightforward. Compared with the case of no reference pricing, the TRP system puts a downward pressure on drug prices, while keeping

equilibrium market shares intact, implying that profits are unambiguously lower in the TRP equilibrium. NRP also outperforms GRP, from the viewpoint of firm 1, since prices *and* market shares are higher in the former case. A comparison between GRP and TRP, on the other hand, shows that prices are higher, but market shares lower, in the former case. Nevertheless, equilibrium profits are always higher under generic reference pricing. The reason is that, under GRP, firm 1 faces drug demand with a lower price elasticity, which enables this firm to charge a considerably higher price while suffering a moderate loss of market shares. All else equal, it follows that expected profits for a potential entrant are always lowest when entering a market that is subject to therapeutic reference pricing, and highest when entering a market with no reference pricing.

This result is not surprising, and tallies well with the popular concern about therapeutic reference pricing with respect to a potential erosion of patent rights, as discussed in the Introduction. However, it is worth noting that a patent-holding firm can be negatively affected by reference pricing even if on-patent drugs are exempted from this particular reimbursement system. In our model, firm 1's profits are lower under GRP, compared with no reference pricing, even if drug 1 is not included in the reference cluster. The reason is that firm 1 offers a drug that is an imperfect substitute to the drugs directly affected by the GRP system. Stronger price competition between firms 0 and G – induced by generic reference pricing – implies that firm 1 is also forced to lower the price of its on-patent drug in order to reduce the loss of market shares.

5. Welfare

In this section, we analyse and discuss the effects of different RP systems on social welfare. We will consider two different welfare perspectives: a *global welfare perspective* where social welfare is given by the sum of consumer and producer surplus net of third-party payment for drugs, and a *public payer perspective* where pharmaceutical profits do not enter the welfare function. In the first part of the section, we make the assumption that expected profits for firm 1 are always high enough to secure entry of the new drug. Subsequently, in the latter part of the section, we discuss policy implications when entry is uncertain. Throughout the welfare analysis, we also make the following two assumptions: First, we assume that the regulator does not take into account the "artificial" vertical differentiation between the branded and generic drugs. In other words, the regulator attaches the same gross utility to objectively homogenous products. We think this is a reasonable (though not trivial) assumption. Second, we assume, for simplicity, that the public payer is able to raise the necessary funds for drug payment in a non-distortionary manner.

5.1. Global welfare

If the welfare function is defined as the sum of consumer and producer surplus, net of third-party payment for drugs, social welfare under reimbursement system k is given by

$$W_k = \bar{U} - C_k, \tag{36}$$

where $\bar{U}$ denotes patients' gross utility of drug consumption [27], while C_k denotes total mismatch costs under reimbursement system k. Thus, maximising welfare amounts to minimising total

---

[27] When the perceived vertical differentiation is not taken into account, patients' gross utility of drug consumption is given by $\bar{U} := v(\lambda + (1 - \lambda)\gamma)$.

mismatch costs, which are given by

$$C_k = \lambda \left(\int_0^{\tilde{x}_H^k} (st)\,ds + \int_{\tilde{x}_H^k}^1 ((1-s)t)\,ds \right) + (1-\lambda) \left(\int_0^{\tilde{x}_L^k} (st)\,ds + \int_{\tilde{x}_L^k}^1 ((1-s)t)\,ds \right).$$

(37)

Clearly, total mismatch costs are minimised if $\tilde{x}_L^k = \tilde{x}_H^k = (1/2)$. In other words, mismatch costs are minimised if all patients located at $x \le (1/2)$ are prescribed either drug 0 or G, while all patients located at $x > (1/2)$ are prescribed drug 1. However, due to price differences, total mismatch costs will never be minimised in equilibrium. We have previously shown that $\tilde{x} \ne (1/2)$ for at least one patient type in all three reimbursement regimes. We also know that equilibrium market shares are equal under NRP and TRP, implying that total mismatch costs must also be equal under these two regimes.

The explicit expression for total mismatch costs in each of the three different regimes, which are quite detailed, are given in the appendix. Based on these expressions, we are able to derive the following unambiguous ranking of reimbursement systems with respect to equilibrium mismatch costs:

$$C_{\text{GRP}} > C_{\text{TRP}} = C_{\text{NRP}}.$$

(38)

This constitutes the main result of our welfare analysis[28]:

Proposition 6. *NRP and TRP yield equal mismatch costs in equilibrium, and these are always lower than under GRP.*

It follows that, if a regulator seeks to maximise global welfare, generic reference pricing should never be implemented; mismatch costs would be lower in equilibrium by choosing either NRP or TRP. In order to explain this result, let us first consider the distortive effects of GRP on each of the two patient types. We know that $\tilde{x}_L^{\text{GRP}} > \tilde{x}_L^{\text{TRP}} = \tilde{x}_L^{\text{NRP}} > (1/2)$, due to the larger price difference between the generic drug and the horizontally (and vertically) differentiated drug 1 under GRP.[29] This implies that GRP always increases total mismatch costs in the L-segment. For H-types, on the other hand, we know that $\tilde{x}_H^{\text{TRP}} = \tilde{x}_H^{\text{NRP}} < (1/2)$ and $\tilde{x}_H^{\text{GRP}} > \tilde{x}_H^{\text{TRP}} = \tilde{x}_H^{\text{NRP}}$. However, since $\tilde{x}_H^{\text{GRP}} \lessgtr (1/2)$, it is possible that GRP reduces aggregate mismatch costs for the H-types if $\tilde{x}_H^{\text{GRP}}$ is sufficiently close to the midpoint of the line segment S. Nevertheless, a possible reduction in mismatch costs for H-types will always be more than outweighed by the increase in mismatch costs for L-types. The reason is two-fold. First, mismatch costs are reduced for H-types only if λ – the fraction of H-types in the population – is sufficiently low (cf. Proposition 3), in which case the contribution of H-types to *total* mismatch costs is also relatively low. Second, since the location of the indifferent L-type is further away from the midpoint of S in all regimes, the effect of a marginal relocation of the indifferent patient on total mismatch costs is – all else equal – larger in the L-segment.

---

[28] The proof, though conceptually straightforward, involves some extremely tedious and detailed algebra and is thus not reported. However, just to give a brief sketch, it is possible to show that $C_{\text{GRP}} - C_{\text{NRP}} = \varphi_1/\varphi_2$, where $\varphi_2 > 0$ and φ_1 is a convex quadratic function of t which crosses zero from below at $t = \underline{t}$. Thus, $\varphi_1 > 0$ for $t > \underline{t}$. It follows that $C_{\text{GRP}} > C_{\text{TRP}} = C_{\text{NRP}}$ for $t > \underline{t}$.

[29] This is not obvious, though, since $c_1^{\text{GRP}} < c_1^{\text{TRP}}$ for $p_1^{\text{GRP}} = p_1^{\text{TRP}}$. Thus, when comparing GRP and TRP, the equilibrium price difference between drug 1 and drug G under GRP is sufficiently large to (more than) compensate for the lower copayment share, leading to a larger distortion in the L-segment towards drug G.

The result stated in Proposition 6 is perhaps somewhat surprising. It certainly runs contrary to the popular concern about the discriminatory effects of therapeutic reference pricing, that this reimbursement system forces a larger number of patients to opt for a less suitable drug – thereby increasing mismatch costs – simply to avoid the extra copayment. However, this is not the case in our model. True, therapeutic reference pricing will increase overall mismatch costs *for given prices*, if we use the NRP-case as a benchmark. But this argument ignores the fact that pharmaceutical firms will adjust their pricing policies according to the drug reimbursement system. In our specific model, we have seen that TRP will lead to a proportionally equal reduction in all drug prices, leaving patients' drug choices unaffected in equilibrium, compared with NRP. Generic reference pricing, on the other hand, will lead to more distorted drug choices, due to larger equilibrium price differences within the therapeutic market. Since the on-patent drug is exempted from reference pricing under GRP, firm 1 faces a less price-elastic demand than its competitors and can thus charge a considerably higher price in equilibrium. This, in turn, induces more patients to choose the drugs that are included in the reference cluster, leading to higher overall mismatch costs.[30]

5.2. Public payer objectives

The welfare function given by (36) is relevant also for countries with a significant pharmaceutical industry. However, in countries where the pharmaceutical industry is absent (or insignificant), it is reasonable to assume that total drug expenditures enter the welfare function directly. Indeed, a stated desire behind the introduction of reference pricing in many countries is precisely to curb total outlays on pharmaceuticals. In the following, we will therefore consider a more narrow public payer objective, where the welfare function is given by consumer surplus net of third-party payments. We can interpret this as national welfare in countries with no pharmaceutical industry. In this case, social welfare under reimbursement system k is given by

$$W_k = \bar{U} - C_k - \pi_0^k - \pi_1^k - \pi_G^k. \tag{39}$$

With this particular welfare function, maximising welfare amounts to minimising the sum of mismatch costs and drug expenditures.[31] It is straightforward to show that total profits, and thus total spending on pharmaceuticals, are lowest under TRP and highest under NRP. This follows, for the most part, directly from the previously derived price and profit rankings.[32] Furthermore, we know from Proposition 6 that TRP always yields equal or lower total mismatch costs in equilibrium, compared with NRP or GRP. Thus, we reach the following unambiguous conclusion:

Proposition 7. *Both mismatch costs and drug expenditures are minimised under TRP*.

It follows that social welfare – as defined by (39) – is always higher under TRP.

---

[30] If we include the perceived quality difference between branded and generic drugs in the welfare function, total mismatch costs will be even higher under GRP (compared with NRP or TRP), since GRP induces more patients to choose the generic drug.

[31] Under the assumption of zero production costs, total drug expenditures (from patients and the public payer) are equal to total profits.

[32] The only possible ambiguity arises in the comparison between NRP and GRP, since both firms 0 and G have lower prices but higher market shares under GRP. However, this ambiguity is easily resolved by the following argument: Going from NRP to GRP, expenditures obviously decrease for the segments where patients consume the same drug in the two regimes, since all prices are lower under GRP. Then there is a segment where H-patients switch from drug 1 to drug 0, and a segment where L-patients switch from drug 1 to drug G. However, expenditures also decrease for both these segments, since (by Propositions 3 and 4) $p_1^{NRP} > p_0^{GRP}$ and $p_1^{NRP} > p_G^{GRP}$.

In the above analysis, we have considered two polar welfare perspectives, where pharmaceutical profits are given either full or no weight in the regulator's objective function. In general, the relative weighting of mismatch costs and drug expenditures in the welfare function is likely to depend on the relative importance of the pharmaceutical industry in the country in question. The more important the pharmaceutical industry is, the less concerned a regulator should be about pharmaceutical spending. In any case, though, as long as the regulator places any weight on pharmaceutical spending at all, the above analysis clearly suggests that a therapeutic reference price system should be implemented, as, compared with the other considered alternatives, this reimbursement scheme minimises both mismatch costs and pharmaceutical expenditures.

However, this result is reached under the assumption of certain entry of drug 1, and the above conclusion is only valid if there is indeed an additional, horizontally differentiated drug version that can be included in the therapeutic cluster. Since equilibrium profits are lowest under TRP (cf. Proposition 5), this reimbursement system makes market entry least likely, for a given level of market entry costs. If the possibility of no market entry is taken into account, then the welfare considerations are no longer clearly in favour of TRP. First, no entry will lead to *maximal* mismatch costs, because only one treatment version (drug 0 and its generic substitute) is offered in the market. Second, the absence of competition from a horizontally differentiated drug will lead to increased drug prices – and thus increased pharmaceutical spending – under both NRP and GRP. In this scenario, the regulator must take into account how the choice of reimbursement system is likely to affect the probability of market entry for new drugs.

No clear-cut conclusions can be made about the optimal choice of reimbursement system. However, based on the above analysis, we can make the following classification of scenarios. *Therapeutic reference pricing* – which minimises both mismatch costs and drug prices – is clearly the socially favourable reimbursement system if market entry costs are low, with a corresponding low risk of no market entry for new drugs. However, if this is not the case, then either NRP or GRP might be necessary to stimulate market entry. There is then a case for *no reference pricing* – which minimises mismatch costs but maximises drug prices – in countries where drug prices do not play an important role for social welfare, due to a dominant pharmaceutical industry. On the other hand, *generic reference pricing* might be the favoured reimbursement system in countries where the pharmaceutical industry is insignificant or non-existent, since GRP leads to lower drug prices than NRP.

6. Extension: The case of no coinsurance

Some regulatory systems – like in the Netherlands and the US Medicaid – do not expose patients to coinsurance. In this section, we therefore consider the special case of $\alpha = 0$, and check whether our main results still apply. It is, however, worth noticing that α could be interpreted as the prescribing physicians' price consciousness (see, e.g., Hellerstein (1998)). Thus, the case of $\alpha = 0$ also implicitly relies on the assumption that the physician is a perfect agent for the patients only, and therefore do not take the payer's potential expenditures into account. We retain all assumptions of our main model, with the exception that α is now set to zero.[33]

---

[33] As previously argued, introducing a fixed copayment f for drug consumption would not affect the results as long as f is equal for all drugs. This also applies when $\alpha = 0$. Thus, for simplicity, we retain our assumption of $f = 0$.

Consider first the case of *no reference pricing*. In this regime, consumers will be insentive to price differences when choosing which drug to purchase; only the drugs' vertical and horizontal characteristics matter. As a consequence, no consumer demands the generic drug, since this is perceived to be of lower quality than the two brand-names. The choice between the two brand-name drugs is solely determined by their horizontal characteristics. Since the two drugs are symmetrically differentiated, each brand-name receives half of the market, i.e., $D_0 = D_1 = 1/2$, irrespective of their price setting.

The profit functions of the brand-name firms are, thus, $\pi_0 = p_0/2$ and $\pi_1 = p_1/2$. Obviously, both firms would charge the maximum possible price. Let $\hat{p}$ denote the maximum price the firms are able to obtain. This price can be interpreted as a price cap imposed by the payer or the outcome of negotiations between payers and firms. Alternatively, if the insurance market is competitive, such that payers have no market power, we can think of $\hat{p}$ as the price that makes the payers' budget constraint bind, i.e., the actuarily fair premium. The determination of this maximum price is beyond the scope of this paper. In the following, we just assume that it exists. Thus, under no reference pricing, the equilibrium prices are simply

$$p_0^{\text{NPR}} = p_1^{\text{NPR}} = \hat{p} \tag{40}$$

The generic firm obtains zero demand and profits, and is therefore not active in this equilibrium. Thus, a copayment system with no coinsurance is harmful for generic competition. The reason is simply that, under this system, demand is perfectly inelastic to prices.

Consider now the case of *therapeutic reference pricing*. In Section 3, we saw that the equilibrium outcomes under TRP are independent of the coinsurance rate. Consequently, the equilibrium presented in Section 3 for TRP applies also for the special case of $\alpha = 0$. As previously shown, since, with TRP, patients are fully liable for any price increase above the reference price, the firms respond to TRP by setting prices as if the patients had no insurance coverage ($\alpha = 1$). This implies that equilibrium prices are independent of the coinsurance rate.

Finally, consider the case of *generic reference pricing*. In this case, patients must pay the price difference $p_0 - p_G$ if they choose the original brand-name drug 0, but not if they choose the horizontally differentiated drug 1. The derivation of the equilibrium is similar to the case of $\alpha > 0$. Note, however, that there is no effectual horizontal competition under GRP when $\alpha = 0$, since firm 1 has no means to attract patients. Thus, the equilibrium strategy for firm 1 is to set the price as high as possible. The equilibrium prices are given by[34]:

$$p_0^{\text{GRP}} = \left(\frac{1 - \sqrt{1 - \lambda}}{\lambda} \right) \left[t - v\gamma(1 - \theta) \right], \tag{41}$$

$$p_1^{\text{GRP}} = \hat{p}, \tag{42}$$

---

[34] The equilibrium exists if

$$t \in \left[\frac{2(1 - \theta)(1 - \sqrt{1 - \lambda})v\gamma}{2(1 - \sqrt{1 - \lambda}) - \lambda}, \tilde{t} \right],$$

where

$$\tilde{t} := \min \left\{ \frac{v(2\lambda - \gamma(1 - \sqrt{1 - \lambda})(1 - \theta))}{2\lambda + \sqrt{1 - \lambda} - 1}, \gamma v(1 + \theta), 2v(1 - \theta) \right\}.$$

$$p_G^{GRP} = 2 \left(\frac{1 - \sqrt{1 - \lambda}}{\lambda} \right) [t - v\gamma(1 - \theta)] - t, \tag{43}$$

yielding equilibrium market shares

$$\tilde{x}_H^{GRP} = \left(\frac{1 - \sqrt{1 - \lambda}}{\lambda} \right) \tilde{x}_L^{GRP} \tag{44}$$

and

$$\tilde{x}_L^{GRP} = \frac{1}{2} - \frac{v\gamma(1 - \theta)}{2t}. \tag{45}$$

Since $(1 - \sqrt{1 - \lambda})/\lambda \in ((1/2), 1)$ for $\lambda \in (0, 1)$, it follows that $\tilde{x}_H^{GRP} < \tilde{x}_L^{GRP} < (1/2)$. Thus, both patient segments are distorted towards drug treatment 1, and the distortion is higher in the H-segment. This is quite intuitive, since, when comparing drug G and drug 0, the brand-name drug is perceived to be of higher quality without being more expensive for patients, while, when comparing drug 0 and drug 1, patients must pay a surcharge equal to $p_0^{GRP} - p_G^{GRP}$ if choosing drug 0.

While equilibrium profits for firm 1 – and thus the profitability of market entry – obviously depends on the magnitude of $\hat{p}$, we notice that equilibrium market shares do not depend on this price. Consequently, it is – in principle – straightforward to compare aggregate mismatch costs under the three different reimbursement schemes also for the special case of $\alpha = 0$. We immediately realise that NRP now yields minimal mismatch costs, since firm 0 and firm 1 split the market equally between them. However, this is only because there is no room for generic competition without coinsurance.[35] Thus, the interesting comparison is between GRP and TRP. A key result in our previous analysis is that GRP yields the highest health risks, measured as total mismatch costs. Does this still apply in a system without coinsurance (i.e., $\alpha = 0$)?

From (41)–(45), we can derive equilibrium mismatch costs under GRP with no coinsurance; these are reported in the appendix. Since $\tilde{x}_H^{GRP} < \tilde{x}_L^{GRP} < (1/2)$, we already know that mismatch costs are never minimised under GRP. Whether GRP or TRP yields the highest health risks with no coinsurance depends on a comparison of (C.1) and (C.10), given in the appendix. Unfortunately, it is infeasible to provide an analyical characterisation of this difference. However, numerical simulations strongly suggest that $C_{GRP}|_{\alpha=0} > C_{TRP}$ for the valid parameter configurations.[36] Thus, total mismatch costs seem to be higher under GRP also for the special case of no coinsurance.

The causes of the larger distortion under GRP are somewhat different from the previous case, though. When $\alpha > 0$, GRP yields higher mismatch costs mainly because of increased distortion towards the generic drug in the L-segment. However, with no coinsurance, GRP creates a relatively strong distortion – in both patient segments – towards drug 1, as explained above. Under TRP, on the other hand, there is an extra copayment, in equilibrium, for consumption of drug 1. This contributes to a more symmetric outcome, compared with GRP, with less distorted drug choices.

7. Discussion and concluding remarks

We have analysed the effects of reference pricing systems for pharmaceuticals, focusing on a specific therapeutic market with potentially three pharmaceutical firms. Two of the firms offer

---

[35] There could be room for generic competition without coinsurance if patients have to pay a fixed fee that is lower for generic drugs than for brand-name drugs. This possibility is not pursued in the present model.

[36] More details are available from the authors upon request.

horizontally differentiated brand-name drugs. One of these drugs is off-patent and faces com-petition from a generic version offered by a third firm. The other drug is on-patent and will be introduced in the market, if the profits are sufficient to cover the entry costs.

This framework has allowed us to compare generic reference pricing and therapeutic reference pricing, as well as the benchmark-case of no reference pricing. We have shown that TRP triggers competition most, resulting in lower equilibrium prices for every drug in the therapeutic market. We have also shown that GRP distorts drug choices most, resulting in a higher level of patient health risks – measured in terms of aggregate mismatch costs – than the other two reimbursement systems. Thus, TRP is preferable from the perspective of both the purchaser (payer) and the patients.

Notably, the beneficial role of TRP crucially relies on the assumption that the new on-patent drug enters the market. If the market entry costs are sufficiently high, TRP may in fact result in a worse outcome than both GRP and NRP, as described above. It has, however, been argued that TRP may induce pharmaceutical firms to invest more in drastic innovations, not subject to reference pricing, rather than non-drastic innovations, which very likely will be included in a reference group. The trade-off with respect to therapeutically similar innovations is thus the following: while innovations of therapeutic substitutes increase competition and reduce patients' mismatch costs by offering a different variant of treatment for the same illness, they might crowd out drastic innovations if they reduce the budget available for R&D. On the other hand, different drug versions are often innovated in so-called R&D-races, implying that therapeutically similar innovations are already in the 'pipeline' when the first drastic innovation enters the market. A thorough analysis of this issue requires an explicit model of drug innovations, which is outside the scope of the present paper.

There are also several other issues related to reference pricing of pharmaceuticals that, naturally, have not been subject to a full treatment within our model framework. As previously discussed, an important and much debated issue is how RP systems affect entry of new drugs. In the present paper, we have focused on entry of therapeutic substitutes. However, RP systems may also affect entry incentives for generic substitutes. We can reasonably assume that the probability of generic entry increases with expected profits for the producer of generic drugs. In our model, it is possible to show that – similar to firm 1 profits – equilibrium profits of the generic producer (firm G) is highest under NRP and lowest under TRP. Thus, based on our analysis, it is possible to argue that RP – in particular TRP – might discourage not only entry of therapeutic substitutes, but also generic entry.

A related issue is how entry impacts the RP level. It has been argued that entry of new on-patent drugs may, in fact, raise the RP level. Our model produces the opposite result. Entry of a new treatment (drug 1) triggers competition, resulting in lower prices and thus a lower reference price level. One can, however, argue that, if the new drug is of substantially higher quality than the existing treatment, the producer can charge a higher price, which in turn may result in a higher RP. This reasoning is correct if we assume existing prices to be given and thus ignore price responses to the entry of a new treatment. However, entry of a new and better treatment is likely to induce the incumbent firms to reduce their prices *even further* (compared with entry of a new but equally good treatment) in order to avoid losing large market shares. If the RP is set equal to the lowest price in the cluster, as we assume in the paper, then surely entry, even of a high-quality treatment, will result in a lower RP level. However, under a more general RP rule, like a weighted average of all drug prices in a cluster, the impact of entry on the RP level may be indeterminate, with the sign depending on the weight attached to the new drug treatment relative to the existing treatments.

Another interesting issue is the optimal clustering of drugs in a therapeutic reference pricing system. Our model – with only three drugs in the therapeutic market – has been constructed

to analyse the policy choice between generic and therapeutic reference pricing. In other words, the question of optimal clustering has been narrowed to a question of generic versus therapeutic reference pricing. However, if there are more than two horizontally differentiated drugs in a therapeutic market, another related policy question arises: which, and how many, of these drugs should be included in a therapeutic reference cluster? In other words, what is the optimal breadth of a therapeutic cluster under TRP? This is a question that cannot be addressed directly in the present model. However, our analysis can still shed some light on the relevant mechanisms and trade-offs involved, and give some indications of the likely results.

In our model, the breadth of a therapeutic cluster can be captured by the parameter t. Under TRP, a lower (higher) value of t implies – all else equal – a narrower (broader) cluster. Obviously, equilibrium prices, with or without reference pricing, are increasing in t. More substitutability (a lower t) intensifies competition and leads to lower drug price. However, from (9)–(11) and (22), it is clear that – compared with NRP – the price effect of TRP is also increasing in t. In other words, the competitive effect of TRP is larger in markets where there is less competition to begin with. This suggests that, if the regulator is mainly concerned about reducing drug spending, the therapeutic clusters should be broadly defined. On the other hand, broader clusters imply a potentially stronger negative effect on the entry of new drugs. What about the concern for minimising health risks (mismatch costs)? Our analysis suggests that, as long as *all* drugs in a therapeutic market are included, the introduction of TRP does not lead to more distorted drug choices. However, if some drugs are excluded from the reference cluster, mismatch costs are likely to increase. This would, in fact, resemble the comparison between GRP and TRP in our model, where the exclusion of one drug from the reference cluster leads to more asymmetric, and thus more distorted, drug choices. We can thus speculate that, in order to avoid large distortions in drug choices under TRP, the reference cluster should either be relatively narrowly defined, in order to maintain sufficient competition between excluded drugs, or very broadly defined, where all drugs in the therapeutic market are included. Intuitively, the former alternative will, to a larger extent, stimulate entry of new drugs, while the dampening effect on drug spending will be more moderate.

Finally, our modelling of RP is also, naturally, stylised to a point where it is not possible to capture the full variety of different reimbursement systems that can be observed across different countries. One important policy option that has not been considered in the present paper is *tiered formularies*, where patients are exposed to different copayments according to a drug's status in the formulary of the benefit plan. Typically, patients face the highest copayment (or even no coverage) for drugs not included in the formulary, medium copayment for listed brand-name drugs, and lowest copayment for listed generic drugs. Tiered formularies are thus close to a coinsurance system in the sense that it aims at inducing patients to purchase cheaper drugs with similar therapeutic effects. Now, if a plan with a three-tiered formulary specifies the different copayments as $c_i = \alpha p_i$, $i = 0, 1, G$, then the NRP regime would in fact be equivalent to a tiered formulary system. Obviously, the correspondence between drug prices and copayments may not be so direct. If the different copayments are not very responsive to drug prices, we are, in effect, back to the case of no coinsurance, with the only difference that relative copayments will shift market shares.[37] In this case, tiered formularies do not affect pricing decisions, they only shift

---

[37] More precisely, the demand structure would be like the following:

$$\tilde{X}_H = \tilde{x}_H - \frac{f_0 - f_1}{2t} \quad \text{and} \quad \tilde{X}_L = \tilde{x}_L - \frac{f_G - f_1}{2t}.$$

demand towards drugs with lower copayments, implying that the analysis of a tiered formulary would not be qualitatively different from the present analysis.

Acknowledgements

Appendix A. Derivation of the price equilibrium

In a vertically separating equilibrium, characterised by a price vector (p_0, p_1, p_G), the following conditions must hold:

Condition 1: $p_G \geq 0$.
Condition 2: $U_L(x, G) \geq U_L(x, 0)$.
Condition 3: $U_H(x, 0) \geq U_H(x, G)$.
Condition 4: $U_H(\tilde{x}_H, 0) \geq 0$.
Condition 5: $U_L(\tilde{x}_L, G) \geq 0$.
Condition 6: $\pi_0(p_0, p_1, p_G) \geq \pi_0(\hat{p}_0, p_1, p_G)$, where $\hat{p}_0$ solves $U_L(x, G) = U_L(x, 0)$.
Condition 7: $\pi_G(p_0, p_1, p_G) \geq \pi_G(p_0, p_1, \hat{p}_G)$, where $\hat{p}_G$ solves $U_H(x, 0) = U_H(x, G)$.

The first condition simply states that the generic price must be non-negative. Conditions 2–3 ensure that the equilibrium really separates, i.e., that H-types choose the brand-name drug 0, while L-types choose the generic substitute. Conditions 4 and 5 secure full market coverage, requiring that the indifferent patients obtain non-negative utility from purchasing and consuming either of the drugs. Finally, Condition 6 (7) ensures that Firm 0 (Firm G) has no incentive to deviate by reducing its price and serve the L-types (H-types).

In the following, we will derive the price equilibrium in detail for the NRP-case. For the two other cases – where the derivation of the equilibrium follows an identical procedure – we will just present the constraints that support the equilibrium.

A.1. No reference pricing

Profit functions are given by (7), with $c_i = \alpha p_i$. Let us first confirm that unconstrained pricing by all three firms cannot constitute an equilibrium. Unconstrained maximisation of the firms' profit functions yields the following reaction functions:

$$p_0 = \frac{1}{2\alpha}(t + \alpha p_1), \tag{A.1}$$

$$p_1 = \frac{1}{2\alpha}[t + (1 - \lambda)(1 - \theta)\gamma v + \alpha p_G(1 - \lambda) + \alpha \lambda p_0], \tag{A.2}$$

$$p_G = \frac{1}{2\alpha}[t + \alpha p_1 - \gamma v(1 - \theta)], \tag{A.3}$$

which yield the following candidate equilibrium prices:

$$p_0 = \frac{1}{\alpha}\left[t + \frac{1}{6}\gamma v(1-\theta)(1-\lambda)\right], \tag{A.4}$$

$$p_1 = \frac{1}{\alpha}\left[t + \frac{1}{3}\gamma v(1-\theta)(1-\lambda)\right], \tag{A.5}$$

$$p_G = \frac{1}{\alpha}\left[t - \frac{1}{6}\gamma v(1-\theta)(2+\lambda)\right]. \tag{A.6}$$

We can show that this price vector always violates Condition 2. In the NRP-case, Condition 2 can be expressed as

$$p_G \le p_0 - \frac{1}{\alpha}\gamma v(1-\theta). \tag{A.7}$$

Using (A.4) and (A.6), this condition reduces to $2 \ge 5$, which is a contradiction. In other words, (A.4)–(A.6) cannot be an equilibrium, because p_G is too high to induce even the L-type patients to buy the generic drug. Consequently, we must look for an equilibrium where the generic drug is priced sufficiently low, so that not only are the L-types not induced to switch to drug 0, but firm 0 must also have no incentive to capture the L-types by lowering its price from the equilibrium level.

Using (A.1) and (A.2), we can express the profit of firm 0 as a function of p_G:

$$\pi_0(p_G) = \frac{\lambda[3t + (1-\lambda)(\alpha p_G + (1-\theta)\gamma v)]^2}{2\alpha t(4-\lambda)^2}. \tag{A.8}$$

Firm 0 can drive the generic competitor out of the market, and capture equal shares of the H- and L-types, by setting a price

$$\hat{p}_0 = p_G + \frac{1}{\alpha}\gamma v(1-\theta), \tag{A.9}$$

which yields a "deviation" profit given by

$$\hat{\pi}_0(p_G) = \frac{[6t - (2+\lambda)(\alpha p_G + (1-\theta)\gamma v)](\alpha p_G + (1-\theta)\gamma v)}{2\alpha t(4-\lambda)}. \tag{A.10}$$

The optimal strategy for firm G is thus to set a price p_G that is just low enough to make such a deviation unprofitable. This price is given by the solution to

$$\pi_0(p_G) = \hat{\pi}_0(p_G). \tag{A.11}$$

We can thus derive the price equilibrium by solving the three Eqs. (A.1), (A.2) and (A.11). The solution is presented as (9)–(11) in Section 3.

It remains to specify Conditions 1–7 for the NRP-case. By construction of the equilibrium, we know that Condition 6 is automatically satisfied. We can also show that Condition 2 is always satisfied. In the NRP-case, this condition is given by

$$\theta\gamma v - \alpha p_G^{NRP} \ge \gamma v - \alpha p_0^{NRP}, \tag{A.12}$$

which, using (9) and (11), reduces to

$$\Delta_0 - \Delta_G \ge 0, \tag{A.13}$$

which is true for all $\lambda \in (0, 1)$. The remainder of the constraints can be expressed in the form of four different conditions on t. From (11), we see that a non-negative generic drug price – Condition 1 – is guaranteed if

$$t \geq t_1^{\mathrm{NRP}} := \frac{(1 - \theta)\gamma v}{3\Delta_{\mathrm{G}}}. \tag{A.14}$$

Furthermore, non-negative utility for the indifferent consumers of the H- and L-type, respectively, is guaranteed if

$$t \leq t_4^{\mathrm{NRP}} := \frac{2v}{1 + 3\Delta_0 + \Delta_1} \tag{A.15}$$

and

$$t \leq t_5^{\mathrm{NRP}} := \frac{2\gamma v}{1 + \Delta_1 + 3\Delta_{\mathrm{G}}}. \tag{A.16}$$

The necessary Condition 7 is not analytically solvable. However, to simplify, we can find a *sufficient* condition on t that satisfies Conditions 3 and 7 simultaneously. By assuming that H-types always prefer drug 0 over drug G for the equilibrium price p_0^{NRP} and a zero-priced generic drug (i.e., $p_{\mathrm{G}} = 0$), it must be true that H-types always prefer drug 0 in equilibrium (for a non-negative generic drug price) *and* that price-undercutting by the generic firm in order to capture H-type consumers is not an option. Using p_0^{NRP} from (9), and setting $p_{\mathrm{G}} = 0$, this condition is given by

$$t \leq t_7^{\mathrm{NRP}} := \frac{v(1 - \theta)}{3\Delta_0}. \tag{A.17}$$

To sum up, a price equilibrium exists in the NRP-case, and is given by (9)–(11), when $t \in [\underline{t}, \overline{t}^{\mathrm{NRP}}]$, where $\underline{t} := t_1^{\mathrm{NRP}}$ and $\overline{t}^{\mathrm{NRP}} := \min\{t_4^{\mathrm{NRP}}, t_5^{\mathrm{NRP}}, t_7^{\mathrm{NRP}}\}$. In general, existence of the equilibrium requires that the share of L-types is relatively low, combined with a sufficiently large difference in gross valuations between the two types. To give an illustrative numerical example, assume that $v = 1$, $\lambda = 0.9$, $\theta = 0.8$ and $\gamma = 0.4$. In this case, $\underline{t} = 0.12$ and $\overline{t}^{\mathrm{NRP}} = t_7^{\mathrm{NRP}} = 0.20$. Note also that the equilibrium exists for an even wider range of mismatch costs, since the upper bound $\overline{t}^{\mathrm{NRP}}$ in this case is a sufficient, but not necessary, condition.

A.2. Therapeutic reference pricing

The price equilibrium under TRP is derived similarly to the NRP-case, and given by (22) in Section 3. As before, Condition 6 is automatically satisfied. Furthermore, Conditions 1 and 2 are identical under NRP and TRP. The remainder of the Conditions – 4, 5 and 3 + 7 – are given by, respectively,

$$t \leq t_4^{\mathrm{TRP}} := \frac{2(1 - \gamma(1 - \theta)(1 - \alpha))v}{1 + 3\Delta_0 + \Delta_1 - 6\Delta_{\mathrm{G}}(1 - \alpha)}, \tag{A.18}$$

$$t \leq t_5^{\mathrm{TRP}} := \frac{2(\theta + \alpha(1 - \theta))\gamma v}{1 + \Delta_1 - 3\Delta_{\mathrm{G}}(1 - 2\alpha)}, \tag{A.19}$$

$$t \leq t_7^{\mathrm{TRP}} := \frac{(1 - \gamma(1 - \alpha))(1 - \theta)v}{3(\Delta_0 - \Delta_{\mathrm{G}}(1 - \alpha))}. \tag{A.20}$$

Thus, under TRP, an equilibrium exists, and is given by (22), when $t \in [\underline{t}, \bar{t}^{\text{TRP}}]$, where $\bar{t}^{\text{TRP}} := \min\{t_4^{\text{TRP}}, t_5^{\text{TRP}}, t_7^{\text{TRP}}\}$. It is worth noting that, due to lower equilibrium prices, the range of mismatch costs for which the equilibrium exists is generally wider under TRP. Using the same numerical example as in the NRP-case, with a 10% copayment rate ($\alpha = 0.1$), the lower and upper bounds on t are given by $\underline{t} = 0.12$ and $\bar{t}^{\text{TRP}} = t_7^{\text{TRP}} = 0.34$.

A.3. Generic reference pricing

The price equilibrium under GRP is derived similarly to the NRP- and TRP-cases, and given by (25)–(27) in Section 3. As before, Condition 6 is automatically satisfied.

Using (25)–(27), we can derive the remainder of the conditions that support the equilibrium under GRP. Once more, it can be shown that Condition 1 is satisfied if $t \geq \underline{t}$, implying that Condition 1 is identical for all three regimes.

Condition 2 is given by

$$t \geq t_2^{\text{GRP}} := \frac{1}{3}(1 - \alpha)\gamma(1 - \theta)v. \tag{A.21}$$

Since $t_1^{\text{GRP}} \geq (1 - \theta)\gamma v$, it follows that $\underline{t} \geq t_2^{\text{GRP}}$. Thus, as long as Condition 1 is satisfied, Condition 2 is also automatically satisfied. Conditions 4 and 5 are given by, respectively,

$$t \leq t_4^{\text{GRP}} := \frac{2\tilde{\Delta}v + (1 - \alpha)(1 - \theta)\gamma v(3(2 + \alpha) + \varkappa - 2\hat{\Delta})}{\tilde{\Delta} + \bar{\Delta} + 15\alpha - 3\lambda\alpha(4 - \alpha - \lambda) - 3(1 - \lambda)(2 - \lambda) + 3\varkappa}, \tag{A.22}$$

where

$$\varkappa := \left(\sqrt{1 - \lambda}\right)[2 + \lambda - \alpha(5 - 2\lambda)], \tag{A.23}$$

and

$$t \leq t_5^{\text{GRP}} := \frac{\tilde{\Delta}(1 + \theta) + (1 - \theta)[(1 - \alpha)[2(2 + \alpha) + \varsigma] - \hat{\Delta}(1 - 2\alpha)]}{\tilde{\Delta} + \bar{\Delta} + 3\alpha[4 - \lambda(3 - \alpha - \lambda)] + 3\varsigma}, \tag{A.24}$$

where

$$\varsigma := \left(\sqrt{1 - \lambda}\right)(\lambda(2\alpha + 1) - 6\alpha). \tag{A.25}$$

Finally, the sufficient condition that simultaneously satisfies Condition 3 and Condition 7 is given by

$$t \leq t_7^{\text{GRP}} := \frac{(1 - \theta)(\tilde{\Delta}v + \gamma v(1 - \alpha)(2 + \alpha + (\sqrt{1 - \lambda})(2 - \alpha(3 - \lambda)) - \hat{\Delta}))}{3(5\alpha + \lambda(1 - \alpha)(3 - \alpha - \lambda) + (\sqrt{1 - \lambda})(2 - \alpha(3 - \lambda)) - 2)}. \tag{A.26}$$

Thus, under GRP, an equilibrium exists, and is given by (25)–(27), when $t \in [\underline{t}, \bar{t}^{\text{GRP}}]$, where $\bar{t}^{\text{GRP}} := \min\{t_4^{\text{GRP}}, t_5^{\text{GRP}}, t_7^{\text{GRP}}\}$. Once more, due to the general price reducing effect of reference pricing, the range of mismatch costs for which the equilibrium exists is generally wider also under the GRP system, compared with the NRP case. Using the same numerical example as previously, the lower and upper bounds on t are given by $\underline{t} = 0.12$ and $\bar{t}^{\text{GRP}} = t_7^{\text{GRP}} = 0.29$.

Appendix B. Equilibrium profits under GRP

Equilibrium profits under generic reference pricing are given by

$$\pi_0^{\text{GRP}} = \frac{(2 + \alpha - (\sqrt{1 - \lambda})(2 - \lambda - \alpha))^2 \Gamma^2 \lambda}{2t\tilde{\Delta}^2}, \tag{B.1}$$

$$\pi_1^{\text{GRP}} = \frac{(3t - \Gamma)(\Omega - 2\sqrt{1 - \lambda}(\lambda - 2\alpha + \alpha\lambda)\Psi) + t^2(6(\sqrt{1 - \lambda})(\lambda - 2\alpha + \alpha\lambda)\bar{\Delta} + \Phi)}{2t\alpha\tilde{\Delta}^2}, \tag{B.2}$$

$$\pi_G^{\text{GRP}} = \frac{\Gamma\left(\alpha(2 - \lambda) + \lambda(3 - \lambda) + \left(\sqrt{1 - \lambda}\right)(2\alpha + \lambda)\right)(1 - \lambda)\Theta}{2t\tilde{\Delta}^2}, \tag{B.3}$$

where

$$\Omega := 2t\omega_1 + \gamma v(1 - \alpha)(1 - \theta)\omega_2, \tag{B.4}$$

$$\begin{aligned}
\omega_1 := {}&64\alpha\lambda + 8\alpha^2 + 2\lambda^2 + 9\lambda^3 - 12\lambda^4 + 6\lambda^5 - \lambda^6 - 86\alpha\lambda^2 - 8\alpha^2\lambda + 40\alpha\lambda^3 + 4\alpha^3\lambda \\
&- 6\alpha\lambda^4 + 19\alpha^2\lambda^2 - 13\alpha^2\lambda^3 - 4\alpha^3\lambda^2 + 3\alpha^2\lambda^4 + 2\alpha^3\lambda^3,
\end{aligned} \tag{B.5}$$

$$\begin{aligned}
\omega_2 := {}&16\alpha\lambda + 8\alpha^2 + 26\lambda^2 - 41\lambda^3 + 26\lambda^4 - 8\lambda^5 + \lambda^6 - 30\alpha\lambda^2 - 16\alpha^2\lambda + 28\alpha\lambda^3 \\
&- 12\alpha\lambda^4 + 2\alpha\lambda^5 + 13\alpha^2\lambda^2 - 5\alpha^2\lambda^3 + \alpha^2\lambda^4,
\end{aligned} \tag{B.6}$$

$$\begin{aligned}
\Psi := {}&\gamma v(1 - \alpha)(1 - \theta)(2\alpha + 5\lambda - 2\alpha\lambda - 4\lambda^2 + \lambda^3 + \alpha\lambda^2) + 2t(2\alpha - 7\lambda + 7\lambda^2 - 2\lambda^3 \\
&- \alpha\lambda^2 + \alpha^2\lambda),
\end{aligned} \tag{B.7}$$

$$\begin{aligned}
\Phi := {}&136\alpha^2 - 16\alpha\lambda + 10\lambda^2 - 5\lambda^3 + 2\lambda^4 - 4\lambda^5 + \lambda^6 + 82\alpha\lambda^2 - 192\alpha^2\lambda - 60\alpha\lambda^3 \\
&+ 40\alpha^3\lambda + 16\alpha\lambda^4 - 2\alpha\lambda^5 + 105\alpha^2\lambda^2 - 13\alpha^2\lambda^3 - 24\alpha^3\lambda^2 - 3\alpha^2\lambda^4 + 4\alpha^3\lambda^3 \\
&+ 4\alpha^4\lambda^2,
\end{aligned} \tag{B.8}$$

$$\Theta := 3t(\alpha\lambda - 3\lambda + \lambda^2 + 4) - \gamma v(1 - \theta)\bar{\Delta} - \left(\sqrt{1 - \lambda}\right)(4 - \lambda)\Gamma. \tag{B.9}$$

Appendix C. Equilibrium mismatch costs

Inserting the expressions for the locations of indifferent patients in the different reimbursement regimes – reported throughout Section 3 – into (37), equilibrium mismatch costs are given by

$$C_{\text{NRP}} = C_{\text{TRP}} = \frac{(\delta - 6(\sqrt{1 - \lambda})(5\lambda + 4)(1 - \lambda)^3)t}{4(3\lambda - 3\lambda^2 + \lambda^3 + 8)^2}, \tag{C.1}$$

where

$$\delta := 104 + 6\lambda - 78\lambda^2 + 53\lambda^3 - 15\lambda^4 + 15\lambda^5 - 4\lambda^6, \tag{C.2}$$

and

$$C_{\text{GRP}} = \frac{(3t - \Gamma)[(3t - \Gamma)(2\sqrt{1 - \lambda}\Upsilon - \Lambda) + 2t(\sqrt{1 - \lambda}\mu - \eta)] + t^2(F + 6\xi\sqrt{1 - \lambda})}{4t\tilde{\Delta}^2},$$

(C.3)

where

$$\xi := 16\alpha\lambda - 12\lambda - 4\alpha^2 + 23\lambda^2 - 17\lambda^3 + 4\lambda^4 - 18\alpha\lambda^2 + 5\alpha^2\lambda + 4\alpha\lambda^3 - 2\alpha^3\lambda + \alpha\lambda^4$$
$$- 3\alpha^2\lambda^2 + 2\alpha^2\lambda^3 + \alpha^3\lambda^2,$$

(C.4)

$$F := 72\lambda + 64\alpha\lambda + 104\alpha^2 - 94\lambda^2 + 74\lambda^3 - 53\lambda^4 + 25\lambda^5 - 4\lambda^6 - 150\alpha^2\lambda - 30\alpha\lambda^3$$
$$+ 20\alpha^3\lambda + 44\alpha\lambda^4 - 10\alpha\lambda^5 + 66\alpha^2\lambda^2 + 7\alpha^2\lambda^3 - 12\alpha^3\lambda^2 - 6\alpha^2\lambda^4$$
$$+ 2\alpha^3\lambda^3 + 2\alpha^4\lambda^2 - 40\alpha\lambda^2,$$

(C.5)

$$\Lambda := 12\lambda^3 - 16\alpha\lambda - 8\alpha^2 - 2\lambda^2 - 8\lambda - 7\lambda^4 + \lambda^5 + 24\alpha\lambda^2 + 14\alpha^2\lambda - 14\alpha\lambda^3 + 2\alpha\lambda^4$$
$$- 8\alpha^2\lambda^2 + \alpha^2\lambda^3,$$

(C.6)

$$\Upsilon := 8\alpha\lambda - 4\lambda + 4\alpha^2 + 5\lambda^2 - 4\lambda^3 + \lambda^4 - 10\alpha\lambda^2 - 5\alpha^2\lambda + 3\alpha\lambda^3 + 2\alpha^2\lambda^2,$$

(C.7)

$$\mu := 24\lambda - 40\alpha\lambda - 8\alpha^2 - 38\lambda^2 + 29\lambda^3 - 7\lambda^4 + 48\alpha\lambda^2 + 10\alpha^2\lambda - 13\alpha\lambda^3 + 2\alpha^3\lambda$$
$$- \alpha\lambda^4 - 3\alpha^2\lambda^2 - 2\alpha^2\lambda^3 - \alpha^3\lambda^2,$$

(C.8)

$$\eta := 24\lambda - 8\alpha\lambda + 8\alpha^2 - 34\lambda^2 + 21\lambda^3 - 12\lambda^4 + 6\lambda^5 - \lambda^6 + 16\alpha\lambda^2 - 14\alpha^2\lambda - 20\alpha\lambda^3$$
$$- 2\alpha^3\lambda + 15\alpha\lambda^4 - 3\alpha\lambda^5 - 5\alpha^2\lambda^2 + 11\alpha^2\lambda^3 + 2\alpha^3\lambda^2 - 3\alpha^2\lambda^4 - \alpha^3\lambda^3.$$

(C.9)

C.1. GRP with no coinsurance

Using (41)–(45), equilibrium mismatch costs under GRP for the special case of $\alpha = 0$ are given by

$$C_{\text{GRP}}|_{\alpha=0} = \frac{\gamma v(1 - \theta)\left(\gamma v(1 - \theta)\left(2\left(1 - \sqrt{1 - \lambda}\right) - \lambda^2\right) + 2t\hat{\lambda}\right) + t^2\tilde{\lambda}}{4t\lambda},$$

(C.10)

where

$$\hat{\lambda} := \sqrt{1 - \lambda}(2 - \lambda) - 2(1 - \lambda)$$

(C.11)

and

$$\tilde{\lambda} := 2 - \lambda(2 - \lambda) - 2\sqrt{1 - \lambda}(1 - \lambda).$$

(C.12)

Appendix D. Proofs

Proof of Proposition 3. In equilibrium, the price difference between the two brand-name drugs are given by

$$p_1^{\text{GRP}} - p_0^{\text{GRP}} = \frac{\gamma v(1-\alpha)(1-\theta)[\hat{\Delta} - 4 + \alpha^2 - (\sqrt{1-\lambda})(\lambda - \alpha^2)] + t\sigma}{\alpha \tilde{\Delta}}, \tag{D.1}$$

where

$$\sigma := 4\alpha + \lambda - 6\alpha\lambda - 3\alpha^2 + 2\lambda^2 - \lambda^3 + \alpha\lambda^2 + 2\alpha^2\lambda + 3(\sqrt{1-\lambda})(\lambda - \alpha^2). \tag{D.2}$$

By the definition of $\hat{\Delta}$, it can easily be verified that the sum of the four terms in the square brackets in the numerator in (D.1) is positive for $\alpha \in (0, 1)$ and $\lambda \in (0, 1)$. The sign of the expression depends thus on the sign of σ. Once more, it is relatively straightforward to verify that $\sigma > 0$ for all $\lambda \in (0, 1)$ if $\alpha < (2/3)$. Thus, $\alpha < (2/3)$ is a *sufficient* condition for $p_1^{\text{GRP}} > p_0^{\text{GRP}}$.

Regarding equilibrium market allocations, we derive from (33) that

$$\tilde{x}_{\text{L}}^{\text{GRP}} > \frac{1}{2} \quad if \quad t > \gamma v(1-\theta)\beta, \tag{D.3}$$

where

$$\beta := (1-\alpha)\frac{(\sqrt{1-\lambda})(2\alpha + \lambda) + 2\alpha + \lambda(3 - \lambda - \alpha)}{3(\sqrt{1-\lambda})(2\alpha + \lambda) - 2\alpha + \lambda + 3\alpha\lambda + 2\lambda^2 - \lambda^3 - 2\alpha\lambda^2 - \alpha^2\lambda}. \tag{D.4}$$

It is fairly straightforward to verify that $\beta < 1$ for $\alpha \in (0, 1)$ and $\lambda \in (0, 1)$. This implies that $t > \gamma v(1-\theta)\beta$ (and thus $\tilde{x}_{\text{L}}^{\text{GRP}} > (1/2)$) as long as Condition 1 (non-negative generic price) is satisfied.

Now consider the indifferent type-H patient. From (32), we can characterise $\tilde{x}_{\text{H}}^{\text{GRP}}$ as a function of t in the following way:

$$\frac{\partial \tilde{x}_{\text{H}}^{\text{GRP}}}{\partial t} > 0 \quad \text{for} \quad t \neq 0, \tag{D.5}$$

$$\lim_{t \to 0^+} (\tilde{x}_{\text{H}}^{\text{GRP}}) \to -\infty, \tag{D.6}$$

and

$$\lim_{t \to -\infty} (\tilde{x}_{\text{H}}^{\text{GRP}}) = \lim_{t \to \infty} (\tilde{x}_{\text{H}}^{\text{GRP}}) = \vartheta, \tag{D.7}$$

where

$$\vartheta := \frac{3(2 + \alpha - (\sqrt{1-\lambda})(2 - \lambda - \alpha))}{2\tilde{\Delta}}. \tag{D.8}$$

It follows that $\tilde{x}_{\text{H}}^{\text{GRP}} < (1/2)$ for $t > 0$ if $\vartheta < (1/2)$ for $\alpha \in (0, 1)$ and $\lambda \in (0, 1)$. On the other hand, if $\vartheta > (1/2)$ for some combinations of λ and α, it must be that $\tilde{x}_{\text{H}}^{\text{GRP}} > (1/2)$ if t is sufficiently high. Solving $\vartheta = (1/2)$ for α yields a function $\alpha^*(\lambda)$, such that $\vartheta < (>)(1/2)$ if $\alpha < (>)\alpha^*(\lambda)$. It is straightforward to verify that $\partial \alpha^*/\partial \lambda > 0$ and that $\alpha^* < 0$ for $\lambda < 0.54$. It follows that $\tilde{x}_{\text{H}}^{\text{GRP}} < (1/2)$ if $\lambda < 0.54$, whereas, for $\lambda > 0.54$, $\tilde{x}_{\text{H}}^{\text{GRP}} > (1/2)$ if λ and/or t are sufficiently high. By numerical simulations, it is also straightforward to verify that both cases, $\tilde{x}_{\text{H}}^{\text{GRP}} < (1/2)$ and $\tilde{x}_{\text{H}}^{\text{GRP}} > (1/2)$, can occur in equilibrium. $\square$

Proof of Proposition 5. A direct analytical comparison of equilibrium profits for firm 1 under the three different regimes is infeasible, since the equilibrium profit expression under GRP is extremely detailed. However, we can prove the proposition via a somewhat more subtle route, by considering how different reimbursement systems affect equilibrium prices and market shares. From Proposition 3, we know that there is a clear-cut ranking of equilibrium prices across the different regimes, where $p_i^{NRP} > p_i^{GRP} > p_i^{TRP}$, $i = 0, 1, G$. Regarding equilibrium market shares, we know that these are identical under NRP and TRP. Furthermore, we also know that $\tilde{x}_j^{GRP} > \tilde{x}_j^{TRP} = \tilde{x}_j^{NRP}$, $j = H, L$. Thus, since $p_1^{NRP} > p_1^{GRP} > p_1^{TRP}$ and demand is at least as high under NRP than under any other reimbursement regime, it follows unambiguously that $\pi_1^{NRP} > \max\{\pi_1^{GRP}, \pi_1^{TRP}\}$. Regarding the comparison between GRP and TRP, it is not immediately obvious that firm 1 earns higher profits under GRP, since prices are higher, but market shares are lower, compared with TRP. Note, however, that equilibrium prices are higher for *all* firms under GRP, compared with TRP. Furthermore, we know that, for given prices, $c_1^{GRP} < c_1^{TRP}$. Thus, if firm 1 unilaterally deviates from the GRP equilibrium by setting a price equal to the equilibrium price under TRP, this firm will increase its market shares, in both consumer segments, *beyond* its equilibrium market shares under TRP, and consequently earn higher profits than under TRP. Such a deviation is not profitable, so firm 1 must earn even higher profits in the GRP equilibrium, where $p_1^{GRP} > p_1^{TRP}$. □

References

Aronsson, T., Bergman, M.A., Rudholm, N., 2001. The impact of generic drug competition on brand name market shares - evidence from micro data. Review of Industrial Organization 19, 425–435.

Bergman, M.A., Rudholm, N., 2003. The relative importance of actual and potential competition: empirical evidence from the pharmaceuticals market. Journal of Industrial Economics LI, 455–467.

Brekke, K.R., Nuscheler, R., Straume, O.R., 2006a. Quality and location choices under price regulation. Journal of Economics and Management Strategy 15, 207–227.

Brekke, K.R., Grasdal, A., Holmås, T.H., 2006b. Regulation and the pricing of pharmaceuticals: reference pricing or price cap regulation?. Norwegian School of Economics & Business Adm, Mimeo.

Cabrales, A., 2003. Pharmaceutical generics, vertical product differentiation, and public policy. Working Paper Ref. 662, Universitat Pompeu Fabra.

Danzon, P.M., 1997. Pharmaceutical Price Regulation: National Policies versus Global Interests. American Enterprise Institute Press, Washington.

Danzon, P.M., 2001. Reference Pricing: Theory and Evidence. In: Lopez-Casasnovas, G., Jönsson, B. (Eds.), Reference Pricing and Pharmaceutical Policy: Perspectives on Economics and Innovation. Springer, Barcelona (Chapter 5).

Danzon, P.M., Ketcham, J.D., 2004. Reference pricing of pharmaceuticals for Medicare: Evidence from Germany, The Netherlands and New Zealand. In: Cutler, D., Garber, M., A.M. (Eds.), Frontiers in Health Policy Research, vol. 7. National Bureau of Economic Research and MIT Press.

Danzon, P.M., Lui, H., 1996. RP and physician drug budgets: the German experience in controlling pharmaceutical expenditures. Working paper, The Wharton School.

Economides, N., 1989. Quality variations and maximal variety differentiation. Regional Science and Urban Economics 19, 21–29.

Frank, R.G., Salkever, D.S., 1997. Generic entry and the market for pharmaceuticals. Journal of Economics & Management Strategy 6, 75–90.

Gabszewicz, J.J., Thisse, J.F., 1979. Price competition, quality and income disparities. Journal of Economic Theory 20, 340–359.

Gabszewicz, J.J., Thisse, J.F., 1980. Entry (and exit) in a differentiated industry. Economic Journal 96, 160–172.

Grabowski, H.G., Vernon, J., 1992. Brand loyalty, entry, and price competition in pharmaceuticals after the 1984 Drug Act. Journal of Law and Economics 35, 331–350.

Hellerstein, J.K., 1998. The importance of the physician in the generic versus trade-name prescription decision. RAND Journal of Economics 29, 108–136.

Hotelling, H., 1929. Stability in competition. Economic Journal 39, 41–57.

Huskamp, H.A., Rosenthal, M.B., Frank, R.G., Newhouse, J.P., 2000. The Medicare prescription drug program: How will the game be played?. Health Affairs 19 (2), 18–23.

Kanavos, P., 2001. Overview of pharmaceutical pricing and reimbursement regulation in Europe. Commission Of The European Communities, Dg Enterprise, November.

Kanavos, P., Reinhardt, U., 2003. Reference pricing for drugs: Is it compatible with U.S. health care. Health Affairs 22 (3), 16–30.

Königbauer, I., 2006. Advertising and generic market entry. Journal of Health Economics 26 (2), 286–305.

Lopez-Casasnovas, G., Puig-Junoy, J., 2000. Review of the Literature on Reference Pricing. Health Policy 54, 87–123.

Ma, C.-T.A., Burgess Jr., J.F., 1993. Quality competition, welfare, and regulation. Journal of Economics 58 (2), 153–173.

Merino-Castelló, A., 2003. Impact of the reference price system on the pharmaceutical market: a theoretical approach. Working Paper Ref. 524, Universitat Pompeu Fabra.

Pavcnik, N., 2002. Do pharmaceutical prices respond to potential patient out-of-pocket expenses?. RAND Journal of Economics 33 (3), 469–487.

Scherer, F.M., 1993. Pricing, profits and technological progress in the pharmaceutical industry. Journal of Economic Perspectives 7, 97–115.

Scott Morton, F.M., 2000. Barriers to entry, brand advertising, and generic entry in the US pharmaceutical industry. International Journal of Industrial Organization 18, 1085–1104.

Shaked, A., Sutton, J., 1982. Relaxing price competition through product differentiation. Review of Economic Studies 49, 3–13.

Shaked, A., Sutton, J., 1983. Natural oligopolies. Econometrica 51 (5), 1469–1483.

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