VALUE-BASED DIFFERENTIAL PRICING: EFFICIENT PRICES FOR DRUGS IN A GLOBAL CONTEXT

PATRICIA DANZON\textsuperscript{a, *}, ADRIAN TOWSE\textsuperscript{b} and JORGE MESTRE-FERRANDIZ\textsuperscript{b}

\textsuperscript{a}The Wharton School, University of Pennsylvania, Philadelphia, PA, USA
\textsuperscript{b}Office of Health Economics, London, UK

ABSTRACT

This paper analyzes pharmaceutical pricing between and within countries to achieve second-best static and dynamic efficiency. We distinguish countries with and without universal insurance, because insurance undermines patients' price sensitivity, potentially leading to prices above second-best efficient levels. In countries with universal insurance, if each payer unilaterally sets an incremental cost-effectiveness ratio (ICER) threshold based on its citizens' willingness-to-pay for health; manufacturers price to that ICER threshold; and payers limit reimbursement to patients for whom a drug is cost-effective at that price and ICER, then the resulting price levels and use within each country and price differentials across countries are roughly consistent with second-best static and dynamic efficiency. These value-based prices are expected to differ cross-nationally with per capita income and be broadly consistent with Ramsey optimal prices. Countries without comprehensive insurance avoid its distorting effects on prices but also lack financial protection and affordability for the poor. Improving pricing efficiency in these self-pay countries includes improving regulation and consumer information about product quality and enabling firms to price discriminate within and between countries. © 2013 The Authors. Health Economics published by John Wiley & Sons Ltd.

Received 12 June 2012; Revised 11 July 2013; Accepted 6 November 2013

KEY WORDS: differential pricing; ICER thresholds; value-based pricing

1. INTRODUCTION

Achieving efficient pricing of pharmaceuticals between and within countries is a complex conceptual and policy problem. In any industry, pricing to maximize social welfare must consider both static efficiency (optimal use of existing products) and dynamic efficiency (optimal investment in research and development [R&D]). Reconciling these objectives is problematic for pharmaceuticals, for three reasons.

First, R&D is roughly 17\% of sales for the US-based pharmaceutical industry, compared with 4\% for other US industries, and other quasi-fixed costs of production are significant. Marginal cost pricing to achieve first-best static efficiency would fail to cover total costs and violate the dynamic efficiency requirement that producers capture the full social surplus produced by innovation. Patents enable firms to price above marginal cost and thus potentially achieve dynamic efficiency. This is 'second best' if pricing above marginal cost reduces utilization.

Second, in the case of pharmaceuticals, the effect of patents is both mitigated and distorted by insurance coverage in most industrialized countries. By lowering out-of-pocket prices to patients, insurance potentially brings utilization closer to first-best levels. However, by making patient demand highly price-inelastic, insurance creates the potential and incentives for manufacturers to set prices above second-best optimal levels. By contrast, patients in self-pay markets (including many middle and lower income countries [MLICs]) lack the financial protection of insurance but also avoid its distorting effects on prices. However, other factors—including uncertain product quality and skewed income distributions—contribute to drug prices that may exceed second-best optimal levels (Flynn \textit{et al.}, 2009; Danzon \textit{et al.}, 2011).

*Correspondence to: Wharton Health Care Systems Department, University of Pennsylvania, 3641 Locust Walk, Philadelphia, PA, 19104, USA. E-mail: danzon@wharton.upenn.edu

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2013 The Authors. Health Economics published by John Wiley & Sons Ltd.
Third, because R&D is a global joint cost benefiting consumers worldwide, efficient global pricing requires appropriate contributions from different countries to this joint cost. Economic theory suggests that price discrimination across countries is welfare superior to uniform pricing, assuming drug utilization increases, which is plausible. Price discrimination and Ramsey pricing theories give criteria for an efficient structure of relative prices but do not address optimal absolute price levels.

Previous literature has addressed components of a theory of efficient drug pricing within a single country (Garber et al., 2006; Claxton et al., 2008; Jena and Philipson, 2008; Lakdawalla and Sood, 2009) and efficient price differentials across countries (Malteg and Schwartz, 1994; Danzon, 1997; Danzon and Towse, 2003; Jack and Lanjouw, 2005; Barros and Martinez-Giralt, 2008).

In this paper, we outline a comprehensive approach to global pricing that simultaneously achieves second-best static and dynamic efficiency within and between countries, given certain assumptions. Specifically, we show that each payer can achieve appropriate prices by setting an incremental cost-effectiveness ratio (ICER) threshold and letting companies choose prices, given their drug’s effectiveness and the payer limiting reimbursement to utilization that meets the ICER threshold. These indirect constraints provide a practical approach to achieve the optimal price and utilization in each country, which we call ‘value-based differential pricing’ (VBDP), because prices would reflect incremental value as perceived in each country. We also outline conditions for optimal pricing in self-pay markets.

2. VALUE-BASED DIFFERENTIAL PRICING: ONE COUNTRY, COMPREHENSIVE INSURANCE

2.1. Optimal incremental cost-effectiveness ratio threshold for a representative consumer: exogenous prices and technology

Our starting point for considering optimal pricing and utilization in a single country is Garber and Phelps (1997) (GP)’s model of an uninsured individual’s optimal allocation of a fixed budget between medical care and other services, treating the availability and price of medical technologies as exogenous.

Assume that each country can be represented by an individual with income \( Y \) that is constant in real terms across time periods. Following GP, period-specific utility of income as viewed from period 0 is \( v = U_0(Y) \), before discounting or quality of life adjustment. Income is spent on medical technologies \( a \) and \( b \), with prices \( w_a \) and \( w_b \), respectively, which for now are assumed to be exogenous, and on other goods. Consumption of medical care in period 0 affects probability of survival and quality of life in future periods. Future utility is discounted by a factor \( \rho \). The expected benefits of medical care can thus be expressed as the sum of discounted expected future quality-adjusted life years (QALYs), \( Q_i \). Expected utility in period 0 can be written

\[
E_0 = U_0(Y - w_a a - w_b b) + v \sum_i \rho^i Q_i(a, b)
\]

The optimal utilization of technology \( a \) using Eq. (1) is defined by the first-order condition:

\[
w_a U'_0 = v \frac{dQ}{da}
\]

Optimal utilization thus implies equating the marginal utility cost of spending on \( a \) in period 0, \( w_a U'_0 \), to the marginal expected utility of future discounted QALYs gained from using \( a \), \( v \frac{dQ}{da} \). Rewriting,

\[
\frac{w_a}{(dQ/da)} = \frac{v}{U'_0}
\]

In Eq. (3), the left hand side (LHS) is technology \( a \)’s ICER, assuming no alternative treatment. The numerator is the incremental cost of using \( a \), which here is just \( w_a \), but it could include other medical costs \( e_a \) if use of \( a \) affects

1Our analysis does not address ‘neglected disease’ drugs, for which global demand is insufficient to incentivize private R&D.

2For a country-specific definition of value-based pricing, see OFT (2007).
other services due to complementarity or substitution. The denominator is the expected QALYs gained from using \( a \). The right hand side (RHS) is the ratio of future, period-specific utility \( v \) to marginal utility in the base period, or willingness-to-pay (WTP) for medical care. Optimal utilization thus requires equating the technology’s ICER to the consumer’s WTP for medical care.

2.2. Application to a universal payer: endogenous prices and utilization

Assume that each country operates a universal insurance system including drugs for all citizens, to provide financial protection and reflect altruistic concerns for access for the poor. Prices charged by manufacturers and technology availability are now endogenous and influenced by insurance design and payer strategies. The analysis can still focus on the representative consumer, assuming that political processes constrain the universal insurance design to reflect consumer preferences for altruism and personal treatment options. A

Reasonable financial protection for patients requires that cost-sharing is modest and capped. Manufacturers therefore face relatively inelastic demand which, in the absence of constraints, could lead to prices above the patent-induced level without insurance. Assume raising funds is costly, and hence, the payer seeks to achieve second rather than first-best static and dynamic efficiency.

In this context, Eq. 3 implies that the payer can indirectly control prices by setting an ICER threshold \( K \) (e.g., £30,000 per QALY) that reflects its citizens’ WTP for medical care and limiting reimbursement to products/utilization that meet this threshold:

\[
\text{ICER} = \frac{(w_a - w_0) + (c_a - c_0)}{dQ} \leq K
\]

We assume a manufacturer is permitted to price freely, subject to this constraint. It would set its price differential over current treatment at the highest level consistent with the ICER threshold, given the product’s incremental cost offsets and effectiveness gain:

\[
w_a = w_0 + dc + KdQ
\]

Thus, a new product with no incremental benefit would be constrained to price at \( w_0 \). A more effective product or one that substitutes for other services could be priced higher and still meet the ICER threshold. The payer’s ICER threshold acts as an indirect control on price, given the product’s incremental effectiveness and cost offsets.

Given the manufacturer’s choice of price, the payer can achieve appropriate use by limiting coverage to those patients for whom the product is cost-effective at this price and ICER threshold. Some products may have a distribution of effectiveness across patient subgroups with different indications or severities. If the firm chooses a high price, the payer would restrict use to patients whose condition/indication implies an expected health benefit sufficient to meet the ICER threshold. A lower price enables the payer to encourage use by patient subgroups with lower expected benefit. The firm thus faces a price-volume trade-off similar to a demand curve. In a single price system, the firm selects the profit maximizing price, given the use that the payer would permit at that price. This outcome is second-best efficient. Static and dynamic efficiency could be enhanced and, at the limit, would be first best, if the firm could vary prices by indication/subgroup, reflecting the drug’s differential effectiveness, and the payer could costlessly distinguish and pay the appropriate prices for each indication. Such differential pricing within product may become increasingly feasible as drugs become more ‘personalized’ on the basis of patient biomarkers, and data systems are improved to provide the necessary information at reasonable administrative cost.

Our approach to value-based pricing is similar to that proposed by Claxton et al. (2008), with important differences. First, because our approach is grounded in overall utility maximization, the payer’s ICER threshold

---

3The analysis also applies if universal coverage is implemented through regulated private insurance, assuming that competition in private insurance markets transmits consumer preferences, constrained or supplemented by subsidies and regulations to address altruistic concerns.
reflects consumers’ WTP for health gain, with the health care system funded accordingly. By contrast, Claxton et al. take the health budget as given, and the ICER threshold reflects the opportunity cost of current resource use. Second, our approach permits prices that transfer all surplus to manufacturers for the duration of the patent, to achieve optimal R&D incentives. (Optimality further assumes that patent terms are designed to achieve an optimal trade-off between current and future consumption and that product-specific R&D investments by government are negligible). Claxton et al. constrain manufacturers to a single price to retain some surplus for payers/consumers during the patent term. With heterogeneous consumers and possibly a non-optimal health budget, their approach could imply weaker incentives for innovation in pharmaceuticals compared with other industries. Third, they focus on pricing in a single country, whereas we address efficient pricing and utilization globally.

Our approach differs from the two part pricing approach put forward by Jena and Philipson (2008) (JP) and Lakdawalla and Sood (2009) (LS) in important respects. We propose that payers constrain prices indirectly, through an ICER threshold, whereas JP and LS assume that payers can observe and use the pre-insurance (counterfactual) consumer surplus for each drug to make an appropriate payment for R&D. We assume that eligibility for reimbursement/utilization is determined by the payer’s criteria, whereas JP/LS assume patients’ cost-sharing is set at marginal cost and patients select first-best utilization. Our approach reflects actual practice in most public and some private insurance systems, where payers define eligibility for costly technologies and only reimburse for patients with approved indications. Jena and Philipson (2013) show how payers setting ICER thresholds (affecting firms’ choice of profit maximizing prices and physicians/patients’ choice of utilization) can lead to inefficient utilization as prices vary above production cost. They recognize, however, that ICER threshold policies to achieve dynamic efficiency would need to be above production cost to encourage efficient R&D investment but do not say what those policies might be.

2.3. Practical issues in defining incremental cost-effectiveness ratio thresholds

2.3.1. Patient heterogeneity. If a payer applies a single threshold ICER across all individuals, this ICER may differ from the WTP for medical care of some individuals. In a world of perfect information, costless insurance and no altruistic concerns, this would violate Pareto efficiency. Under more realistic assumptions of asymmetric information, moral hazard, and equity/altruistic concerns, any insurance imposes significant restrictions on individual choice. Citizens choose to establish national health insurance and regulate/subsidize private insurance because the expected benefits, in financial protection and improved equity, exceed the utility cost of constrained choice. In such contexts, setting payer rules to reflect ex ante consumer preferences is likely to be superior to ignoring them. How best to elicit such preferences is an important issue for future research.

Moreover, there are ways of accommodating consumer heterogeneity, even in universal insurance systems with a single payer or competing insurers, to provide for both ‘voice’ and ‘exit’. In a pluralist system of competing insurers such as the USA, different health plans could choose different ICER thresholds, subject to minimum coverage requirements and subsidies as required by the 2010 Affordable Care Act, implying different levels of patient access and different drug prices. Individuals would choose a health plan that best reflects their preferences and WTP for health.

Private and public payers, in single payer or competing payer systems, could also vary ICERs by health condition to address social preferences across conditions, reflecting both altruistic and personal preferences. An illustrative example is the UK NICE ‘end of life’ ICER threshold (NICE, 2009), reflecting perceptions that society’s WTP for health increases with disease severity. Further, in any insurance plan, patients deemed ineligible for coverage of a given product could be permitted to pay out-of-pocket. Thus, the potential welfare losses associated with use of uniform ICERs across patients with heterogeneous preferences can be minimized.

---

See Danzon et al., (2012) for more detail. For private constraints on use in the USA, see Danzon and Taylor(2010).
3. VALUE-BASED DIFFERENTIAL PRICING: PRICE DIFFERENCES ACROSS COUNTRIES

This framework can be applied to determine optimal prices in each country with universal insurance coverage and optimal differentials across such countries. Specifically, if each public and private payer defines its ICER threshold unilaterally, based on its citizens’ WTP for health gain, manufacturers are permitted to price up to that threshold, and payers manage utilization to reimburse for cost-effective use at that price, then the resulting prices and utilization should be (approximately) consistent with second-best static and dynamic efficiency within each country and across countries.

3.1.1. Income effects. To consider how WTP—and therefore ICER thresholds and VBDP prices—might vary with income across countries, assume that all factors other than income and income-related preferences are invariant across countries and that each country makes its ICER choice $K$ unilaterally. Differentiating the RHS of Eq. (3) with respect to income yields

$$\frac{d}{dY} \left( \frac{v}{U_0} \right) = \frac{dK}{dY} = 1 - \frac{E_{U_Y}}{E_{U_Y}} \quad (4)$$

$E_{U_Y}$ is the elasticity of utility with respect to income ($\frac{dU/dY}{U}$) and is expected to be positive. $E_{U_Y}$ is the elasticity of marginal utility with respect to income ($\frac{dU/dY}{dU/dY}$), also known as relative risk-aversion. If $E_{U_Y} < 0$, individuals are risk-averse and the optimal ICER $K$ rises more than in proportion to income. If $E_{U_Y} = 0$, individuals are risk neutral and $K$ rises in proportion to income. Thus, if two countries differ in per capita income but are otherwise similar, our model suggests WTP for medical care and the resulting ICER thresholds and price levels will likely be higher in the higher income country, but the precise relationship to income cannot be predicted a priori.

3.1.2. Incentives to free ride. A country that accounts for a small share of global drug sales might recognize that setting its ICER threshold below its true WTP reduces the prices that it pays for drugs with at most modest effect on its access to existing drugs or on global incentives for R&D to develop new drugs. Such free riding incentives exist in any price regulation scheme and are not unique to VBDP. Free riding tends to undermine appropriate price differentials between countries and would likely lead to suboptimal R&D.

3.1.3. Relation between value-based differential prices and Ramsey pricing. Ramsey pricing has been proposed as an alternative framework for determining optimal pharmaceutical price differentials between countries (Danzon and Towse, 2003; Jack and Lanjouw, 2005). The question thus arises whether/how VBDP price differentials between countries differ from Ramsey optimal price (ROP) differentials?

Certain differences between ROP and VBDP are implied by their respective objective functions. Ramsey pricing is designed to determine welfare-maximizing price differentials across consumer groups, given an exogenous ‘joint’ investment and a normal return-on-investment (ROI) constraint. Formally, ROP prices are designed to minimize the single-period welfare loss from consumption below first-best levels, subject to the ROI constraint, but ignoring future utility, incentives for R&D and any consumer budget constraint. The resulting ROP prices vary across countries inversely with price elasticity of demand, assuming uniform marginal cost and one price per country. Absolute price levels are indeterminate without knowing cost and demand parameters and the ROI constraint.

By contrast, VBDP is designed to address both dynamic and static efficiency; it incorporates consumers’ lifetime utility and lifetime budget constraints and sets prices to transfer all surplus to the innovator firm, to incentivize future R&D. Thus, both VBDP and ROP are designed to achieve second-best static efficiency but subject to different constraints and with VBDP also designed to achieve dynamic efficiency. ROP incorporates the jointness of consumption, whereas VBDP as presented here does not. Only VBDP offers a practical approach to determine absolute prices, as well as relative prices.
To examine the implications of these two models for income-related price differentials across countries, note that ROP prices vary inversely with $E$, the uncompensated price elasticity of demand (absolute value). From the Slutsky equation (using absolute values), the uncompensated price elasticity for good $j$ ($E_j$) is equal to the compensated price elasticity $\epsilon_j$ plus the income elasticity of demand $\eta_j$ times the income share of $j$:

$$E_j = \epsilon_j + s_j \eta_j$$  \hspace{1cm} (5)

Plausible assumptions are that $\epsilon_j$ is invariant with income but $\eta_j$ increases with income. Differentiating (5) with respect to income yields

$$\frac{dE_j}{dY} = \frac{ds_j}{dY} \eta_j + \frac{d\eta_j}{dY} s_j$$  \hspace{1cm} (6)

The share $s$ is positive and, for a given price, decreases with income, ceteris paribus. The first term on the right is therefore negative. The second term is uncertain. The overall expression is likely to be negative, that is, price elasticity decreases with income, unless the last term is large and positive. Thus, under plausible assumptions, $E$ varies inversely with income, and therefore, ROPs vary directly with income across countries. Comparing Eqs. 4 and 6, the cross-national relationship between prices and income is not necessarily the same under ROP and VBDP. In the ROP formulation, income affects prices via the income elasticity of demand for health, whereas in the VBDP model, income affects prices via the elasticity of the marginal utility of non-medical consumption. Obviously, there is no unique a priori correct specification of utility functions and hence no determinate relationship between optimal drug prices and income across countries from these two models. However, both the ROP and VBDP approaches suggest that optimal prices will plausibly increase with income.

4. SELF-PAY MARKETS

Countries without universal insurance coverage lack its distorting effects on prices. Patients’ self-pay demand should therefore reflect their WTP for expected incremental QALY benefits, given good information on product quality and effectiveness. Because profit maximizing price discrimination leads to the same relative prices (inversely related to demand elasticities) as ROP prices, market incentives should lead unregulated price discriminating firms to set optimal price relativities across markets, provided they can segment markets between and within countries. Absolute price levels might be constrained to yield only a normal return by competitive entry to achieve a monopolistically competitive equilibrium (Danzon, 1997; Danzon and Towse, 2003). Whether actual cross-national price differentials in self-pay MLIC markets approximate ROP differentials cannot be determined because true demand elasticities and marginal costs are not observable. However, empirical evidence across a sample of MLICs shows that price elasticities with respect to average per capita income are close to zero and price competition is weak, for both on-patent and generic drugs (Danzon et al., 2011). This finding of higher prices, relative to per capita income, in lower income countries, seems inconsistent with optimal differentials under either ROP or VBDP prices.

Theory and empirical evidence suggest that two necessary conditions for second-best static and dynamic efficiency in self-pay markets are violated in many MLICs. First, highly skewed income distributions create incentives for single price originators to charge prices that are high, relative to average per capita income (Flynn et al., 2009), and price discrimination between income groups within countries is generally not feasible (Danzon et al., 2011). Second, quality of generic ‘copies’ is uncertain in many MLICs because such generics are not required to meet regulatory standards of bioequivalence to the originator. Uncertain product quality leads to competition focused on brand, rather than price. Thus, achieving prices closer to VBDP optimal prices in these self-pay markets requires regulatory requirements to assure product quality and purchasing mechanisms that facilitate differential pricing between market sectors based on income. Procurement mechanisms for HIV, TB, and malaria drugs provide an interesting prototype, at least for these drugs (Danzon et al., 2011).
5. CONCLUSIONS

Optimal pharmaceutical pricing is complicated by high R&D and patents and by extensive insurance in industrialized countries. We show that for countries with universal insurance, if each country/payer unilaterally and non-strategically sets an ICER threshold based on its citizens’ WTP for health gain and permits firms to price up to the ICER, while the payer assures reimbursement for all patients whose expected health gain meets the ICER effectiveness threshold, the resulting prices and utilization would be ‘valued-based’ and yield second-best static and dynamic efficiency within and across countries. In other words, the resulting price levels and differentials would be appropriate across countries to achieve second-best optimal global incentives for utilization and innovation. If data and information systems enable payers to implement price differentials between patient subgroups/indications to reflect their differences in incremental effectiveness, then VBDP would approximate first-best static and dynamic efficiency. Such value-based prices are likely to vary cross-nationally with per capita income and be broadly consistent with ROP differentials but the precise relationship of optimal prices to per capita income is indeterminate apriori.

Countries lacking comprehensive insurance avoid the distorting effect of insurance on manufacturer prices but also lack insurance’s role in financial protection and assuring medical care affordability, regardless of income. In such self-pay markets, unregulated price discrimination by firms could in theory lead to ROP differentials approximating second-best static and dynamic efficiency, provided that consumers are well-informed about product quality and firms can price discriminate within as well as between countries, to income disparities. Designing systems in MLICs to deliver consistent quality, facilitate price competition, and increase affordability while preserving some contribution to and incentives for R&D are important issues for future research.

6. FUNDING SOURCES

This research was supported by funding from Eli Lilly Inc.’s Project on Fair Prices for Pharmaceuticals. We would like to thank these sponsors for making the research possible. We are responsible for all conclusions and any errors.

7. ORIGINAL PUBLICATION

This is an original paper and is not being sent anywhere else. An earlier version has been published as NBER Working Paper No. 18593.

CONFLICT OF INTEREST

The authors have no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank Alan Garber and Michele Pistollato for comments on an earlier draft. We are responsible for all remaining errors.

REFERENCES


