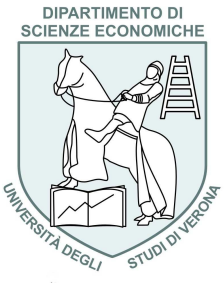

Working Paper Series
Department of Economics
University of Verona

Pricing policies when patients are heterogeneous: a welfare analysis

Rosella Levaggi, Paolo Pertile

WP Number: 17 October 2016

ISSN: 2036-2919 (paper), 2036-4679 (online)

Pricing policies when patients are heterogeneous: a welfare analysis*

Rosella Levaggi† Paolo Pertile‡

Abstract

We use a simple theoretical model to compare alternative regulation regimes for the reimbursement of medical innovations when responses to a new treatment (effectiveness) are heterogeneous within the eligible population. We study two dimensions: *i)* efficiency in selecting sub-groups of patients for which the new technology is reimbursed, *ii)* distribution of the rent between firm and payer. We show that, when rational behaviour of profit maximizing firms is taken into account, stratified cost-effectiveness analysis and *marginal* value-based prices lead to the same equilibrium, which is efficient only if the population is sufficiently homogeneous. Inefficiency arises because some patients that should be treated are not. On the other hand, prices based on the *average* value may allow for an efficient solution even when heterogeneity is large. With this pricing policy, efficiency may be achieved even when part of the rent is retained by the payer, provided that the degree of heterogeneity is sufficiently small.

1 Introduction

The quest to increase *value for money* of innovation in health care has generated huge interest in the study of adoption rules and pricing policies. This is

*We are grateful to the participants to the 17th European Health Economics Workshop, and in particular to the discussant, Lars Schwettmann, for helpful comments. The usual disclaimer applies.

†Department of Economics and Management, University of Brescia.

‡Department of Economics, University of Verona.

crucial in countries where the provision of health care is mainly public, but it is gaining increasing attention also in countries where the private sector plays a prominent role. According to WHO: "Achieving fair pricing and ensuring long-term sustainability of health care systems and access for patients is one of the biggest challenges for health and pharmaceutical systems in Europe and worldwide" (WHO, 2015).

In response to this challenge, an increasing number of countries are introducing into the decision process some form of health technology assessment (HTA), by combining economic inputs with data from clinical trials. While in the past the focus was essentially on the cost-effectiveness of the new technology on the whole eligible population (i.e. it was based on full-sample averages) the attention toward results in different sub-groups has been recently growing. This is also related to recent developments of genomic medicine and the identification of biomarkers to be used as predictors of the effectiveness of a new technology. Personalized medicine is being widely debated as the possible next revolution in medical practice, but it might also provide new opportunities to increase *value for money* through an efficient allocation of specific types of patients to different treatments.

Mainly two regulatory approaches have been discussed in relation to heterogeneity in effectiveness across the eligible population: stratified cost-effectiveness analysis (SCEA) and value-based prices. The former is based on the determination of sub-group specific incremental cost-effectiveness ratios (ICER) and adoption only for those groups for which the new technology is cost-effective. Based on a case study, and assuming fixed prices, Coyle et al. (2003) argue that SCEA is superior to non-stratified cost-effectiveness analysis. Hawkins and Scott (2011) use the same example to show that, if a profit maximizing firm is allowed to propose a price, the outcome of SCEA is suboptimal, because it is dominated by a combination of price and size of population to treat for which both profits and net health benefits are larger. The importance of considering behavioural responses to cost-effectiveness rules by firms is also highlighted by Jena and Philipson (2013) in a slightly different context.

Value-based prices have received huge attention since the Office of Fair Trading recommended the adoption of a value-based price regulation scheme (Office of Fair Trading, 2007), but how this should be exactly implemented is still uncertain (McGuire et al., 2008). Before that, other proposals had been made to link the amount of revenues obtained by the firm to the social value of the innovation (Kremer, 1997; Gravelle, 1998). One point of discussion in

the current debate is whether the price should be based on the value of the new product for the marginal or the average patient. This could make a big difference, especially when the population is very heterogeneous. So far, most of the attention has been on *marginal* value-based prices (MVBP), mainly because these would allow the payer to retain part of the rent (Claxton, 2007). The way the rent is split between the industry and the payer is indeed crucial, because it determines the balance between static efficiency – making drugs accessible to all those who need them – and dynamic efficiency – ensuring that firms' profits are robust enough to sustain R&D investments (Leibenstein, 1966). Danzon et al. (2015) adopt a multi-country perspective and show that MVBP are *"roughly consistent with second-best static and dynamic efficiency"*(Danzon et al., 2015, p. 294).

We aim to combine some of the features of previous studies that we believe are important, such as the active role of the firm in price setting (Hawkins and Scott, 2011) and the analysis of the balance between static and dynamic efficiency (Danzon et al., 2015) within a simple, but still general, theoretical framework. Moreover, we study how the degree of heterogeneity in the population influences the properties of the alternative schemes. We are not aware of other theoretical analyses sharing this feature. Our model provides a theoretical framework to generalise some of the results from the previous literature and to define under which conditions those results (do not) hold. Finally, we explore the properties of a regulatory scheme where prices are defined as a fraction (possibly, but not necessarily, equal to one) of the *average* value for patients treated (AVBP), as an alternative to SCEA and MVBP.

We show that the outcome under SCEA and MVBP is the same and it is not Pareto efficient, unless effectiveness is sufficiently homogeneous within the eligible population. On the other hand, AVBPs allow for an efficient solution, irrespective of the degree of heterogeneity in the population. However, with sufficiently heterogeneous populations, an efficient solution requires the whole rent to be left to the firm. The fraction of rent that must be left to the firm is non-decreasing in the degree of heterogeneity.

Section 2 introduces the model. In Section 3 we define an efficiency benchmark by characterizing the first-best solution, which is then compared with stratified cost-effectiveness analysis (Section 4), *marginal* value-based prices (Section 5) and *average* value-based prices (Section 6). Section 7 explores how the rent is distributed between the firm and the payer under the three schemes considered and Section 8 provides some concluding remarks.

2 The model

A new technology is considered for reimbursement. The size of the eligible population is 1, but the treatment of $n \in [0,1]$ patients is reimbursed. It is assumed that only patients whose treatment is reimbursed receive it and that reimbursement, if granted, covers the whole price. Let the marginal health benefit of the new technology be:

$$MB(n) = (1-z) + (2z-1)n, \tag{1}$$

where $z \in \left[0, \frac{1}{2}\right)$ is the marginal effectiveness of the 'last' patient, i.e. $z = MB(1)$. To simplify the analysis, but without loss of generality, we assume that no (active) treatment is currently available for this population. This implies that incremental values of benefits and costs equal absolute values.

The definition of the marginal benefit function as in Eq. 1 allows us to study the role of heterogeneity in effectiveness within the eligible population by letting z change, while keeping average effectiveness constant and equal to $\frac{1}{2}$. The degree of heterogeneity is inversely related to the value of z. For $z \to \frac{1}{2}$ the MB tends to a constant value, while $z = 0$ corresponds to the maximum degree of heterogeneity. The lower bound on z means that the marginal benefit of the technology is non-negative for the whole eligible population.

Figure 1 clarifies the relationship between the value of z and the degree of heterogeneity: z_1 and the corresponding MB function (solid line) corresponds to the case of comparatively large heterogeneity, because the difference between the effectiveness of the "first" patient ($\lim_{n\to 0^+} MB(n)$) and the "last" patient ($\lim_{n\to 1} MB(n)$) is large. When $z = z_2$ (dashed line) this gap is smaller and the population is comparatively homogeneous.

Let $c \in \mathbb{R}^+$ and $p \in \mathbb{R}^+$ be respectively the marginal cost to produce the drug and its price. The technology is provided by a firm whose objective is the maximization of profits, defined as:

$$\Pi(p,n) = (p-c)n. \tag{2}$$

The marginal health benefit is turned into a monetary measure by multiplying Eq. 1 by $\lambda \in \mathbb{R}^+$, the shadow value of health. Hence, the total expected health gain in monetary terms, as a function of n, is $\lambda \int_0^n [(1-z) + (2z-1)s]\, ds$; this is also the incremental money health gain, given the assumption that there are no other drugs available for the treatment of these patients.

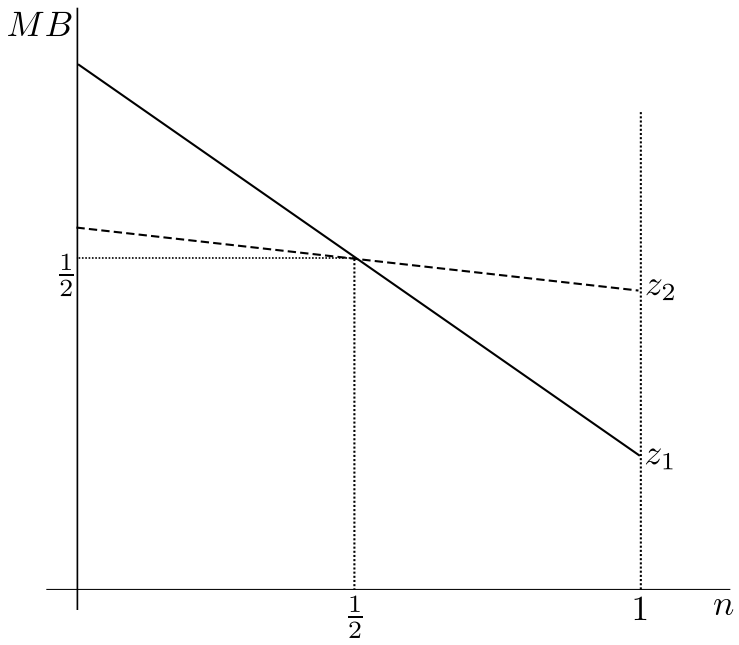

Figure 1: Marginal benefit as a function of the size of population treated for different values of z ($z_2 > z_1$).

The consumer surplus can be written as:

$$CS(p, n) = \lambda \int_0^n [(1 - z) + (2z - 1)s] \, ds - p \cdot n. \tag{3}$$

Eq. 3 defines the monetary value of the health gain from treating n patients with the new technology, net of the amount paid to the firm.[1]

3 First Best

As is typical in the economic analysis of regulation, we start by defining a benchmark, corresponding to the case where a benevolent social planner maximises welfare. We define this as a log-linear function of consumer surplus

[1]We refer to Eq. 3 as consumer surplus by applying the standard definition. However, it might be argued that the actual surplus enjoyed by consumers (patients) should account for the fact that they benefit from insurance, meaning that the price actually paid is less than the price paid to the firm (Lakdawalla and Sood, 2009).

(Eq. 3) and profit (Eq. 2):

$$W(p,n) = \alpha \ln[CS(p,n)] + (1-\alpha)\ln[\Pi(p,n)], \tag{4}$$

where $\alpha \in [0,1]$ is the relative weight of consumer surplus in the welfare function. In particular, for $\alpha = 1$ only consumer surplus matters, while at the other extreme ($\alpha = 0$) welfare maximisation corresponds to profit maximisation. The definition of a welfare function allows for an explicit characterization of the trade-off between two key objectives: profits for the firm, which in turn produce incentives to R&D investment, and *value for money* of adoption decisions. Moreover, the marginal contribution to welfare of increases in either dimension is decreasing.

The first-best (FB) solution is determined by maximizing Eq. 4 with respect to p and n, under the constraint on the maximum size of the population to treat ($n \leq 1$). The solution is:

$$\begin{cases} n^{fb} = 1; \quad p^{fb} = \frac{1}{2}\left[2c\alpha + \lambda(1-\alpha)\right] & \text{if} \quad z \geq \frac{c}{\lambda} \\ n^{fb} = \frac{\lambda(1-z)-c}{\lambda(1-2z)}; \quad p^{fb} = \frac{1}{2}\left[c(1+\alpha) + (1-z)(1-\alpha)\lambda\right] & \text{if} \quad z < \frac{c}{\lambda}. \end{cases} \tag{5}$$

We assume that $\frac{c}{\lambda} < \frac{1}{2}$, which ensures the existence of a range of values for z such that in the FB the whole population is treated. The economic intuition for the condition under which it is optimal to treat the whole population ($z \geq \frac{c}{\lambda}$) is straightforward: the marginal benefit of the last patient in the eligible population must exceed the health equivalent of the marginal cost. It is interesting to note that n^{fb} is independent of α, meaning that, irrespective of the relative weight of the two components in Eq. 4, social welfare maximization requires that the size of the population treated is (static) efficient, and the desired distribution of the rent is achieved by adjusting the price.

Typically, regulators are unable or unwilling to directly fix both price and size of the population to treat, meaning that the FB solution cannot be directly implemented. In what follows we investigate the welfare properties of regulation schemes that are under discussion (Sections 4 and 5) or feasible (Section 6).

4 Stratified cost-effectiveness analysis

In this section, we discuss a regulation regime based on SCEA and the following sequence of actions:

1. the regulator announces that the treatment will be reimbursed only for patients for whom the ICER is below the threshold λ, given the price. It is assumed that the regulator can commit to this rule.

2. The firm knows the rule and sets the profit maximizing price.

The rule described in point 1 means that the treatment for the n^{th} patient will be reimbursed if and only if:

$$\frac{p}{(1-z)+(2z-1)n} \leq \lambda. \tag{6}$$

Hence, the number of patients treated, given p, is:

$$n^s(p) = \begin{cases} 1 & \text{if} \quad p \leq z\lambda \\ \frac{p-\lambda(1-z)}{\lambda(2z-1)} & \text{if} \quad p > z\lambda. \end{cases} \tag{7}$$

The following proposition defines an important property of SCEA:

Proposition 1 *Adoption based on the results of Stratified Cost-Effectiveness Analysis is an optimal policy for a regulator willing to maximize consumer surplus, if prices are exogenous.*

Proposition 1 follows immediately from checking that Eq. 7 is also the solution to the following problem:

$$\max_n \quad \lambda \int_0^n \left[(1-z)+(2z-1)s\right] ds - p \cdot n \quad \text{s.t.} \quad n \in [0,1]. \tag{8}$$

i.e. it corresponds to the maximization of CS with p exogenous, under the constraint on the size of the eligible population.

Proposition 1 generalizes and extends the results by Coyle et al. (2003) by showing that if prices are fixed, SCEA is not only better than non-stratified cost-effectiveness analysis, but it is the policy that maximizes consumer surplus (net health benefits). Their analysis is based on a case study with eight sub-groups of patients, each of which has a different level of effectiveness. The continuous case that we consider can be seen as a limit case of the discrete one, and it allows to simplify the generalization of results.

Let us now move to the study of the optimal behaviour of a firm facing the adoption rule defined by Eq. 7. Since $z < \frac{1}{2}$, the second line of Eq. 7 defines a downward sloping demand curve that the firm will take into account in defining its optimal pricing strategy. In particular, this demand function corresponds to the MB function, expressed in monetary terms. The firm solves:

$$\max_{p} \quad n^s(p)(p-c), \tag{9}$$

where $n^s(p)$ is defined in Eq. 7. The solution is:

$$p^s = \begin{cases} z \cdot \lambda & \text{if} \quad z \geq \mathring{z} \\ \frac{1}{2}\left[c + \lambda(1-z)\right] & \text{if} \quad z < \mathring{z}, \end{cases} \tag{10}$$

with $\mathring{z} = \frac{1}{3}\left(\frac{c}{\lambda} + 1\right)$. The equilibrium solution is then:

$$\begin{cases} n^s = 1; \quad p^s = z \cdot \lambda & \text{if} \quad z \geq \mathring{z} \\ n^s = \frac{1}{2}\left[\frac{c-\lambda(1-z)}{\lambda(2z-1)}\right]; \quad p^s = \frac{1}{2}\left[c + \lambda(1-z)\right] & \text{if} \quad z < \mathring{z}. \end{cases} \tag{11}$$

According to the first line of Eq. 11, if z is sufficiently large relative to c, the firm will find it optimal to set $p = z \cdot \lambda$ and treat the whole population. Note that the assumption $\frac{c}{\lambda} < \frac{1}{2}$ implies $\frac{c}{\lambda} < \mathring{z} < \frac{1}{2}$. Recalling that for $z > \frac{c}{\lambda}$ the FB solution entails the adoption of the new technology for the whole population (first line of Eq. 5), this means that when the effectiveness of the new treatment is sufficiently homogeneous across patients, basing adoption decisions on SCEA leads to a FB solution even if the firm is free to set the price.

When the second line of Eq. 11 is relevant, the size of the population treated is no longer efficient. In particular, it corresponds to half the FB population size. The firm will act as a monopolistic price setter and will then aim at equalizing marginal revenue with marginal cost. Recalling that, given the optimal policy for the regulator defined by Eq. 7, the demand function faced by the firm corresponds to the money equivalent of the marginal benefit function, a comparatively low value of z implies a more inelastic demand function, which allows the firm to exploit some monopolistic power in price setting. Given the linearity of the demand (marginal benefit) function, the standard result that the absolute value of the slope of the marginal revenue function is twice the value of the slope of the demand function applies, and the size of the population treated is half the FB value.

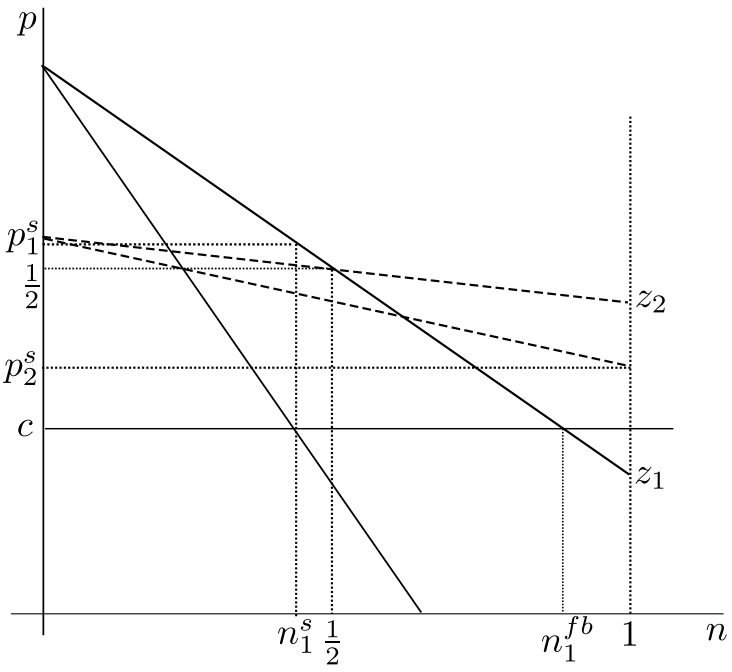

Figure 2: Price and number of patients treated in equilibrium under SCEA, for different degrees of heterogeneity (solid: high heterogeneity; dashed: low heterogeneity).

The implications of the above results are summarized in the following proposition:

Proposition 2 *If adoption decisions are based on stratified cost-effectiveness analysis and the firm is free to set the price, the number of patients treated is inefficiently low, unless responses are sufficiently homogeneous within the eligible population.*

Figure 2 provides a graphic illustration. The two solid lines represent marginal benefit (higher curve) and marginal revenue (lower curve) for a situation with comparatively large heterogeneity ($z = z_1$). In this case, the firm faces a comparatively inelastic demand curve. As a result of the application of the standard rule of equalization between marginal revenue and marginal cost, the resulting equilibrium is characterized by $n_1^s < n_1^{fb}$ and p_1^s. With less heterogeneity ($z = z_2$) marginal benefit and revenue correspond to the two dashed lines. In this case, the firm finds it optimal to set a price

($p_2^s = z \cdot \lambda$) such that the treatment is cost-effective for the whole population ($n^s = 1$), as in the FB solution.

Hawkins and Scott (2011) move from the results of Coyle et al. (2003) and use the same example to show that SCEA is not necessarily efficient. In particular, they show that there exists a combination of price and size of population to treat with the new technology such that both profits and the total net benefit in the population increases. A key difference with Coyle et al. (2003) is that they allow the firm to set the price, as we do. Hawkins and Scott (2011) show that in their example there exists a value of the price, lower than the price the provider would optimally set, given an adoption rule based on SCEA, such that both firm's profits and patients' total net benefit are higher. This is a special case of our framework, for which we have shown that if heterogeneity in individual effectiveness across population is sufficiently large ($z < \mathring{z}$), SCEA leads to an inefficient outcome. The fact that it is possible to find an alternative combination of p and n such that both parties are better off, i.e. it is a Pareto improvement, is an immediate implication. A further contribution from our model is the result that for a given cost of provision of the treatment, c, this is no longer true when patients are sufficiently homogeneous.

Hawkins and Scott (2011) conclude that '*stratified cost-effectiveness analysis may not be the last word*'. In Section 6 we will explore the possibility of finding a regulatory solution to ensure that the FB is achieved, irrespective of the degree of patient heterogeneity.

5 *Marginal* value-based prices

The scheme is characterized in the following way:

1. the regulator defines a pricing rule such that the price equals the monetary value of the benefit of the treatment for the marginal patient:

$$p^m(n) = \lambda \left[(1 - z) + (2z - 1)n \right]. \tag{12}$$

   It is assumed that the regulator can commit to this rule.

2. Knowing the pricing rule defined in Eq. 12, the firm sets n in order to maximize profits.

Assuming that the firm plays a role in the definition of n looks reasonable in this case. In practice, this will take place at the time when the firm decides which indications to propose for reimbursement. The firm's problem is:

$$\max_{n} \quad n(p^m(n) - c) \quad s.t. \quad n \in [0, 1]. \tag{13}$$

The equilibrium solution is:

$$\begin{cases} n^m = 1; \quad p^m = z \cdot \lambda & \text{if} \quad z \geq \mathring{z} \\ n^m = \frac{1}{2}\left[\frac{c-\lambda(1-z)}{\lambda(2z-1)}\right]; \quad p^m = \frac{1}{2}\left[c + \lambda(1-z)\right] & \text{if} \quad z < \mathring{z}, \end{cases} \tag{14}$$

which is identical to the one obtained for SCEA (Eq. 11). This is because under MVBP the price per unit of effectiveness (λ) corresponds to the ICER threshold under SCEA, meaning that, given n, the product will be always cost-effective at the margin (i.e. for the last stratum treated, in a discrete framework). Hence, under our assumptions, the fact that under SCEA the firm sets n and under MVBP it sets p is irrelevant for the determination of the equilibrium. This allows us to extend the results of Section 4 to MVBP:

Proposition 3 *With Marginal Value-Based Prices, if the firm can decide the size of the population to treat, the number of patients treated is inefficiently low, unless responses are sufficiently homogeneous across the eligible population.*

One argument in favour of MVBP is its ability to extract rent from the producer. Our results show that this gain for the payer may come at the price of static efficiency, due to a suboptimal size of the population treated and a high price. Danzon et al. (2015) suggest that static and dynamic efficiency could be enhanced with respect to MVBP if the payer could vary prices by subgroup to reflect the specific level of effectiveness. This form of price-discrimination is equivalent to a price based on the average value of effectiveness for the population treated (Claxton, 2007). The efficiency properties of a similar approach will be investigated in the following section, whereas Section 7 will study its redistributive properties.

6 *Average* Value Based Prices

In this section, we consider the implications of introducing AVBP according to the following sequence of actions:

1. the regulator defines a pricing rule such that the price equals a fraction $\rho \in (0, 1]$ of the monetary value of the mean benefit of the treatment within the population treated:

$$p^a(n) = \frac{\rho\lambda}{n} \int_0^n [(1 - z) + (2z - 1)s] \, ds. \tag{15}$$

   It is assumed that the regulator can commit to this rule.

2. Knowing the pricing rule, the firm sets n in order to maximize profits.

The problem solved by the firm is:

$$\max_n \quad n(p^a(n) - c) \quad s.t. \quad n \in [0, 1]. \tag{16}$$

The equilibrium solution is:

$$\begin{cases} n^a = 1; \quad p^a = \frac{\lambda\rho}{2} & \text{if} \quad z \geq \frac{c}{\rho\lambda} \\ n^a = \frac{\rho\lambda(1-z)-c}{\rho\lambda(1-2z)}; \quad p^a = \frac{1}{2}\left[c + \lambda\rho(1 - z)\right] & \text{if} \quad z < \frac{c}{\rho\lambda}. \end{cases} \tag{17}$$

It is easy to see that with $\rho = 1$, $n^a = n^{fb}$. In other words, the regulator can always implement a solution involving an efficient size of the population to treat, even when the decision on the breath of the indication is decided by the firm, by setting the price equal to the average monetary value of benefit for patients:

Proposition 4 *With prices equal to the average value for the population treated, the number of patients treated is efficient, irrespective of the degree of heterogeneity in the population.*

A price equal to the average value means that the whole rent is left to the firm, which in turn implies dynamic efficiency (see e.g. Tirole (1988)). Hence, this situation ensures both static and dynamic efficiency.

In the FB, the whole eligible population is treated if $z > \frac{c}{\lambda}$. Therefore, when z is sufficiently large, it is possible to ensure $n^a = n^{fb}$ even with values of ρ smaller than one. This observation will be developed further in the next section.

7 Distributional issues

The analysis of distributional issues is closely related to some key aspects of the policy debate. On the one hand, there is growing pressure on insurers to keep prices as low as possible and curb health care expenditure. On the other hand, the concern that this may jeopardize incentives for the industry to invest in R&D to bring further innovation to the market in the future (dynamic efficiency) is also growing (Giaccotto et al., 2005; Civan and Maloney, 2009; Blind, 2012).

In the three previous sections we focussed on static efficiency and showed that a FB solution can be achieved with SCEA and MVBP only when the response to the treatment is sufficiently homogeneous, whereas it is always achievable with AVBP. In both cases, there are constraints that the price must satisfy for the FB solution to be achieved. In this section, we investigate how surplus is distributed between firm and payer under the alternative regulatory frameworks in those situations where an efficient solution can be achieved. According to Eq. 5, the FB prices, unlike the FB population sizes, depend on α. The question we aim to answer in this section is: for which values of α are those prices equal to the equilibrium prices obtained for the different schemes, conditional upon achieving a FB solution? This is going to inform about dynamic efficiency implications of FB solutions implemented through the three regulatory schemes.

7.1 Stratified cost-effectiveness analysis and *marginal* value-based prices

Given that SCEA and MVBP lead to the same equilibrium, they can be studied together. In this case, the firm is free to set p (under SCEA) or n (under MVBP), given the rule announced by the regulator. The combination of Eq. 5, 11 and 14 means that a FB is achieved if and only if $z \geq \mathring{z}$; the corresponding price is $p = z \cdot \lambda$. By equating this price to the FB price for this range of values of z (first line of Eq. 5) it is possible to define the corresponding value of α:

$$\hat{\alpha} = \frac{\lambda(2z-1)}{2c-\lambda}, \tag{18}$$

which is linearly dependent and decreasing in z. In this case, there is a single price compatible with the FB, meaning that there will also be a single value of $\alpha = \hat{\alpha}$. For $z \to \frac{1}{2}$, $\hat{\alpha} \to 0$; for the minimum value of z compatible with a

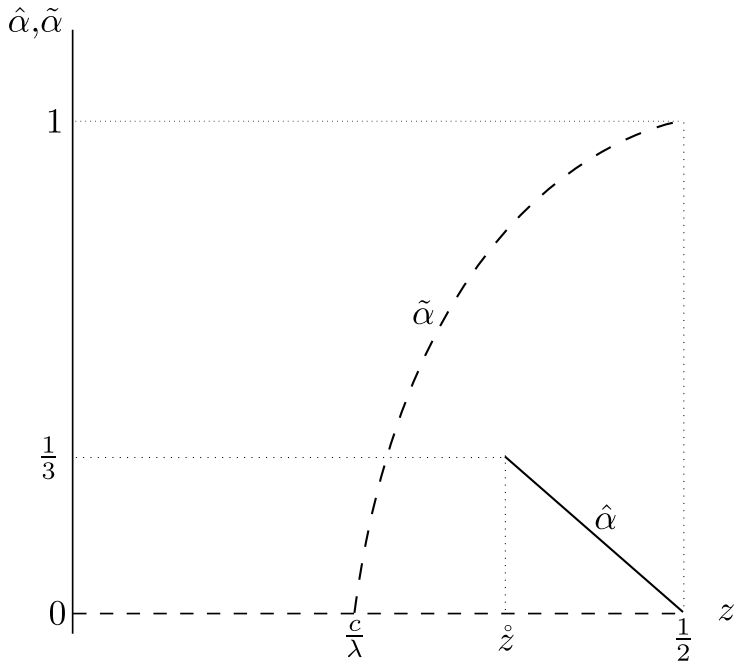

Figure 3: Ranges of values α corresponding to FB solutions under decentralized solution (solid line) and VB prices (area between dashed lines), as a function of z.

FB ($z = \mathring{z}$), $\hat{\alpha} = \frac{1}{3}$. The relationship between $\hat{\alpha}$ and z is illustrated in Figure 3 (solid line). Note that $\hat{\alpha}$ is only defined for $z > \mathring{z}$, because a FB can only be achieved in this range.

7.2 *Average* value-based prices

Let us define $\tilde{\rho}$ as the minimum value of ρ such that $n^{vb} = n^{fb}$. It follows immediately from Eq. 17 and Eq. 5 that $\tilde{\rho} = 1$ for $z \leq \frac{c}{\lambda}$. For $z > \frac{c}{\lambda}$, all values of ρ such that $z \geq \frac{c}{\rho\lambda}$ imply $n^{vb} = n^{fb} = 1$. Therefore,

$$\tilde{\rho} = \begin{cases} \frac{c}{\lambda z} & \text{if} \quad z > \frac{c}{\lambda} \\ 1 & \text{if} \quad z \leq \frac{c}{\lambda}. \end{cases} \tag{19}$$

The achievement of a FB with $\rho < 1$ means that, unlike in the case of Section 7.1 and unlike with AVBP when heterogeneity is comparatively large, there exists a range, not a single price, compatible with the FB. By replacing

$\tilde{\rho}$ into p^{vb} as defined in the first line of Eq. 17 and equating this to p^{fb} as defined in the first line of Eq. 5, it is possible to define $\tilde{\alpha}$ as the maximum value of α corresponding to a FB implemented through AVBP. In particular, it can be shown that, for $z > \frac{c}{\lambda}$,

$$\tilde{\alpha} = \frac{c - \lambda z}{z(2c - \lambda)}. \tag{20}$$

Eq. 20 shows that $\tilde{\alpha}$ is strictly increasing in z for $z > \frac{c}{\lambda}$ and $\tilde{\alpha} \to 1$ as $\tilde{z} \to \frac{1}{2}$. This means that in the limit case of a perfectly homogeneous population, the FB can be achieved even with a price equal to the marginal cost, i.e. extracting the whole rent from the firm.

The situation is illustrated in Figure 3. The upper dashed line is $\tilde{\alpha}$. Recalling that $\rho = 1$, i.e. $\alpha = 0$, always leads to a FB solution under AVBP, the area between the two dashed lines defines the range of values of α compatible with a FB. For $z \leq \frac{c}{\lambda}$, i.e. when heterogeneity in effectiveness is comparatively large and it is efficient to treat only part of the population ($n^{fb} < 1$), this is only possible if $\alpha = 0$. In the limit case of a perfectly homogeneous population ($z \to \frac{1}{2}$) the regulator can extract the whole rent from the firm ($\alpha = 1$).

Proposition 5 *With average value-based prices, the fraction of surplus that needs to be left to the firm to achieve efficiency is non-decreasing in the degree of heterogeneity of patients.*

It is interesting to see that over the range of values where both SCEA/MVBP and AVBP lead to a FB solution, an increase in the degree of homogeneity (z higher) has opposite effects in terms of rent distribution: in that case, more homogeneity will be preferred by the regulator under AVBP, and by the firm under SCEA/MVBP.

8 Conclusion

There seems to be growing attention toward the correlation between observable patient characteristics and effectiveness of new treatments. This phenomenon, which is also related to the rapid development of genomic medicine, has both clinical and economic implications. On the economic side, this tendency stimulated some research on the efficiency of common approaches to

market access regulation, when effectiveness varies across sub-groups of patients.

We let the firm play an active role in defining either the price or the size of the population to treat and study the resulting equilibria under SCEA and MVBP. We show that, unless the population is sufficiently homogeneous, heterogeneity across patients provides the firm with monopolistic power, leading to an equilibrium where the number of patients treated is inefficiently low. Our model allows to generalize the result by Hawkins and Scott (2011) that the outcome of SCEA is not Pareto efficient, and defines conditions under which that result does not hold. In particular, we find that if the population is sufficiently homogeneous, the number of patients treated with the new technology is efficient.

We then study the properties of a value-based price based on the average effectiveness in the population treated. We show that this tool is superior to SCEA and MVBP in terms of static efficiency, because it allows for an efficient number of patients treated, no matter what the degree of heterogeneity in the population is. However, when the population is sufficiently heterogeneous, this can only be achieved if the whole rent is left to the firm. This implies dynamic efficiency, but also a potential threat of higher expenditure for the payer. Whether it would be feasible to adjust the price to the degree of heterogeneity in the population is an open question. However, our model highlights a serious limitation of SCEA and MVBP in terms of static efficiency when patients are heterogeneous and firms play a role in setting prices or indications for new pharmaceutical products.

Overall, our results suggest that the degree of heterogeneity in the eligible population and firms' optimal reactions cannot be neglected in assessing efficiency and distributional implications of alternative regulatory schemes for market access. Hopefully, the simple model proposed will be used and further developed to investigate other schemes in addition to those discussed so far and deployed by policy makers. The model could clearly be enriched to take several additional aspects into account. For example, one could explicitly introduce R&D costs in early stages of a dynamic model of innovation discovery, development and commercialization, or address the issue of who benefits the most from the possibility of personalizing treatments.

References

Blind, K. (2012). The influence of regulations on innovation: A quantitative assessment for OECD countries. *Research Policy*, 41(2):391–400.

Civan, A. and Maloney, M. T. (2009). The effect of price on pharmaceutical R&D. *The BE Journal of Economic Analysis & Policy*, 9(1).

Claxton, K. (2007). OFT, VBP: QED? *Health economics*, 16(6):545–558.

Coyle, D., Buxton, M. J., and O'Brien, B. J. (2003). Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Economics*, 12(5):421–427.

Danzon, P., Towse, A., and Mestre-Ferrandiz, J. (2015). Value-based differential pricing: Efficient prices for drugs in a global context. *Health economics*, 24(3):294–301.

Giaccotto, C., Santerre, R. E., and Vernon, J. A. (2005). Drug prices and research and development investment behavior in the pharmaceutical industry. *Journal of Law and Economics*, 48(1):195–214.

Gravelle, H. S. (1998). Ex post value reimbursement for pharmaceuticals. *Medical decision making*, 18(2):S27.

Hawkins, N. and Scott, D. A. (2011). Reimbursement and value-based pricing: stratified cost-effectiveness analysis may not be the last word. *Health Economics*, 20(6):688–698.

Jena, A. B. and Philipson, T. J. (2013). Endogenous cost-effectiveness analysis and health care technology adoption. *Journal of health economics*, 32(1):172–180.

Kremer, M. (1997). Patent buy-outs: A mechanism for encouraging innovation. Working Paper 6304, National Bureau of Economic Research.

Lakdawalla, D. and Sood, N. (2009). Innovation and the welfare effects of public drug insurance. *Journal of Public Economics*, 93(3-4):541–548.

Leibenstein, H. (1966). Allocative efficiency vs."X-efficiency". *The American Economic Review*, 56(3):392–415.

McGuire, A., Raikou, M., and Kanavos, P. (2008). Pricing pharmaceuticals: Value based pricing in what sense? *Eurohealth*, 14(2):3–5.

Office of Fair Trading (2007). The pharmaceutical price regulation scheme. an OFT market study. Technical report, WHO, European Regional Office.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

WHO (2015). Access to new medicines in Europe: technical review of policy initiatives and opportunities for collaboration and research. Technical report, WHO, European Regional Office.