

A BARGAINING APPROACH

## A theory on ICER pricing and optimal level of cost-effectiveness threshold

Mikel Berdud  
Jimena Ferraro  
Adrian Towse

[ohe.org](http://ohe.org)

CONSULTING | REPORT  
JULY 2020

**JULY 2020**

**A BARGAINING APPROACH**

# A theory on ICER pricing and optimal level of cost-effectiveness threshold

**Mikel Berdud**

Office of Health Economics, London, UK

**Jimena Ferraro**

University of Buenos Aires, Buenos Aires, Argentina

**Adrian Towse**

Office of Health Economics, London, UK

---

**Please cite this report as:**

Berdud M., Ferraro J., Towse A. 2020. A theory on ICER pricing and optimal level of cost-effectiveness threshold: a bargaining approach. OHE Consulting Report, London: Office of Health Economics. Available at: <https://www.ohe.org/publications/bargaining-approach-theory-icer-pricing-and-optimal-level-cost-effectiveness-threshold>

**Corresponding Author:**

Mikel Berdud  
[mberdud@ohe.org](mailto:mberdud@ohe.org)

For further information please contact:

**Professor Graham Cookson**

Chief Executive, OHE  
Honorary Visiting Professor in Economics at City, University of London

**Tel** +44 (0)207 747 1408

**Email** [gcookson@ohe.org](mailto:gcookson@ohe.org)

# About OHE Consulting Reports

Many of the studies OHE Consulting performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications. Publication is at the client's discretion.

Studies published by OHE as OHE Consulting Reports are subject to internal quality assurance and undergo external review, usually by a member of OHE's Editorial Panel. Any views expressed are those of the authors and do not necessarily reflect the views of OHE as an organisation.

## Funding and Acknowledgements

This consulting report was commissioned and funded by Roche Products Ltd.

# Table of Contents

---

<b>Executive Summary</b> .....	<b>iv</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>2 The model</b> .....	<b>5</b>
2.1 Baseline assumptions.....	5
2.2 The timing of the model.....	6
2.3 Micro-foundations of the demand, the supply, and the market outcome: the Nash Bargaining Solution .....	8
2.3.1 The developer.....	8
2.3.2 The payer.....	10
2.3.3 The bargaining problem between the developer and the payer .....	10
2.3.4 Delimiting the set of possible bargaining outcomes and agreements.....	11
2.4 Macro level implications of the ICER pricing.....	14
2.4.1 The aggregate demand of new health technologies: the payer perspective.....	15
2.4.2 The aggregate supply of innovative medicines: the industry perspective .....	20
2.4.3 The social optimum: maximising the social welfare.....	23

---

<b>3 The functioning of the market for new medicines</b> .....	<b>26</b>
3.1 The baseline model.....	26
3.2 R&D cost of innovation, reserve ICERs and dynamic efficiency .....	29
3.3 Restrictions based on non-value-based criteria .....	32
3.4 Competition in pharmaceutical markets for patented products.....	35
3.5 Distribution of reserve ICERs and the value of health displaced .....	38
3.6 Non-binding flexible health budgets in the long run.....	40

---

<b>4 Concluding remarks</b> .....	<b>44</b>
<b>References</b> .....	<b>47</b>
<b>Appendix 1</b> .....	<b>49</b>
<b>Appendix 2</b> .....	<b>52</b>
<b>Appendix 3</b> .....	<b>54</b>

# Executive Summary

With the growing pressure on governments over rising health expenditures, health budgets and pharmaceutical pricing have increasingly come under scrutiny. Governments and payers seek to inform their decisions using value-for-money or value-based criteria, in order to make efficient use of limited resources. Thus, in many health systems around the world, decisions about the reimbursement of – and patient access to – medicines are based on health technology assessments (HTA). These assessments typically include a measure of cost effectiveness based on the concept of the Incremental Cost-Effectiveness Ratio (ICER), which calculates incremental costs per unit of incremental health gain. In some cases, decision makers will compare the ICER of a new medicine against a pre-specified value and decide in favour or against the reimbursement of medicines with ICERs respectively below or above this value. This pre-specified value is known as the Cost-Effectiveness Threshold (CET).

Increasingly, decision makers are using the CET as a tool for price regulation – that is, they will only reimburse a new medicine at a price (or treatment cost) where the ICER is equal to or below the CET. This approach – commonly known as value-based pricing (VBP) (Danzon, et al., 2015; Danzon et al., 2011) – is referred to as *ICER pricing* in the paper<sup>1</sup>. Thus, the value of the CET has important implications not only for drug pricing and drug spending, but also for patients' access to novel medicines and their health outcomes, and developers' R&D investment decisions on new medical interventions including medicines.

Despite its growing importance, there is no agreement on how the value of the CET should be set, and how it should be used by decision makers or government. Notably, most CETs that have been suggested – as well as those currently in use – are determined using different approaches and limited data (Cubi-Molla et al. 2019). In addition, there is a lack of theoretical economic models exploring the effect of CET application on the allocation of value or benefits between consumers and developers of new medicines. Whether these allocations are efficient in the short and/or long-run (in terms of patient access to novel medicines and R&D investment in medical innovation) is an outstanding policy question.

This paper proposes a supply and demand model of pharmaceutical markets to analyse the relationship between the value of the CET and the distribution of the health benefits and economic value of new medicines between consumers (payers) and developers (life science industry). This model is novel in that it incorporates a bargaining process to analyse the impact of different degrees of consumers' and developers' bargaining power on the distribution of the health and economic value of new medicines between the two parties.

The model proposed in the paper builds on the approach of Pandey et al. (2018), with a number of adjustments incorporated to address the following limitations:

- the bargaining power of consumer and the bargaining process for pricing negotiations is not considered;

---

<sup>1</sup> We use *ICER-pricing* instead of *VBP* because while *ICER-pricing* is a form of *VBP*, *VBP* does not imply *ICER-pricing* as other forms of measuring value are possible.

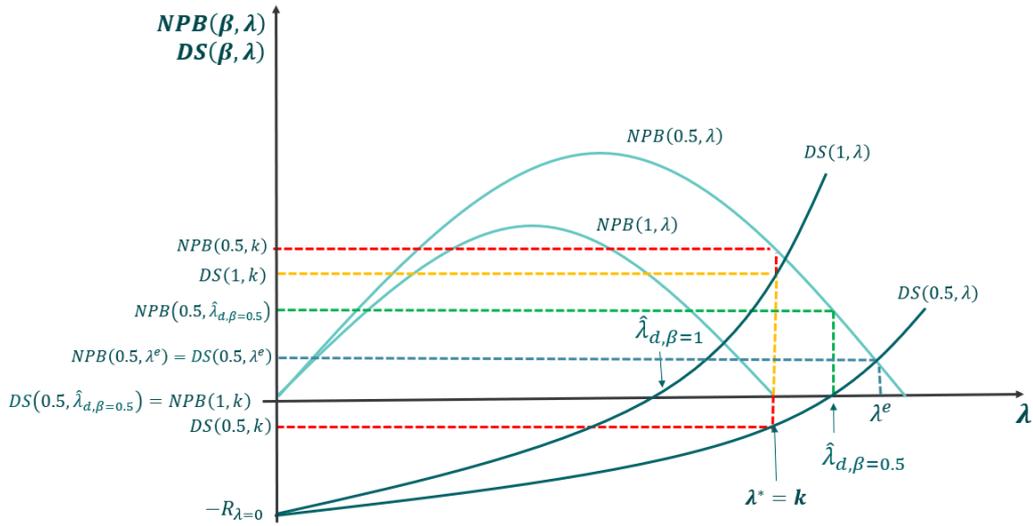
- R&D costs, an essential component for the sustained development of new medicines, are not fully factored;
- the shape of the supply curve for new medicines as a function of the CET is assumed to be linear, hence constant increases in the CET value have a constant positive effect on the resources invested in R&D and the number of new medicines supplied in the market, whatever the level of the CET is;
- health budgets are assumed to be fixed, hence budget changes in the mid-/long-term are not considered.

Following these assumptions, an efficient CET for the payer is lower than the CET based on the opportunity cost in the health system (the so-called supply-side CET). This is the maximum amount payable per unit of health gain so that the health added to the system by the new drug is equal to the health foregone, as resources used in current health technologies have to be reallocated to pay for the new drug.

The general framework developed in this paper aims to model the economic behaviour of both the payer and the developers. We incorporate a bargaining process where the bargaining power is shared between the payer and developers. We also incorporate sunk costs of R&D (i.e. costs that have already been incurred before the medicine is approved for use) into the developers' decision-making process. We consider non-linear supply curves, which imply that the number of new medicines launched in the market in response to changes of the CET varies depending on the level of the CET, not just the size of the change. We also incorporate the possibility of a flexible budget in the medium-term. We explore the implications of these factors on the level of the CET that can be considered optimum from a society perspective – that is, the CET for which the total health and economic value obtained by the payer/consumers and developers is maximised with none of them making losses or alternatively both of them obtaining a share of such value.

The key result is that, if certain conditions hold, the optimal CET value (i.e. the one that reconciles incentives for developers to invest in future medical innovations with the maximum possible patient access to currently available innovations) can be higher than the supply-side CET. Implementing allocations that distribute the health and economic value more equally can be an efficient solution under these circumstances as shown in Figure S1.

In Figure S1, the vertical axis measures the payer's benefit  $NPB(\beta, \lambda)$  and developers' benefit  $DS(\beta, \lambda)$ , both as a function of the bargaining power distribution – developers' bargaining power is  $\beta$  and payer bargaining power is  $(1 - \beta)$  – and the level of the CET  $\lambda$ . The horizontal axis measures the level of the CET. The  $NPB(\beta, \lambda)$  curve measures, for a given distribution of the bargaining power ( $\beta$ ), how much monetised net health benefit the payer obtains for each level of the CET ( $\lambda$ ). The  $NPB(\beta, \lambda)$  curve first increases on the CET because the monetary value of the health incorporated by all the additionally accepted medicines at the increased CET (the access effect) is higher than the monetary value of the health foregone by the highest price that must be paid for all medicines under the higher CET (the price effect). It is decreasing when the price effect exceeds the access effect. The  $DS(\beta, \lambda)$  curve measures, for a given distribution of the bargaining power, the industry's profits for each level of the CET. The  $DS(\beta, \lambda)$  is increasing on the CET because the higher the CET, the higher the prices and the returns on R&D investment, which increases profits as well as incentives for the production and supply of more new medicines.



**Figure S1: OPTIMAL CET WITH PAYER BARGAINING POWER AND DEVELOPERS' R&D COST**

Note: We use  $\lambda^* = k$  to denote the CET level that maximises the total benefit and  $k$  is the supply-side CET. The minimum acceptable CETs for developers are represented by  $\hat{\lambda}_{d,\beta=1}$  and  $\hat{\lambda}_{d,\beta=0.5}$  on the horizontal axis, which are the values that makes developers' profit equals to zero when they have all the bargaining power ( $\beta = 1$ ), or when the bargaining power is equally distributed between them and the payer ( $\beta = 0.5$ ). The level of the CET  $\lambda^e$  corresponds to an allocation that split the total benefit equally between the payer and the developers. The size of the sunk cost of R&D for a CET level  $\lambda = 0$  is represented by  $-R_{\lambda=0}$ .

The dashed yellow lines in Figure S1 represent the case where all bargaining power is held by the developer ( $\beta = 1$ ). The benefit functions for the payer and the developers are  $NPB(1, \lambda)$  and  $DS(1, \lambda)$ , respectively. In this case, the CET that maximises total benefit of the payer and the developers is  $\lambda^* = k$  (the supply-side CET), where all benefit is accrued by the developers  $DS(1, k)$  while the payer's benefit  $NPB(1, k)$  is equal to zero.

The dashed red lines in Figure S1 represent the case where bargaining power is distributed equally between the payer and the developers ( $\beta = 0.5$ ). The benefit functions for the payer and the developers are  $NPB(0.5, \lambda)$  and  $DS(0.5, \lambda)$ , respectively. Again, the CET that maximises the total benefit of the payer and the developers is  $\lambda^* = k$ , where the payer obtains positive benefits  $NPB(0.5, k)$ , and the developers make losses  $DS(0.5, k)$ . As a result, long-run investment in R&D is disincentivised.

Therefore, under the scenario of shared bargaining power ( $\beta = 0.5$ ), the CET must be set equal to (or higher than)  $\hat{\lambda}_{d,\beta=0.5}$  to provide sufficient incentives for R&D investment and at the same time facilitate maximum possible patient access to medicines. This is represented by the dashed green line in Figure S1. Another possible allocation that shares the benefit equally between the payer and the developers, also with a bargaining power distribution  $\beta = 0.5$ , requires a CET equal to  $\lambda^e$ , represented by the dashed blue line in Figure S1.

Our additions to the theoretical demand and supply model leads to four main implications:

- i. First, **the incorporation of developers' sunk cost of R&D** (section 3.2 of the paper) shows which values of the CET should be considered to ensure sufficient returns in the long-term for dynamic efficiency, i.e. sustain an optimal level of long-term

investment in pharmaceutical innovation. It also shows which CET values will be efficient in the short run to maximise access but will force industry to supply at cost to cover, in part, the R&D investment. In such cases, the optimal CET may be higher than the one that maximises the short-term benefit of the payer. To restore long-term efficiency, some **short-term population benefit needs to be traded off via a higher CET to incentivise developers to keep investing in pharmaceutical R&D in the long-term, and thereby obtain a higher level of long-term population benefit.**

- ii. Second, as single buyers in the market, **payers have a significant degree of bargaining power** (sections 3.3 and 3.4). This is reinforced by the fact that they are also empowered to design market policy and regulation to meet their objectives. We describe a staged bargaining process and apply the Nash Bargaining Solution (Grennan, 2013; Jelovac, 2015) to explain market prices resulting from negotiations at the procurement stage. Results show that **an efficient CET value, which distributes health and economic value evenly between the payer and developers, could be higher than the supply-side CET with sufficiently large payer bargaining power.**
- iii. Third, **how potential market entrants respond to different levels of the CET** (section 3.5) has an impact on the shape of the developers' supply and the payer's demand of new medicines. Therefore, different distribution functions of the developers' reserve ICERs (the minimum ICER at which each developer is willing to accept a price to sell the medicine) will have an impact on the optimal CET level. **If the reserve ICERs of the developers concentrate around a certain value of the CET, the optimal value of the CET will be close to that value.** Lower CET values would be inefficient as many cost-effective medicines will be denied access. Higher CET values would also be inefficient because too much will be paid for new medicines.
- iv. Fourth, our model explores **the impact of budget increases in the medium-term** (section 3.6). Although the assumption of fixed budgets is plausible in the short-term, we consider that a significant degree of flexibility is possible in the medium-term. Health budget growth in the medium-term might also result in increased funding for medicines assessed in the present, as a share of their cost will be borne by the system in the future. Along with previously discussed changes, **the incorporation of flexible budgets increases the CET level that splits the surplus equally between both players in the market.**

In conclusion, under the four implications discussed above, applying a supply-side CET to determine the maximum price for medicines reimbursement may result in inefficient allocations of health resources, where new and cost-effective technologies could be denied access in the long term. This could subsequently reduce overall benefit to society in both the short run – patients not having access to valuable interventions, and developers making a loss – and the long run – reduced investment in R&D. Regulation of pharmaceutical prices should be informed by a clear understanding of both the market structure, and the procurement and contracting environment.

# 1 Introduction

Health systems commonly use health technology assessments (HTA) to inform new medicines reimbursement decisions. Many HTA bodies around the world – such as the National Institute of Health and Care Excellence (NICE) in England – apply a form of HTA based on Economic Evaluations (EE) and cost-effectiveness criteria. Other HTA bodies use a Therapeutic Value-Added approach, where the emphasis is on assessing benefit, with pricing and value-for-money issues addressed outside of the HTA process.

In the case of HTAs using cost-effectiveness analysis, new health technologies are appraised to determine whether or not they are considered cost-effective. Decision makers – typically the government or another payer, but may sometimes be the HTA body when its decision is binding – compare the incremental cost-effectiveness ratio (ICER), as estimated by the HTA, for the treatment submitted by the new medicine developer, with the health system’s cost-effectiveness threshold (CET). This threshold represents the cut-off point below which a new technology would be considered cost-effective. The decision to fund a new technology is contingent on the new medicine having an ICER that is equal or below the CET. The CET may be implicit or explicit, and is often adjusted to take into account factors relating to the target population or characteristics of the disease under consideration that are relevant to the decision. Thus, the role of the CET is not only to serve as a resource allocation guide, but also to regulate price<sup>2</sup>.

Informing reimbursement decisions by means of HTAs via ICERs submitted by drug developers with the accepted CET, has several economic implications. Typically, effectiveness is measured in Quality-Adjusted Life Years (QALY), and costs are those incurred by the health system. Although this is a well-established standard procedure, there is debate as to whether all relevant dimensions of cost-effectiveness are captured by this framework. For instance, many HTA bodies (and countries) consider costs beyond those incurred by the health system, including those for family caregivers, or indirect costs, such as absences from work (Cubi-Molla et al. 2019). Other elements of value that are beyond pure clinical effectiveness – such as unmet need, insurance value, value of hope, or rarity of the disease – could in theory be considered in addition to health gains. This approach will provide a wider perspective than the exclusive use of the QALY (Garrison et al., 2019; Lakdawalla et al., 2018; Sculpher, Claxton and Pearson, 2017).

Even when health economists and decision makers agree on the method of measuring the cost-effectiveness of a treatment there exists a follow-up question: how should the CET be set? The CET can be interpreted as the maximum monetary value that the country or the system is willing to pay for each incremental health gain (i.e., QALY), and it needs to be determined to be implicitly or explicitly binding, or informative for health technology funding decisions. There is on-going controversy in the scientific community regarding the optimal level of this threshold for health systems, patients, or society in theory or in practice in any particular health system. A key point of debate is whether CETs should be determined using supply- or demand-side approaches, with advocates on both sides

---

<sup>2</sup>In Canada, for example, the Patented Medicines Prices Review Board (PMPRB) is establishing nationwide guidelines, whereby the maximum (rebated) price of patented medicines, with sales or treatment costs above a certain amount, will be set using factors including incremental QALYs, valued using a CET of CA\$60,000 per QALY gained.

(Vallejo-Torres et al., 2016; Thokala et al., 2018). However, there is a gap in the literature on integrating both approaches (demand- and supply-side), to determine the CET. Attempts to reconcile the two perspectives are rare.

Health systems using *supply-side* approaches may seek to determine the level of the threshold as the health system opportunity cost. In theory, this reflects the amount of money the health system needs to spend at the margin to produce one unit of health gain – a proxy for a measure of the quantity of health displaced when new technologies are funded. Under this approach, if the objective of the system is to maximise population health, society would never be willing to pay an amount higher than the CET as the health foregone exceeds health gained (Brouwer et al., 2019; Lomas, Martin and Claxton, 2019; Claxton et al., 2015). However, the logic of this approach requires a set of assumptions to hold, these include: a fixed health budget, an objective measure of the health gain (the QALY), a judgement to treat all units of health as being of equal weight ('a QALY is a QALY'), and prices of medicines included in the calculation of the ICER to remain unchanged throughout their life-cycles<sup>3</sup>. Under *demand-side* approaches, the level of the CET reflects society's monetary valuation of incremental health gain. Again, this may vary by social context. Its advocates argue that this approach is in-line with those taken in other sectors (Baker et al., 2011; Smith and Richardson, 2005). In the UK for example, the Department of Transport, the Rail Industry, the Department for the Environment and other government agencies use willingness-to-pay (WTP) based value of safety in their cost-benefit analyses (Mason et al., 2009)

For HTAs using either a supply- or a demand-side CET, the pricing of new medicines is determined according to the way by which the ICER relates to the system's CET. This is what has come to be widely known as *value-based pricing* (Danzon, Towse and Mestre-Ferrandiz, 2015; Danzon, Towse and Mulcahy, 2011). In this paper, we refer to this as *ICER pricing*. Although the mechanism is well defined, there is a lack of theoretical economic models exploring the allocation of consumer and producer surplus, as well as social welfare generation, under ICER pricing<sup>4</sup>.

There is a recent attempt by Pandey and colleagues (Pandey et al., 2018) to model the demand and supply of new medicines under ICER pricing. The main assumptions in this novel approach are: (i) the supply-side CET is the payer's maximum WTP, which includes all implicit assumptions previously discussed; (ii) the developers have all the bargaining power, and will hence price the new technology where the value of the company submitted ICER equals the CET; and (iii) the reserve ICERs of developers – the minimum ICER at which they are willing to sell the new technology – are uniformly distributed along the different values of the CET, ranging from zero to a maximum value above the supply-side CET<sup>5</sup>.

---

<sup>3</sup> We note that some advocates of a supply-side threshold – based on an estimate of the incremental value of health gain – separate this from a 'decision threshold', which would take into account this number but also other social welfare judgements.

<sup>4</sup> As we noted earlier, the assumption is that the prices of drugs (and comparators) at the time of review are used (typically at the launch of the product.) The fact that prices change over the life cycle is often not considered. There is an implicit or explicit assumption that gains from lower prices at later points in the life cycle should accrue to the health system. This is now being more explicitly considered.

<sup>5</sup> The work by Pandey and colleagues (Pandey et al., 2018) includes an appendix where the role of different distribution and density functions of the developers' reserve ICERs are discussed as options for further research (see: APPENDIX A: Strategic behaviour and the cost-effectiveness threshold – a new conceptual model).

This paper generalises the Pandey et al. (2018) model by incorporating additional structure based on economic theory, as well as observational evidence from pharmaceutical market functioning, medicines procurement, and pricing. Market competition – or the use of procurement and contracting mechanisms that encourage competition – when other similarly effective treatments become available will reduce medicine prices. Such mechanisms include tenders, price-volume agreements and/or the pairwise negotiations between the budget holder (e.g., regional health system, hospital, retail chain) and the innovator. Other regulatory tools and policy levers for pharmaceutical expenditure control – such as budget limits – have similar effects on prices. All these features increase the bargaining power of single-buyer payers. We relax some key assumptions based on these features and explore the implications of different assumptions on the optimal CET and the distribution of consumer and producer surplus.

We introduce research and development (R&D) costs as sunk costs for developers. We identify the implications of incorporating R&D costs in the reserve ICERs of the developers<sup>6</sup>. Our model shows that, for some positive values of the CET, developers would make a loss because revenues would be lower than the combined costs of R&D, manufacturing, and distribution. The effective reserve ICER for developers is the one that ensures R&D investment is recouped. Therefore, we assume in our model that all costs of R&D are recovered in the long run<sup>7</sup>.

We also assume that both consumers (payers) and producers (developers) have some bargaining power. We follow the Nash Bargaining Solution (NBS) to incorporate the effect of bargaining power on the final price and the effective ICER (which corresponds to the final agreed price). Given some of the characteristics of developers and payers, there may be situations in which the payer can use their bargaining power to extract some of the developer surplus. We note that there is likely to be a difference between short-run and long-run optimal CETs for payers<sup>8</sup>. We modelled two situations where the bargaining power of the payer increases: (i) when policy and regulation is used (e.g., budget cap, profit control, discounts, price volume) and (ii) when there is in-class market competition.

Another consideration in our model is the possibility that developers' reserve ICERs concentrate around certain values of the CET, which means that the distribution of reserve ICERs is not uniform in the value of the CET – that is, the supply curve of new medicines as a function of the CET is not linear<sup>9</sup>.

---

<sup>6</sup> In the APPENDIX A of Pandey et al. (2018), Dr Mike Paulden, PhD discusses the possibility of some developers not supplying at a loss. This specification partially addresses the issue of R&D cost, but it does not recognise the sunk nature of R&D expenditure. The main implication of this sunk cost is that developers will sell the medicine even at a loss in the short term at the expense of a dynamic inefficiency (suboptimal investment in R&D for future innovation) in the long term.

<sup>7</sup> Note that R&D is a global sunk cost. How much R&D is relevant for a market reflects the WTP for health gain in a market. If we have a model where the CET is set depending on that allocation of R&D in the form of a reserve price, then there is a potential circularity in the Pandey et al. approach as well as in ours. We can assume for now that we have only one market so there is no global R&D attribution issue.

<sup>8</sup> Higher short run profits from innovation means more investment in innovation over time. We will seek to illustrate not just the effects on the distribution of social surplus as between consumer and producer of different relative bargaining powers, but also the implications for long run societal optimums and the long run maximisation of consumer surplus of different bargaining outcomes.

<sup>9</sup> It is likely that R&D effort is endogenous to the prices and revenues that can be achieved. Thus, wherever the CET is set, reserve prices will move towards that level over time as companies compete in R&D effort to realise supra-normal profits. Over time in any therapeutic area we might expect competitive entry to reduce

Finally, we explore the short-term implications of having flexible budgets in the medium-term. By allowing a rate of growth in the budget over time, the current opportunity cost of the health system can be altered. This means that the health gains that new medicines deliver in the future should be considered at a different CET.

Our general theoretical model of cost-effectiveness of ICER pricing shows that, when the payer has effective bargaining power, the CET can – in some circumstances – be set at a level above the supply-side threshold without involving a net health loss for the system. The same implication arises where health budgets are flexible and allowed to increase. The difference between these two scenarios is that in the latter, there is an additional surplus obtained by the two players due to the extra funding, while in the former, the benefit comes from transferring some of the surplus of the developer to the payer via a price effect. The incorporation of R&D cost leads to an increase in the CET to incentivise developers to invest in future pharmaceutical innovation – i.e., to ensure dynamic efficiency. Finally, if reserve ICERs of the developers concentrate around a narrow range of the potential CET values, then the optimal threshold - or alternatively, the threshold that equalises payer developer surplus – will also be within that range. It will depend on the skewness and the kurtosis of the associated density function whether the optimal CET is higher or lower than with a uniform distribution.

The paper is organized as follows: section 2 will provide a description of the model. Section 3 will explain the functioning of the market under our model based on the different cases presented above. Section 4 will present a discussion of the results and provide concluding remarks.

---

*profits to normal levels, providing the science and regulatory environment permits competitive development and providing payers are able to increase their bargaining power over price when there are competing suppliers of on-patent products in that therapeutic area.*

## 2 The model

We analyse the implications of the ICER pricing in markets characterised by single third-party payers. Whilst we recognise that private health and out-of-pocket markets can make up a significant share of demand in some countries, we have excluded them from our analysis as they are rarely driven by ICER pricing. We begin with the simplest specification of the model which will be used as the baseline for the analysis, then develop the model further, with different sets of assumptions.

### 2.1 Baseline assumptions

The baseline assumptions upon which we build our model are as follows:

1. There is an efficient, publicly funded health system, focused exclusively on delivering benefits for patients, within a constrained fixed budget. Adopting new technologies will displace existing health care services, resulting in forgone benefits for other patients.
2. There is an accepted measure of the benefit patients derive from health care. This is exclusively a health benefit. We have designated  $h_i$  as the health gains per patient being treated with manufacturer  $i$ 's medicine. This is the benefit of use of the medicine.
3. New technologies are costly to produce, and manufacturers will not supply at a loss<sup>10</sup>.
4. A single threshold  $\lambda$  is publicly specified by a health system decision maker. New technologies are adopted if and only if the Incremental Cost-effectiveness Ratio (ICER) is less than  $\lambda$ .
5. Manufacturers of new technologies are protected from price competition through the patent system and an assumed lack of competing products with similar health effects, thus allowing for super-normal profits.
6. Each manufacturer has a minimum 'reserve price', or 'reserve ICER', that must be met before supplying a new technology. This price must be sufficient to cover the variable costs of production and distribution in the short run, and the fixed costs of production and – most importantly – the manufacturer's R&D investment (a sunk cost) – in the long run. There is a broad, continuous, linear distribution of 'reserve ICERs'<sup>11</sup> between

---

<sup>10</sup> This assumption comes from the work of Pandey et al (2018). They do not specify if they have one market or a global market. In the real-world companies may make losses in certain markets while global profits are positive. The model only analyses a single market. We can therefore think of it as either some sort of aggregated global market represented by one ICER-CET regulated market with a single third-party payer. When we include R&D costs we allow for certain levels of supplier losses relative to total cost when the size of such a loss is smaller than the R&D investment, thus continued operation is at a surplus over short run costs.

<sup>11</sup> It is the minimum level of ICER at which it is profitable to the developer to provide the technology. The assumption that these reserve ICERs are broadly distributed is an important one and we explore alternative possibilities later in the paper.

zero and  $\lambda^M$ , where  $\lambda^M$  represents the upper-bound of the distribution, or the maximum value of a reserve ICER.

7. We assume that  $\lambda^M > k$ , where  $k$  is the health system's opportunity cost, (estimated by the conventional supply-side model) which represents the relationship between marginal reductions in expenditure on existing health care services and forgone benefits for other patients.
8. The developers (life-sciences industry) and the payer (consumers) have a distribution of relative bargaining power when setting the price. We define this as follows:  $\beta$  is the manufacturer's bargaining power and  $1 - \beta$  is the payer's bargaining power. The result of bargaining is that the payer may end up paying less than the threshold,  $\lambda$ , for some or all medicines.
9. Each new technology is independent and developed by a different manufacturer.

These baseline assumptions reflect those of the Pandey et al. (2018). model, with two exceptions:

- In assumption 6, we have sought to distinguish between short-run and long-run reserve ICERs. The short-run developer profit is negative and equal in size to the R&D cost when  $\lambda = 0$ . The long-run reserve ICER is fixed at a level where the R&D sunk cost is recovered, and developer profits are equal to zero.
- Assumption 8 is unique to our model. Pandey et al. (2018) assumes that the developer holds all the bargaining power, hence the final price of the new treatment is set where the ICER is equal to the CET. We include a component which allows bargaining power to be shared between the payer and the developer.

## 2.2 The timing of the model

The timing of players' (i.e. the payer and the developers) decisions and outcomes is separated in five different stages:

**Stage 1:** nature fixes the level of the system's maximum WTP i.e. the value of  $k$ , the system's opportunity cost of the marginal health gain.

**Stage 2:** the developer, with knowledge of the system's WTP level and its own reserve ICER, decides whether or not to invest in R&D to develop a new medicine<sup>12</sup>

**Stage 3:** the payer (health care decision maker) commits to a CET level  $\bar{\lambda}$  to inform reimbursement recommendations and pre-bargaining decision making. The value of  $\bar{\lambda}$  will be equal or lower than the system's maximum WTP as set by nature in stage 1.

**Stage 4:** the developer sets the list price and the corresponding ICER using both clinical evidence and health costs to determine reimbursement status.

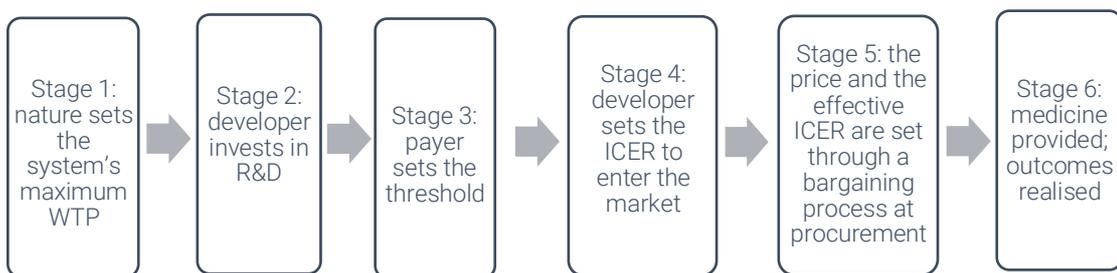
---

<sup>12</sup> We are aware that this is a repeat game and developers will decide which investments to undertake given their expectations about the likely CET that will be applied by the payer. Although we are assuming R&D effort is exogenous, we are aware that in practice it is endogenous. In other words, the payer setting a maximum WTP via a CET will determine R&D effort such that the marginal project will have an expected return for the developer equal to a normal rate of profit.

**Stage 5:** the net final price of the new technology (and therefore the agreed effective ICER) is set through a bargaining process at the procurement stage, taking into account the bargaining power of both players. This price may therefore be below that implied by the threshold ( $\bar{\lambda}$ ) committed to by the payer at stage 3.

**Stage 6:** medicine is sold by the developer to the payer and provided to patients. Outcomes are realised.

Figure 1 shows the sequence of stages of the bargaining process in the market for new medicines.



**Figure 1: TIMING OF THE BARGAINING PROCESS**

In practice, behaviour at each stage will anticipate behaviour in the next stage. For example, R&D investment will depend on expectations about the threshold, which in turn will be set by taking into account expectations of how developers will invest in R&D and price new medicines accordingly. With the proposed timing, companies' R&D investment behaviour will depend on the system's WTP (fixed by nature at stage 1). This is assumed to be  $k$ , the supply-side defined CET<sup>13</sup>. Therefore, the R&D investment is a function of an exogenously determined CET by the system's health production function.

Switching stages 2 and 3, the developer's decision on whether or not to invest in R&D becomes a function of the payer's decision on the level of the CET. At the same time, if the payer is aware that the amount of R&D investment is a response to its decision about the CET, the point at which to fix the CET becomes a crucial policy question if the objective is to maximise innovation. Therefore, the reserve ICERs and the CET become endogenous to each other. The developers will use backward induction of payer behaviour to decide their level of R&D investment. Likewise, the payer will use forward induction of developer behaviour to set the level of the CET.

Such endogeneity is not the case under the timing of Figure 1, where developers take their decisions based on nature's settings, and the payer does not consider the R&D investment already committed. When payers decide the value of CET to inform reimbursement decisions, developers have already invested in R&D. The payer could potentially get the most of existing innovation by simply reducing prices via lower CETs, up to the point where the consumer surplus is maximised. We have calculated both, the social maximising threshold and the dynamic reserve ICER under the alternative timing generated by switching stages 2 and 3. Calculations are presented in the following section, and results compared and discussed in Appendix 3.

<sup>13</sup> We follow the Pandey et al., (2018) model where the key assumption is that the system maximum WTP per unit of health gain is determined by the supply-side threshold.

## 2.3 Micro-foundations of the demand, the supply, and the market outcome: the Nash Bargaining Solution

In this section, we study the economic behaviour of both players: the developer and the payer. We characterise the reserve ICER for the developer and the reserve ICER (i.e., the CET)<sup>14</sup> for the payer.<sup>15</sup>

We depart from stage 2 where the payer commits to  $\bar{\lambda}$ , (i.e. the value of the threshold that is used to inform funding and reimbursement decisions). The payer and the manufacturer then bargain over the difference between the price of the new medicine at the reserve ICER of the developer (the minimum willingness-to-accept price per QALY gained), and the price of the new medicine at the CET of the payer (the amount the payer has committed to pay per QALY gained). Where the reserve ICER of the developer is lower than the reserve ICER of the payer, the difference may be interpreted as the set of all possible solutions to the bargaining problem.

### 2.3.1 The developer

Let the profit function of the developer of medicine  $i$  be:

$$\Pi(p_i) = p_i q_i - C_i(q_i) \quad [1]$$

Where  $q_i$  is the number of patients treated (the intention-to-treat population), which is taken as given;  $p_i$  is the final price per unit of medicine, and  $C_i(q_i)$  is the production cost function of the developer which we define as:

$$C_i(q_i) = R_i + c q_i \quad [2]$$

Where  $R_i$  represents the cost of R&D incurred by the developer to develop medicine  $i$ <sup>16</sup>, and  $c q_i$  is the variable cost the developer needs to incur to sell  $q_i$  units of medicine  $i$ , with  $c$  being the marginal cost of production. The cost of R&D is a sunk cost; hence, the developer has incentives to sell the medicine in the short run starting at a price that provides a benefit higher than  $-R_i$ . This level of price is determined by solving equation [3]:

$$-R_i \leq p_i q_i - (R_i + c q_i)$$

Rearranging, we have:

$$\underline{p} \geq c \quad [3]$$

In the long run, however, the developer does not have incentives to develop the medicine if the final price does not (at a minimum) ensure the cost of R&D – in addition to the cost of

<sup>14</sup> Other value-based and non-value-based approaches (i.e. budget-impact thresholds, clinical effectiveness, unmet need, rarity) are often used in value frameworks or HTA and they could also be considered.

<sup>15</sup> We assume that we have only one market with one payer. Revenues from this payer are the only way to recover global R&D costs. In practice the R&D costs that “need” to be recovered from any particular market ex ante to provide a particular return may differ from those recovered ex post, even if total revenues and margins from all markets are in line with ex ante expectations.

<sup>16</sup> For the sake of simplicity, we do not consider any fixed cost in addition to the R&D investment. This assumption does not change results but eases the definition of the role played by the R&D cost in the result.

manufacturing<sup>17</sup> – will be recovered. This long run price is calculated by solving equation [4]:

$$0 \leq p_i q_i - (R_i + c q_i) \quad [4]$$

Rearranging, we have:

$$\hat{p} \geq c + \frac{R_i}{q_i}$$

We now define the short-run and long-run reserve ICERs of the developer. First, let the developer's ICER of a new medicine be defined as:

$$ICER_d = \frac{p_i - p_j}{h_i - h_j} \quad [5]$$

Where  $p_j$  is the price of the comparator (usually the standard care at the time of the adoption of the new medicine  $i$ ;  $h_i$  and  $h_j$  are the health gains per patient of the new medicine and the comparator respectively<sup>18</sup>. As the ICER is defined as a function of the price of the medicine, we can define the short- and long-run reserve ICERs. Let  $\underline{\lambda}$  be the short-run reserve ICER of the developer which corresponds to a price level  $p_i = \underline{p}$ , and let  $\hat{\lambda}$  be the long-run reserve ICER of the developer which corresponds to a price level  $p_i = \hat{p}$ . In addition, assuming that the patient population is fixed, the profit of the developer can also be expressed as a function of the ICER, where acceptable profits are only achieved with ICERs equal to or above the long-run reserve ICER, that is,  $\lambda_i \geq \hat{\lambda}$ , where  $\lambda_i$  is the ICER corresponding to the final price  $p_i$ .

Finally, we define the outside option of the developer – denoted as  $\Theta^d$  – as the payoff to the developer where an agreement on a price is not successfully reached. We assume that there are no private health markets and no out-of-pocket demand for the medicine. Thus, the outside option of the developer, conditioned on having developed the product, is a negative profit equal to the magnitude of the sunk R&D cost,

$$\Theta^d = -R_i \quad [6]$$

---

<sup>17</sup> We assume that in the long run a developer that does not expect enough return on investment to at least recover the R&D investment would not start the project of a new medicine. These disinvestment decisions are hard to take when projects are ongoing and about to complete, but they are taken for planning the optimal investment level in the future.

<sup>18</sup> For the sake of simplicity, we assume that for the health system all differences in the cost of using medicine  $i$  or  $j$  are due to the price difference, if any, between them. In a more general specification, the ICER should also account for the differences in direct health costs, such as costs of treatment administration, inpatient costs, acute and emergency costs, primary care costs, personnel costs, etc. Formally that would be represented by  $ICER = \frac{HC_i - HC_j}{h_i - h_j}$  where  $HC_i$  is the function representing total health care cost of using the medicine  $i$  which we assume it is additively separable as follows  $HC_i = p_i q_i + T_i(q_i)$  and  $T(q_i)$  represents the total direct health care cost of providing medicine  $i$  except the cost of the medicine. Assuming  $T_i(q_i) = T_j(q_i)$  implies that all differences in the cost of using treatment  $i$  instead of  $j$  are due to the difference in price.

### 2.3.2 The payer

In stage 1, nature sets the level of the threshold  $\lambda$  and determines the maximum WTP of the payer  $\lambda = \{k, v\}$ <sup>19</sup>. It determines the maximum amount payable per unit of health gain. The (gross) payoff function of the payer is then:

$$B_i(p_i) = \lambda h_i q_i - (p_i q_i + T_i(q_i)) \quad [7]$$

Where the first addend on the right-hand side of the equation [7] represents the monetary value of total health gains obtained by patients treated with medicine  $i$ . The second addend – in parenthesis – represents the total cost to the system of treating patients with medicine  $i$ . This includes both the drug cost ( $p_i q_i$ ) and the direct cost for the health system ( $T_i(q_i)$ , e.g., visits to the GP, drug administering cost, inpatient days, and visits to emergency).

To define the reserve price  $p$  (the maximum price the payer is willing to pay for the medicine  $i$ ) and the corresponding reserve ICER  $\lambda$  at that price, we need to first define the outside option of the payer  $\Theta^p$ . This outside option is the payoff to the developer where a price  $p_i$  for the medicine  $i$  is not successfully agreed upon, or the payoff obtained by treating  $q_i$  patients with an existing medicine  $j$  (the comparator). Equation [8] represents the outside option of the developer:

$$\Theta^p = B_j(q_i) = \lambda h_j q_i - (p_j q_i + T_j(q_i)) \quad [8]$$

By calculating  $B_i(q_i) - \Theta^p$ , we measure the net monetary value (the surplus) obtained by the payer from using medicine  $i$ ,

$$NB_i(p_i) = B_i(p_i) - \Theta^p = \lambda(h_i - h_j)q_i - (p_i - p_j)q_i - (T_i(q_i) - T_j(q_i)) \quad [9]$$

The system will then adopt the new medicine whenever  $NB_i(q_i) \geq 0$ , that is, when the monetary value of the incremental health benefit exceeds the incremental cost. We define the exact cut-off point in equation [9],

$$B_i(q_i) - \Theta^p = 0 \quad [10]$$

Again, assuming that  $T_j(q_i) = T_i(q_i)$ , then substituting [7] and [8] in [10] and rearranging we obtain:

$$\lambda = \frac{p_i - p_j}{h_i - h_j} \quad [11]$$

This corresponds to a price  $p_i = p$ , the reserve price of the payer given the reserve ICER for the payer of  $\lambda$ .

### 2.3.3 The bargaining problem between the developer and the payer

At stage 2, before the game reaches the bargaining process (stage 5), the payer commits to a maximum level of the threshold  $\bar{\lambda} \in [0, \lambda]$ , which determines their reserve ICER. If the short run reserve ICER of the developer is lower than the reserve ICER of the payer ( $\underline{\lambda} < \bar{\lambda}$ ), then both parties can agree on a final price of the medicine  $p_i$  which is related to a

<sup>19</sup> At this stage we assume that the value of the maximum WTP set by nature can be either the supply side threshold or the demand side threshold. Later in the paper, we will explore the impact of using the supply-side cost-effectiveness threshold  $k$  as in Pandey et al., (2018).

particular level of the ICER  $\lambda_i$  such that  $\underline{\lambda} \leq \lambda_i \leq \bar{\lambda}$ . Such an agreement secures non-negative profits to the payer, and a loss of a smaller size than the sunk cost  $-R_i$  to the developer. Therefore, in a model of value-based pricing with an explicit (or implicit) CET, price negotiations at health system's different contracting levels are equivalent to negotiations over the level of the ICER  $\lambda_i$ . The result of this bargaining process is a price for the medicine and its corresponding ICER i.e. the effective ICER.

We follow the Nash Bargaining approach to model the final accepted ICER (or final price) negotiation. The Nash Bargaining solution divides the total profit of an economic interaction between negotiating parties in such a way that the product of all players' net benefit over the disagreement point is maximised for a given value of their bargaining power. In economic negotiation contexts where players' 'impatience' plays a key role on the final outcome, the Nash Bargaining solution is an efficient approach for modelling (Binmore et al., 1986). This is the case in pharmaceutical markets where the payers' impatience is defined by the clinical need, and developers' impatience defined by the sunk nature of R&D cost. The Nash Bargaining solution has been applied to model pharmaceutical markets in different settings (Dubois et al., 2019; Jelovac, 2015; Grennan, 2013; Garcia-Mariño et al., 2011)

In the context of our model, the level of the agreed ICER net from the reserve ICER determines the share of the pie each player obtains. The Nash Bargaining approach over the level of the ICER can be represented as:<sup>20</sup>

$$\text{argMax}_{\lambda_i} (\lambda_i - \underline{\lambda})^\beta (\bar{\lambda} - \lambda_i)^{1-\beta} \quad [12]$$

Where  $\beta$  is the bargaining power of the developer,  $(1 - \beta)$  is the bargaining power of the payer and  $\alpha \in [0,1]$ . Solving [12] by applying first order conditions, we obtain  $\lambda_i$  as a function of  $\beta$ . As the price is a function of the agreed ICER,  $p_i$  is also obtained. Substituting  $p_i$  into [1] and [9], we obtain the payoffs of the developer and the payer respectively.

Solving [12], we obtain that the Nash Bargaining solution of the ICER:

$$\lambda_i^* = \beta \bar{\lambda} + (1 - \beta) \underline{\lambda} \quad [13]$$

Where  $\lambda_i^* \in [\underline{\lambda}, \bar{\lambda}]$  with  $\frac{d\lambda_i^*(\beta)}{d\beta} > 0$ ;  $\Pi(p_i) \in [0, \Pi(\bar{p})]$  with  $\frac{d\Pi(p_i(\lambda_i^*))}{d\lambda_i^*(\alpha)} > 0$ ; and  $B(p_i) \in [0, B(\bar{p})]$  with  $\frac{dB(p_i(\lambda_i^*))}{d\lambda_i^*(\alpha)} < 0$ .

### 2.3.4 Delimiting the set of possible bargaining outcomes and agreements

To analyse possible solutions of the bargaining process, and the economic and social welfare implications of ICER pricing at micro level, we establish the following set of assumptions which allow us to simplify the payoff functions of both the payer and the developer, and show how bargaining power affects both parties' benefits and social welfare:

<sup>20</sup> Using equations [1], [6] and [9] we write the Nash product as  $(\Pi_i(p_i) - \theta^d)^\beta (B_i(p_i) - \theta^p)^{1-\beta}$ . Given that the product ICER is linear in the price, using [5] and [11] we can define the effective ICER for developers and the payer as a function of the price  $\lambda_i(p_i)$ . Rearranging we write the Nash product as in equation [12]. A detailed proof of this is provided in Appendix 1.

1. We normalise incremental cost and incremental health benefit by assuming  $h_j = 0$ ,  $p_j = 0$  and  $T_j(q_i) = 0$
2. The new medicine always produces positive incremental health benefits:  $h_i > 0$
3. The new medicine is always served at positive incremental cost:  $\underline{p} > 0$
4. Let the direct health care cost of using medicine  $i$  be equal to the direct health care cost of using medicine  $j$ :  $T_i(q_i) = T_j(q_i) = 0$
5. Let the total treatment patient population be  $q_i = Q$

Following the ICER definition in equation [5] and applying assumptions above we can now define the price as a function of the ICER as in the following equation:

$$p_i(\lambda_i) = \lambda_i h_i \quad [14]$$

Substituting [14] into [1] and [9] and applying all assumptions, we have the following payoff functions for the developer and the payer, respectively:

$$\Pi(\lambda_i) = (\lambda_i h_i - c)Q - R_i \quad [15]$$

$$NB(\lambda_i) = (\lambda - \lambda_i)h_i Q \quad [16]$$

Figure 2 graphically delimits the bargaining set – that is, all potential agreements over the final ICER  $\lambda_i^*$  that are incentive compatible (or provide mutual positive payoffs) for both the payer and the developer. The possible mutually beneficial agreements are all values of  $\lambda_i$  such that  $\lambda_i \in [\underline{\lambda}, \bar{\lambda}]$  where  $\bar{\lambda} < \lambda$ . Figure 2 shows that for values of  $\lambda_i$  such that  $\lambda_i \in [\underline{\lambda}, \hat{\lambda}]$ , the profit of the developer is negative but higher than sunk costs ( $-R_i < \Pi < 0$ ). For values of  $\lambda_i$  such that  $\lambda_i \in [\hat{\lambda}, \bar{\lambda}]$ , the payoff to the developer is zero or higher ( $0 \leq \Pi \leq \Pi(\bar{\lambda})$ ).

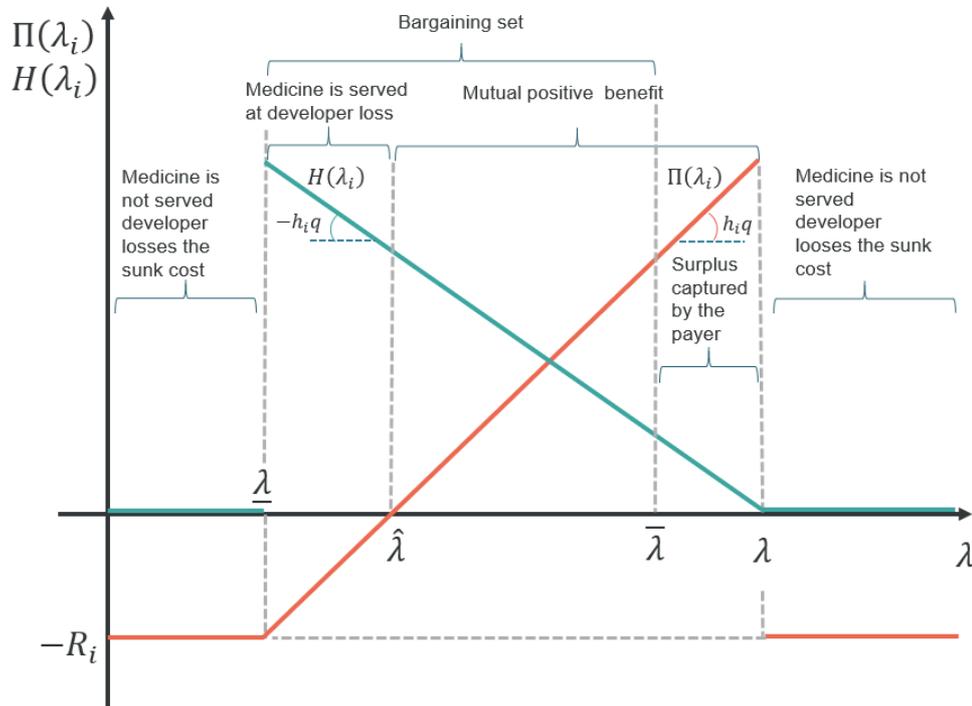
Introducing the Nash bargaining solution [13] into both [15] and [16] we obtain the following:

$$\Pi(\lambda_i^*) = ((\beta\lambda + (1 - \beta)\underline{\lambda})h_i - c)Q - R_i \quad [17]$$

$$NB(\lambda_i^*) = ((1 - \beta)(\bar{\lambda} - \underline{\lambda}) + (\lambda - \bar{\lambda}))h_i Q \quad [18]$$

Where the total surplus  $TS(\lambda_i^*)$  at the micro level – one payer and one developer – is given by the sum of [17] and [18]:

$$TS(\lambda_i^*) = \Pi(\lambda_i^*) + NB(\lambda_i^*) \quad [19]$$



**Figure 2: THE SET OF POSSIBLE AGREEMENTS ON THE ICER**

Note: the light blue line represents the payoff to the payer; the orange line represents the payoff to the developer. The slope of developer's payoff function is the health gain per patient treated times the number of patients treated. The slope of the payer's payoff function is minus the health gain per patient treated times the number of patients treated.

A few notable cases:

1.  $\beta = 1$  : The developer holds all bargaining power. This leads to a result where the developer obtains all surplus<sup>21</sup>. The agreed ICER is equal to  $\bar{\lambda}$ , which corresponds to a price  $p_i(\bar{\lambda}) = \bar{p} = \bar{\lambda}h_i$ , or the maximum price that the developer can charge that the payer is willing to pay. The payoff to the payer is equal to zero  
 $NB(\bar{\lambda}) = \underline{NB} = (\lambda - \bar{\lambda})h_i Q$ , and the developer gets  $\Pi(\bar{\lambda}) = \bar{\Pi} = (\bar{\lambda}h_i - c)Q - R_i$
2.  $\beta = 0$ : The payer holds all bargaining power. This leads to a scenario where the payer gets all surplus. The agreed ICER is equal to  $\underline{\lambda}$  which corresponds to a price  $p_i(\underline{\lambda}) = \underline{p} = \underline{\lambda}h_i$ , or the minimum price at which the developer is willing to sell and the payer is able to buy. The payoff to the developer is  $\Pi(\underline{\lambda}) = \underline{\Pi} = -R_i$ , and the payer gets  $NB(\underline{\lambda}) = \overline{NB} = (\lambda - \underline{\lambda})h_i q$
3.  $\beta \in (0,1)$ : the payer and the developer split the total surplus and the share each player gets is determined by [13]. The payoff to the developer is:

$$\Pi(\lambda_i^*) = ((\beta\bar{\lambda} + (1 - \beta)\underline{\lambda})h_i - c)Q - R_i$$

<sup>21</sup> The developer obtains all the surplus except the surplus created by the difference between the health system's maximum WTP ( $\lambda$ ) and the CET ( $\bar{\lambda}$ ) which is captured by the payer in any case.

and the payoff to the payer is:

$$NB(\lambda_i^*) = (\lambda - \underline{\lambda} - \beta(\bar{\lambda} - \underline{\lambda})) h_i Q$$

Where all possible shares of the surplus are defined by the relative bargaining power of the two players, and the total size of the surplus is determined by the CET ( $\lambda$ ) and health gains ( $h_i$ ).

Table 1 shows some of the different values of  $\lambda_i^*$  corresponding to different levels of developer negotiation power.

**Table 1: BARGAINING POWER OF PLAYERS, EXAMPLES OF AGREED ICER AND PAYOFFS**

DEVELOPER: $\beta$	PAYER: $1 - \beta$	ICER: $\lambda_i(\beta)$	DEVELOPER: $\Pi(\lambda_i)$	PAYER: $NB(\bar{\lambda}, p_i)$
1	0	$\bar{\lambda}$	$(\bar{\lambda} h_i - c) Q - R_i$	$(\lambda - \bar{\lambda}) h_i Q$
0.75	0.25	$0.75\bar{\lambda} + 0.25\underline{\lambda}$	$((0.75\bar{\lambda} + 0.25\underline{\lambda}) h_i - c) Q - R_i$	$(\lambda - \underline{\lambda} - 0.75(\bar{\lambda} - \underline{\lambda})) h_i Q$
0.5	0.5	$0.5\bar{\lambda} + 0.5\underline{\lambda}$	$((0.5\bar{\lambda} + 0.5\underline{\lambda}) h_i - c) Q - R_i$	$(\lambda - \underline{\lambda} - 0.5(\bar{\lambda} - \underline{\lambda})) h_i Q$
0.25	0.75	$0.25\bar{\lambda} + 0.75\underline{\lambda}$	$((0.25\bar{\lambda} + 0.75\underline{\lambda}) h_i - c) Q - R_i$	$(\lambda - \underline{\lambda} - 0.25(\bar{\lambda} - \underline{\lambda})) h_i Q$
0	1	$\underline{\lambda}$	$-R_i$	$(\bar{\lambda} - \underline{\lambda}) h_i Q$

In addition to the effect of bargaining power ( $\beta$ ) and the consequential agreed ICER as the solution of the bargaining, there are other parameters in the result worth exploring. These are the CET ( $\bar{\lambda}$ ) and the sunk cost of R&D ( $-R_i$ ). We explore the economic implications of each of these in the following sections.

## 2.4 Macro level implications of the ICER pricing

To analyse the welfare effects of ICER pricing, we need first to understand the market behaviour at the aggregate level. In previous sections, we analysed the payer perspective in the bargaining model at a micro level, where all income and costs that drive decisions are related to the adoption of a single new technology. To perform a welfare analysis of pharmaceutical markets regulated through a CET and ICER pricing requires additional analysis of the following:

- i. Aggregate demand: how payers behave when using the whole health budget to provide health care to taxpayers, especially to what extent they are concerned about providing access to new medicines
- ii. Aggregate supply: how the health industry (i.e. all developers together) responds to a pricing and reimbursement environment based on ICERs and the CET
- iii. Final equilibrium: how the final price resulting from ICER pricing at the reimbursement decision stage and the bargaining process of medicine at the commercial procurement stage affect the final distribution of the surplus between the payer and the industry

Aggregate demand is determined by the extent of adopted pharmaceutical innovation and total patients treated with new technologies at every price/cost. At the same time, prices paid for new pharmaceutical products depend on the decision of the payer. This decision, in the context of this study, follows a cost-effectiveness rationale. It implies that payers will not pay above a maximum price that reflects some amount of resources per unit of health gain such that overall, the total health care provided meets some predefined objective (e.g., maximise total health gains, provide treatment options to patients with no medical alternatives, not exceed a given budget). The payer sets the CET or the maximum price to payable per unit of health which determines the decision regarding the adoption of each individual new medicine. The CET then determines the total amount of innovation accessed by all patients (all pharmaceutical products eventually adopted), and consequently the total health gain of the society – i.e. the utility of the payer. In the second instance, these medicines have to be made available for patients through commercial arrangements and other procurement and contracting procedures. They are then subject to another round of price negotiations before the medicines are made available to patients. At this point, purchasers may exert their bargaining power to push prices down further. This feature of the market, and the payer-supplier market behaviour, determine the effective ICER at the level  $\lambda^*$ , which we have modelled by applying the Nash Bargaining solution.

The number of new technologies that the industry brings to market at any given price level determines the aggregate supply. At the micro level, the minimum price a developer accepts to sell its product is determined by the reserve ICER. For each level of the CET, all developers whose reserve ICER is below the threshold will supply the new medicine to the payer or the health system, and patients will have access. Therefore, at each level of the threshold there will be a number of new medicines that become accessible to patients, these jointly form the supply. Suppliers are also subject to a second round of price negotiations at procurement stage.

There are two points at which relevant pricing behaviour occurs: the HTA or pricing and reimbursement stage (value-based price), and the procurement stage (final market price). In this paper we introduce the second stage to generalise the model of the CET regulated pharmaceutical markets.

The distribution of bargaining power at the second stage can be case specific. It depends on several factors such as the existing degree of competition, the market power of the two parties, volume and purchasing capacity of the contracting body, product innovativeness and/or clinical need. However, for the purpose of this paper we make the simplifying assumption that purchasers' and developers' bargaining power distribution is homogeneous across all technologies under negotiation.

#### 2.4.1 The aggregate demand of new health technologies: the payer perspective

To understand the aggregate demand for (all) new medicines, and the implications of the bargaining game, we need to focus on the payer utility function at the macro level. In other words, the overall utility the payer gains by using the health care budget to provide and finance health services to the whole of society (or at least those citizens who are enrolled in or entitled to access care within the relevant health system).

In ICER-threshold based markets for medicines, the overall utility of the payer – while the medicine is still on-patent<sup>22</sup> – depends on the following factors: the maximum WTP of the payer defined in stage 1 by nature, the level of the CET, or the ICER at which the new technology is recommended at the reimbursement decision stage, and the outcome of the market bargaining at the commercial and procurement level.

In our model, the ability to set the CET rests on the payer (demand) side, with the maximum acceptable level being the payer's maximum WTP. There are two main approaches discussed in the literature regarding how the payer should set this maximum WTP of the system (Vallejo-Torres et al., 2016; Thokala et al., 2018): using the health system opportunity cost of generating an additional unit of health gain (the 'supply side' threshold), or using society's preferences and WTP per additional unit of health gain (the 'demand side' threshold).

We follow the Pandey et al., (2018) model where the key assumption is that the system maximum WTP per unit of health gain is determined by the supply-side threshold (referred to as  $k$ ). The health budget of the payer is fixed, implying that the adoption of a new technology necessarily displaces other technologies or treatments already in use in the health system. Therefore, the CET reflects the health system opportunity cost. If the health system spends (at the margin) an amount  $k$  to produce a unit of health gain – the value of the supply-side CET – the price for a unit of health gain a new health technology delivers to the health system should never exceed that amount. Under a fixed health budget, paying more implies that health foregone exceeds health added thus leading to a decrease in total health produced with the same level of resources. Paying less than  $k$  means that more health is added than health foregone – an increase of the total health produced with the same resources. Furthermore, we assume that it is a policy option for the payer to fix the CET below the maximum WTP in the attempt to maximise total health produced<sup>23</sup>.

Whether the threshold that maximises the health production is somewhere close to the maximum WTP (the supply-side CET) or far below it depends on how much medical innovation developers generate at each level of the threshold, which in turn depends on R&D costs, manufacturing costs, and developers' reserve ICERs distribution<sup>24</sup>. With values of the threshold closer to the health system opportunity cost, developers with higher reserve ICERs enter the market. The health gains delivered by new medicines are obtained at a higher price. However, this also implies that the marginal net health gain for the payer with a fixed budget decreases as the level of the threshold approaches the supply-side CET, where the health gained net of health foregone is zero. The result depends on which of the two effects dominates: the increase in total health gain obtained through more cost-effective technologies entering the market, or the decrease in total health gain due to increased price per unit of health gain.

---

<sup>22</sup> Note that the surplus from any medicine is the discounted stream of surpluses over the useful lifetime of the treatment (i.e. the period during which it is used by the health system and not superseded by a superior or more cost-effective treatment). Given that for most products, the loss of intellectual property protection leads to generic or biosimilar entry, payers can expect to pay lower prices from this point. If competition is great, then prices will fall to long run production and distribution costs.

<sup>23</sup> In Pandey et al. (2018) the consumer surplus is measured by the net health gain of the payer hence maximising the total net health gain, the payer also maximises the consumer surplus.

<sup>24</sup> We did not discuss this policy option in section 2.3.3 when we characterise the Nash bargaining solution. In that section, we assume that the value of the Cost-effectiveness threshold of the payer is fixed exogenously at the maximum WTP and the effective ICER is set by the relative bargaining power of players.

Let the relationship between the net health gain and the threshold level (the maximum level of the ICER the payer imposes on the developer) be measured by the following linearly decreasing function:

$$h(\lambda_i) = a - aF(\lambda_i) \quad [20]$$

where  $\lambda_i \in [0, k]$  and  $F(\lambda_i)$  is the distribution of the reserve ICERs of the developers. Following Pandey et al. (2018) we assume that the distribution function of the reserve ICERs of the developers is uniform. We can write the distribution function, entirely defined within the interval  $[0, k]$  as:

$$F(\lambda_i) = \frac{\lambda_i}{k} \quad [21]$$

When we introduce equation [21] into equation [20], we show that at a level of the threshold equal to the supply-side CET,  $\lambda_i = k$ , the net health gain is zero. Equation [20] measures the marginal net health gain for the system by adopting the new treatment for every threshold level below or equal to the supply-side level ( $\lambda_i \leq k$ ). This is positive for all values of the threshold below the opportunity cost, decreasing with the value of the threshold and equal to zero when the value of the threshold is equal to the opportunity cost.

The form of the  $h(\lambda)$  is determined by the underlying technological possibilities, the health production function of the system, and by the distribution of the reserve ICERs of the developers. The linear relationship between net health gain and threshold value is implied by the uniformity of the reserve ICERs of the developers. We also analyse the implications of a case with a non-uniform distribution of the reserve ICERs of developers – which may be caused by the production technology or the production cost function – which produces a non-linear  $h(\lambda)$  function.

For a payer using the supply-side perspective, the outcome of interest is the health benefit to patients provided by the new technologies net of health cost to patients due to the displacement of existing treatments or health services. The measure of interest can be stated as the Net Population Benefit (NPB), defined as the monetary value of the health benefit provided by new technologies, net of the monetary value of the health foregone by the displacement of existing technologies<sup>25</sup>. It represents the consumer surplus gain from adopting the new technology.

Pandey and colleagues (2018)<sup>26</sup> define a ‘consumer threshold curve’ (or CTC) as the curve that plots the relationship between the payer’s CET, and NPB. We adopt the same approach to define our baseline model.

We assume that the reserve ICERs of developers  $\lambda$  are uniformly distributed along the interval  $\lambda \in [\lambda^m, \lambda^M]$ . For the sake of simplicity, we also assume that the minimum reserve ICER among all of the developers in the market is equal to zero (i.e.  $\lambda^m = 0$ ). The maximum

---

<sup>25</sup> Net population benefit is a measure of the consumer surplus that implicitly assumes that consumer surplus is strictly positive if and only if the threshold level is strictly below the opportunity cost of the health system. For values of the threshold that are above the opportunity cost, the consumer surplus is necessarily negative as more health is foregone than incorporated. This is not the case when payers have bargaining power, in whose case, the net monetary value of using CETs above the opportunity cost of the health system can still be positive because part of the benefit that companies make can be recovered through price reductions following price negotiations. We explore these cases later in the paper.

<sup>26</sup> Pandey H., Paulden M., McCabe C., 2018. Theoretical models of the cost-effectiveness threshold, value assessment, and health system sustainability. Edmonton (AB): Institute of Health Economics.

reserve ICER of all of the developers potentially entering the market is represented by  $\lambda^M$ . Developers assume that the payer has a CET  $\lambda$ , which is set at  $k$  (the maximum WTP of the health system) and that  $\lambda^M > k$  i.e., the maximum reserve ICER among all of the developers in the market is higher than the maximum WTP of the payer. Again, for the sake of simplicity we also assume that all relevant reserve ICERs in this analysis are within the interval  $\underline{\lambda} \in [0, k]$ , as the payer maximum WTP is  $k$  and new health technologies with ICERs higher than  $k$  are never recommended.

We represent the CTC with the following function,

$$NPB = (1 - \beta) \int_0^{\lambda} h(x) dx + \beta \lambda h(\lambda) \quad [22]$$

Where  $\beta$  is the bargaining power of the developer;  $\lambda$  is the threshold fixed by the payer to inform health technology adoption decisions (the maximum WTP per health gain); and  $x$  represents the reserve ICER of each developer  $\underline{\lambda}$ .

The interpretation of [22] is as follows: the first addend measures how much surplus the payer can extract from the developer by exploiting their bargaining power. As the maximum surplus that each developer can gain is measured by the difference between the fixed CET of the payer and its individual reserve ICER ( $\lambda - \underline{\lambda}$ ), the integral captures the sum of all individual maximum surpluses. A proportion equal to its bargaining power ( $1 - \beta$ ) is extracted by the payer. This is an additional structure we have incorporated into the CTC used in Pandey et al., (2018) where they assume developers have all bargaining power ( $\beta = 1$ ).

The second addend measures the surplus obtained by the payer with the level of the CET fixed at  $\lambda$ , which is below the level of the supply side CET,  $k$ . This is obtained entirely by the payer regardless of the bargaining power of the developers, and is equal to zero when the payers' CET is equal to the health system maximum WTP ( $\lambda = k$ ). As the  $(1 - \beta)$  proportion of this surplus is accounted for in the integral of the first addend, the second addend adds the proportion  $\beta$  that is not captured in the first addend.

Substituting [20] into [21] and solving the integral we have:

$$NPB = (1 - \beta)\lambda\left(a - \frac{a\lambda}{2k}\right) + \beta\lambda\left(a - \frac{a\lambda}{k}\right) \quad [23]$$

Applying the first order conditions with respect to  $\lambda$  to [23] we obtain:

$$(1 - \beta)\left(a - \frac{a\lambda}{k}\right) + \beta\left(a - \frac{2a\lambda}{k}\right) = 0 \quad [24]$$

Rearranging [24,] we obtain [25], which shows that the CET that maximises the payer's surplus is the one that leads to an equalisation of (i) the marginal benefit of the last new technology entering the market in response to the marginal increase of the CET, and (ii) the marginal cost to the payer now that all technologies paid for at a value-based price are at a higher price corresponding to the increased CET<sup>27</sup>.

---

<sup>27</sup> We note that only medicines (and perhaps medical devices) are likely to be priced at the CET such that an increase in the CET leads to an increase in the prices paid for all new interventions. Even in the case of medicines, already launched products are usually not eligible for price increases – the US being an exception. However, we are assuming that payers are anticipating that all new medicines will be launched at prices that reflect the CET that has been set by the payer.

$$ak = a\lambda(1 + \beta) \quad [25]$$

The left-hand side of [25] reflects the marginal health gain of the last new technology adopted. An increase in the CET implies that new technologies are adopted, as their reserve ICERs are now considered cost effective. These new adoptions increase net population benefit as they provide health benefits to patients at any distribution of the bargaining power. This effect is what we call the *supply effect* on net population benefit, which represents the increase in net population benefit resulting from additional medicines being adopted, which in result from a marginal increase in the threshold.

The right-hand side of [25] reflects the higher price that the payer needs to pay per unit of health gain after an increase in the CET. We call this effect *the price effect - or the demand effect* - on the net population benefit. It represents the increase in the cost of delivering net health gains to society per marginal increase in the threshold. The size of the price effect to the payer is higher for higher values of the developers' bargaining power  $\beta$ . Higher values of  $\beta$  increase the right-hand side of the equation, hence the payer ends up paying a higher price not only for the new technology, but also for all technologies that they were already buying before the threshold increase. When  $\beta = 0$ , this latter effect does not exist, and the payer extracts the surplus of the developers up to where the price of the marginal developer (i.e., the last developer entering the market) is set at its marginal cost of production.

The value of the threshold that maximises the NPB is such that the *supply effect* is equal to the *demand effect*. The point at which these two effects are equal will depend on the elasticity of  $h(\lambda)$ , or how much the net health gain decreases in percentage terms per 1% increase in the threshold  $\lambda$ . Solving equation [25] for the value of  $\lambda$ , we obtain the level of the CET that maximises the net population benefit, the consumer surplus of the payer:

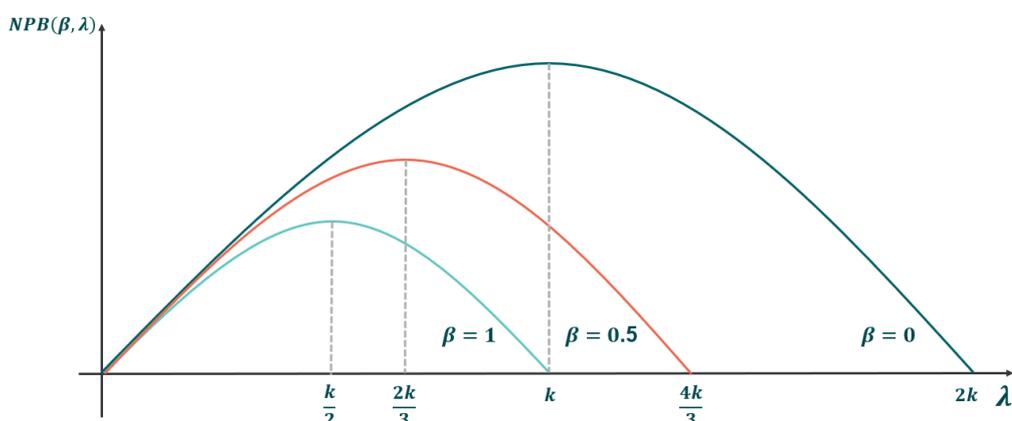
$$\lambda_p^* = \frac{k}{1+\beta} \quad [26]$$

The value  $\lambda_p^*$  is increasing in  $k$  and decreasing in  $\beta$ . An increase in  $k$  implies that payer is willing to pay a higher maximum price per unit of additional health benefit and therefore the level of its surplus-maximising CET will be higher as well, because the Maximum WTP has increased. In the model, this means that the  $h(\lambda)$  function pivots to the right on  $a$ , increasing the surplus contribution for the payer of the last adopted technology. In contrast,  $\lambda^*$  decreases as developer bargaining power increases. Higher bargaining power on part of the developer means that the new medicine can be priced closer to the CET level, thus reducing the surplus to the payer, which is the difference between the CET and the agreed final ICER. Each additional health gain costs more to the payer. Table 2 shows the values of  $\lambda^*$  and maximum WTP of the payer, expressed in terms of the supply-side CET  $k$  and for different bargaining powers  $\beta$ . Table 2 also includes the value of the developer's bargaining power that leads to the scenario of  $\lambda^* = k$ .

**Table 2: SURPLUS MAXIMISING CET AND WTP FOR THE PAYER BY BARGAINING POWER**

Developer: $\beta$	Payer: $1 - \beta$	Optimal payer CET: $\lambda_P^*$	Maximum WTP
1	0	$k/2$	$k$
$3/4$	$1/4$	$4k/7$	$8k/7$
$1/2$	$1/2$	$2k/3$	$4k/3$
$1/4$	$3/4$	$4k/5$	$8k/5$
0	1	$k$	$2k$

Figure 3 illustrates how changes (of different magnitudes) in the relative bargaining power of the payer and the developer affects the CTC and NPB. The surplus of the payer is maximised at higher values of the ICER when their own bargaining power increases and the bargaining power of the developer decreases. The maximum WTP also increases when the payer's bargaining power increases. This is due to the fact that, when the threshold is increased to obtain new technologies, the extra amount paid for non-marginal existing interventions will be lower.



**Figure 3: SURPLUS OF THE PAYER (NPB) BY BARGAINING POWER AND CET**

#### 2.4.2 The aggregate supply of innovative medicines: the industry perspective

The producer surplus, at the industry level, is the sum of the surplus of each producer that develops, manufactures and sells a new technology with a reserve ICER below (or equal to) to the CET set by the payer<sup>28</sup>. The formal expression for the aggregate surplus of the industry is:

$$DS = \beta \left[ \int_0^\lambda \left( a - \frac{a}{k}x \right) dx - \lambda \left( a - \frac{a}{k}\lambda \right) \right] - F(\lambda)R \quad [27]$$

<sup>28</sup> In a more general specification, we could also incorporate the cost of "failed" projects in R&D cost  $R_i$ .

The first addend of the equation (in brackets) represents the share of the total surplus that the industry obtains at each level of the threshold, given their bargaining power  $\beta$ .

The second addend represents the proportion of the total cost of R&D incurred together by all developers with a threshold equal to the highest reserve ICER that is relevant for the payer, this is also equal to its maximum WTP ( $\lambda = k$ ), given the distribution function of the reserve ICERs.

Following the timing of the model presented in section 2.2, this proportion is equal to one as developers take their decisions about whether or not to invest in R&D following the CET value determined by nature in stage 1, which we assume is equal to the supply-side CET  $k$ . Therefore, we have that  $F(k) = 1$  as per equation [21].

By substituting  $F(k) = 1$  in equation [27] and rearranging, we have:

$$DS = \frac{\beta a \lambda^2}{2k} - R \quad [28]$$

The distribution of bargaining power also affects producer surplus. If developers have all the bargaining power, i.e.  $\beta = 1$ , then the surplus of the industry is the sum of all individual developers' surpluses when they all set the final price and ICER to be equal to the CET. If the payer has all the bargaining power, i.e.  $\beta = 0$ , then the final price is set at the level of each developer's reserve ICER  $\lambda$  by the bargaining process, or at each developer's minimum willingness-to-accept. By previous analysis (see equation [3]), at the reserve-ICER price, each producer covers manufacturing costs, but all R&D costs will be lost<sup>29</sup>.

If the cost of R&D is excluded from developer surplus, then only the surplus relevant for the static efficiency is measured and previous cost of R&D is not considered in the calculation. Where surplus is zero, developers experience negative profits equal to the size of the sunk R&D costs. This is captured by removing the R&D cost component from equation [27], which gives:

$$DS' = \beta \left[ \int_0^\lambda \left( a - \frac{a}{k} x \right) dx - \lambda \left( a - \frac{a}{k} \lambda \right) \right] \quad [30]$$

Calculating integrals and rearranging, we have:

$$DS' = \frac{\beta a \lambda^2}{2k} \quad [31]$$

The industry surplus then is calculated by [31], which is strictly higher than [28] by an amount equal to the total industry investment in R&D at the level of the threshold<sup>30</sup>. We compare  $DS$  to  $DS'$  to explain how the incorporation of R&D costs affects the level of the socially optimal CET, as well as the level of threshold that individually maximises payer surplus.

Both  $DS$  and  $DS'$  are upward concave functions, increasing in  $\lambda$  (the threshold). Therefore, there is no finite value of  $\lambda$  that maximizes  $DS$  and  $DS'$ . The shape of  $DS$  and  $DS'$  can be explained in two steps:

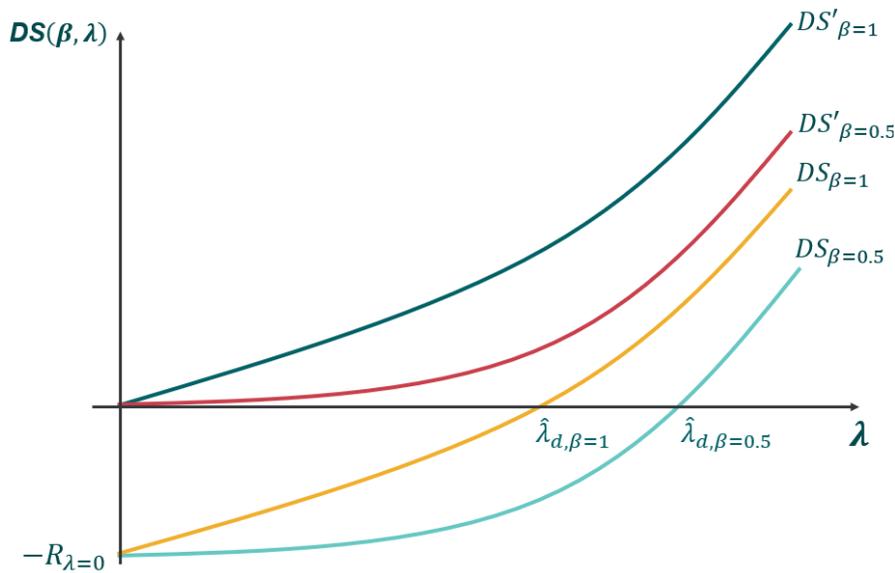
<sup>29</sup> Given that the cost of R&D is sunk cost, the aggregate curve of industry's surplus should also include the costs of those R&D projects that never reach the market.

<sup>30</sup> Using [29] into [28] and comparing to [31] we can calculate the difference between  $DS$  and  $DS'$  to be  $\frac{\lambda}{\lambda M} R$ , or the total industry's R&D investment when the level of the threshold is  $\lambda$ .

1. If there is no entry of new technologies, an incremental increase of the threshold produces an increase of the same proportion in the profit of developers who are already supplying their technologies. This effect guarantees that  $DS$  and  $DS'$  are, at least, linearly increasing in  $\lambda$ .
2. Additionally, as we have assumed that reserve ICERs are uniformly distributed within the threshold level range  $[0, k]$ , an incremental increase of the threshold causes the entry of new technologies with reserve ICER is below the new threshold level. This adds additional positive value to the total surplus obtained by the industry.

These two effects together shape the industry's surplus (profit) function, which is represented by  $DS$ , an increasing and concave curve that depends on  $\lambda$ . At the aggregate level, the increase in the surplus of the whole industry as a response to an increase of the threshold of a constant size is increasing.

Figure 4 illustrates the relationship between  $\lambda$ , R&D costs and developers' surplus. To examine how bargaining power affects the  $DS$  curve and its graphical shape, we look at the value of  $\beta \in [0, 1]$ . As it is a non-negative number with lower and upper bounds at zero and one respectively, it is effectively a proportion. The  $DS$  increases/decreases proportionally with higher/lower values of  $\beta$ . With changes of the bargaining power,  $DS$  curve pivots over the origin ( $-R$  when  $\lambda = 0$ ). It pivots down (up) when  $\beta$  decreases (increases).



**Figure 4: SHAPE OF THE DS AND DS' FOR DIFFERENT BARGAINING POWER**

Note:  $-R_{\lambda=0}$  is the loss of R&D of the industry at a threshold level equal to zero.

If the (ex-ante) R&D investment of the industry is incorporated into the surplus analysis, the level of the threshold necessary to produce positive profits to developers is strictly positive. Below this level, there are developers that are willing to sell the new technology because the total loss is lower than the invested R&D costs. However, although this may be statically efficient because access to patients is maximised, it is not dynamically efficient as the industry will stop investing in R&D in the long term and the resulting level of innovation will be suboptimal. Dynamic efficiency is only met when the threshold level is

set at non-negative values of  $DS$ . This condition is met when equation [28] is set equal to zero, which produces the following solution:

$$\hat{\lambda}_d = \sqrt{\frac{2kR}{\beta a}} \quad [32]$$

Higher values of the developers' bargaining power mean that the level of CET necessary to meet the condition in [32] is lower. This is because more bargaining power means greater ability to extract surplus by pricing new medicines closer to the threshold. However, when payers have some bargaining power, the prices of new medicines may fall below the threshold and part of developers' profit will be extracted by the payer. To avoid dynamically inefficient market outcomes, the threshold should be increased until the condition in [32] is met. This is shown in Figure 4 by the values  $\hat{\lambda}_{d, \beta=1}$  and  $\hat{\lambda}_{d, \beta=0.5}$  which are the threshold levels that ensure dynamic efficiency for values of developer's bargaining power of  $\beta = 1$  and  $\beta = 0.5$ , respectively.

### 2.4.3 The social optimum: maximising the social welfare

We now assume that the objective is to maximise the social surplus, which is the sum of developer and payer surplus:

$$SW(\lambda) = NPB(\lambda) + DS(\lambda)$$

Using equation [23] for  $NPB(\lambda)$  and equation [28] for  $DS(\lambda)$ , the objective function can be rewritten as:

$$argMax_{\lambda} (1 - \beta)\lambda \left(a - \frac{a}{2k}\lambda\right) + \beta\lambda \left(a - \frac{a\lambda}{k}\right) + \frac{\beta a \lambda^2}{2k} - R \quad [33]$$

Applying the first order conditions and rearranging, we get:

$$\lambda^* = k \quad [34]$$

The optimum level of threshold from a societal perspective is the payer's maximum WTP per unit of health gain. For the case considered here, this is equal to the supply-side threshold. This results from the fact that the surplus of the developers is an exponentially increasing function of the threshold level. The social optimal solution is the maximum threshold level for which the payer surplus is non-negative<sup>31</sup>. The level of innovation is maximised at the CET corresponding to the payers'/consumers' maximum WTP. If we assume that the developer has all the bargaining power ( $\beta = 1$ ), then all the surplus generated by innovation will be captured by the industry by equating the ICER to the supply-side CET (in stage 4). For cases where bargaining power is shared between the payer and the developer, part of the surplus is recovered by the payer through negotiations of the price at the procurement and contracting stage (stage 5). This last case is relevant for further exploration. Medicines face on-patent and class competition during patent term, as well as generic competition once the patent expires. Both pre- and post-patent

<sup>31</sup> This is the result that Danzon et al. (2015) and Danzon et al. (2011) arrive at, which leads them to propose value-based differential pricing with  $\lambda^* = k$  for the period of patent protection. Note that our baseline model (following Pandey et al. (2018)) does not separate pre-and post- patent time periods – indeed it does not have a post-patent period.



where surpluses of both players (i.e. payer and developers) are equal. That happens with  $\lambda = \lambda^e$  and social surplus is evenly split  $NPB(0.5, \lambda^e) = DS(0.5, \lambda^e)$ .

## 3 The functioning of the market for new medicines

In the previous section, we characterised the behaviour of supply – at both the individual and industry level – and of demand from a single payer at both a micro and macro perspective. From both analyses, we have characterised some of the economic implications of a system using CETs to inform decisions on the adoption of new health technologies. We depart from the theoretical approaches proposed in the current debate and incorporate additional structure to address two well-established characteristics of the market of new medicines. These are:

1. The sunk cost of R&D for developers: this was incorporated to explore how such costs affect the long-run efficiency of developing innovation via reserve ICERs and surplus of the developers.
2. A bargaining approach that defines the negotiation over the net price of a new medicine. We modelled the negotiation of the net price as a negotiation over the effective ICER (i.e. the actual ICER resulting from the application of the net price agreed in the negotiation)<sup>32</sup>. The outcome of the bargaining process is determined by the distribution of the bargaining power.

In this section, we present a theoretical discussion of the economic implications of using CETs and ICER pricing to underline the decision-making process for new medicines. We produce different scenarios characterised through the parameters of the model using either the individual micro-level framework, or the aggregate macro-level framework.

### 3.1 The baseline model

For the baseline, we first consider the case of monopolistic developers who have complete bargaining power ( $\beta = 1$ ), and whose R&D investment is equal to zero ( $R_i = 0$ ). Additionally, we assume that all reserve ICERs of the developers are uniformly distributed along the entire range of their possible values ( $\lambda \in [0, \lambda^M]$ , where  $\lambda^M > k$ ). This particular case corresponds to the general result by Pandey et al. (2018) with the incorporation of a formal expression for the consumer threshold curve (CTC). The CTC measures net population benefit (NPB) at every value of the threshold ( $\lambda$ ). We explore the economic implications in a one-to-one bargaining framework, then extend the discussion to the aggregate level.

We solve the model using backward induction. In stage 4 developers decide the ICERs of their new medicines. Which in turn determines their final price. Developers will propose an ICER equal to the payer's announced CET or  $ICER_d = \bar{\lambda}$ . In stage 3, the payer is aware of the developer's strategic pricing and sets a level of the ICER that maximises their own surplus  $\lambda_p^* = \frac{k}{2}$ .

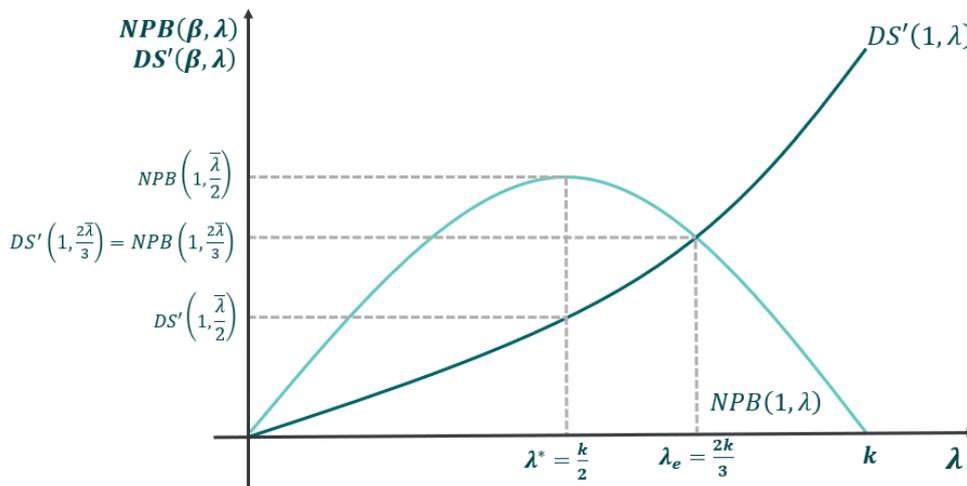
---

<sup>32</sup> A proof of equivalence is provided in Appendix 3.

Going back to Figure 1, we can see that Stages 2 and 5 are not applicable because the cost of R&D and the bargaining of the effective ICER are not considered. The combination of stages 3 and 4, plus the decision from nature about  $k$  in stage 0, produce the result in stage 6. Figure 6 illustrates the level of the effective ICER ( $\lambda^* = \lambda_p^*$ ), and both consumer and producer surplus at this.

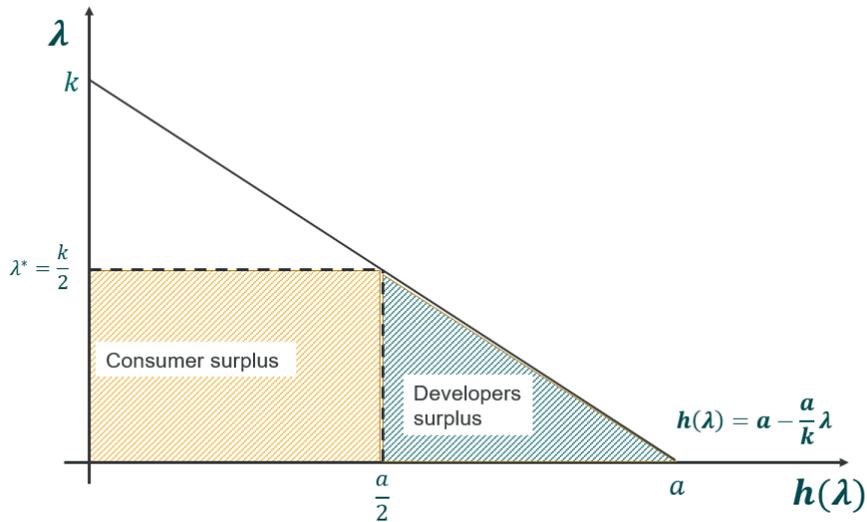
With this result, consumers (and the payer acting as their agent) would obtain the maximum surplus through maximising net health benefits (health added to the health system less the health foregone) by choosing the level of threshold at which all new health gains are willing to be paid by the payer. The payer uses their authority to secure a target level of surplus, while developers are able to obtain a positive surplus by exerting their full bargaining power to price all new medicines launched at the level of the effective threshold. Developers whose reserve ICER is above the effective threshold will not supply the medicine. Figure 7 shows these two areas of surplus, which are defined by the net health benefit function  $h(\lambda)$ .

Alternatively, we also consider other potentially appropriate divisions of the combined surplus between consumers and developers aimed at balancing the needs of the health system and those of innovators<sup>33</sup>. We assume, initially, that this division is achieved through a level of the threshold (e.g. selected as a policy option) that equates the surplus of consumers and developers. This level is calculated by equating equations [31] and [22] under the assumption of  $\beta = 1$ . Rearranging, we find this threshold to be  $\lambda_e = \frac{2k}{3}$ . This result is also represented in Figure 6 by the intersection of CTC and DS'.



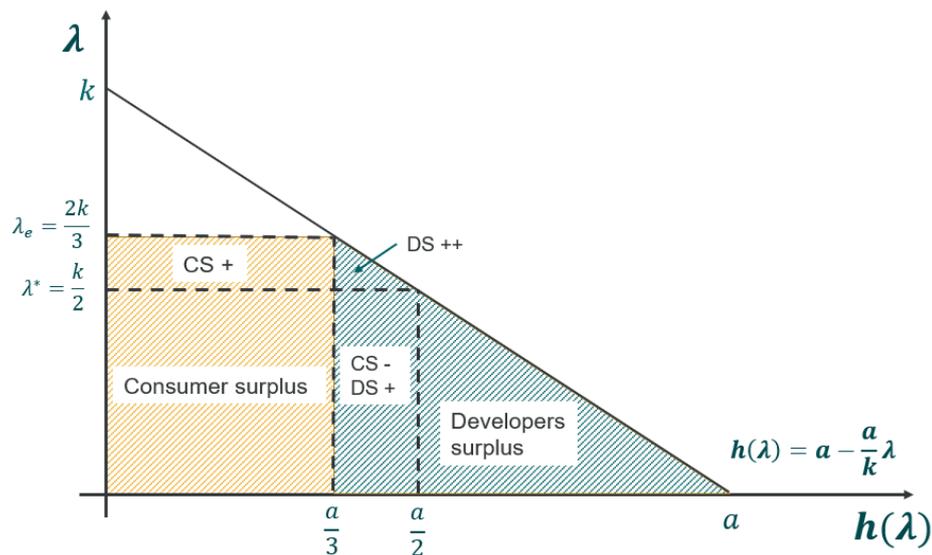
**Figure 6: NPB MAXIMISING PAYER AND NO R&D COST CONSIDERATION**

<sup>33</sup> Such alternative allocations of the surplus like the equal distribution of the market surplus between developers and the payer are also discussed in Pandey et al. (2018)



**Figure 7: CONSUMER AND PRODUCER SURPLUS WITH NPB MAXIMISATION OBJECTIVE**

In Figure 8 below, we can see that consumers lose the area CS- which is transferred to the developers (DS+). Consumers gain the area CS+ which is smaller than the one lost CS-. This is the reason the equal surplus solution does not maximise the NPB. Developers increase their surplus by DS+ and DS++.



**Figure 8: CONSUMER AND PRODUCER SURPLUS WITH EQUAL SURPLUS OBJECTIVE**

Opting for a solution that maximises the total surplus requires – as equation [32] shows – the CET to be set equal to the maximum WTP of the system. By paying the highest price consumers are willing to pay, the number of new medicines launched is maximised. All

surplus generated by health gains is accrued by developers, and – under the assumption of fixed budgets – the total amount of health gain remains constant.

There is a continuum of CETs within the range  $\lambda \in \left(\frac{k}{2}, k\right)$  that can be implemented in response to a predefined social objective on how to appropriately share the surplus – i.e. an equal allocation of the total surplus between stakeholders.

### 3.2 R&D cost of innovation, reserve ICERs and dynamic efficiency

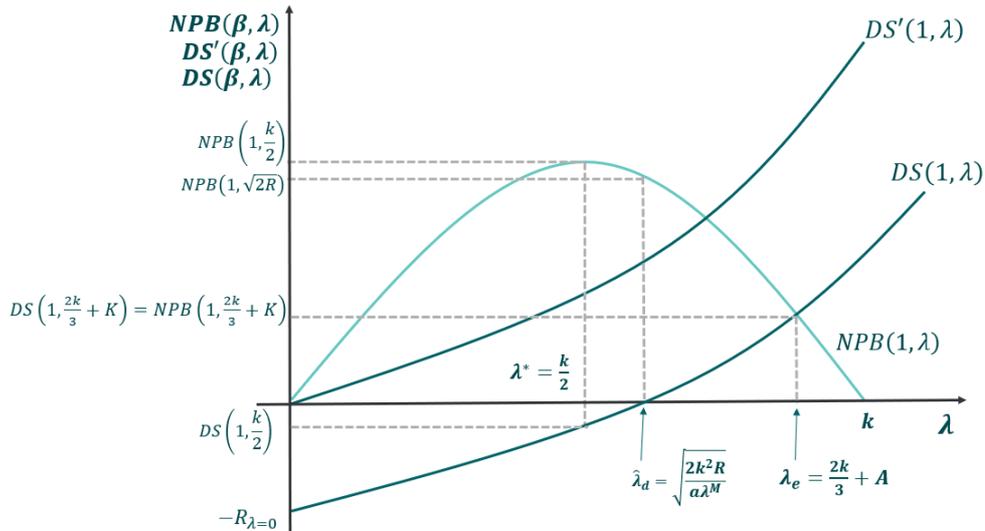
Let us assume now that developers have sunk costs of investing in R&D in stage 3. These sunk costs will affect the developer's long-term reserve ICER. As in the baseline case, when considering the aggregate industry supply and the CET regulated market for new medicines, there is a strictly positive value of the CET below which no developer is willing to sell. This level of the threshold corresponds to a price level lower than the marginal cost of production. In this situation, developers (or manufacturers) make a loss and hence choose not to supply the new medicine.

However, the main implication of the incorporation of the R&D cost has to do with the nature of sunk costs. For a range of prices above the level of the reserve ICER ( $\underline{\lambda}$ ), the developer incurs losses (negative profits) but will continue to supply the medicine as the size of this loss is smaller than losing all the R&D investment by not selling the medicine. Up to the point where the CET reaches  $\hat{\lambda} > \underline{\lambda}$  (see Figure 1) where the price exceeds the marginal cost of production sufficiently to generate a level of profit that covers the R&D cost, developers will supply at a loss.

We will refer to the reserve ICER  $\underline{\lambda}$  as the '*statically efficient*' threshold - that is, the level at which the price of the new medicine is equal to marginal cost and maximises access in the short-term. We will refer to  $\hat{\lambda}$  as the '*dynamically efficient*' threshold where the price exceeds the marginal sufficiently to allow the developer enough commercial margin to cover the R&D cost as well. Effective thresholds within the range  $\lambda \in [\underline{\lambda}, \hat{\lambda})$  generates economic incentive to launch the new medicine but undermines the incentive to keep investing in R&D for the development of future medicines. The closer the threshold is to  $\underline{\lambda}$  (the lower bound of the range), the closer we get to *static efficiency* as access to medicines is maximised in the short run. However, to sustain a healthy level of investment in future pharmaceutical innovation, the threshold should be above  $\hat{\lambda}$ . Otherwise, the intertemporal level of innovation produced and accessed will be suboptimal. A threshold above (or equal to)  $\hat{\lambda}$  is necessary to send the right signal that investment in pharmaceutical innovation receives a competitive rate of return.

If the objective for the payer is to maximise NPB, it is possible that where R&D investment is large enough, the market outcome will be dynamically inefficient. Figure 9 shows this case. At the level  $\lambda^*$ , developers do not obtain a positive rate of return. Although medicines already developed (and launched) would not be withdrawn from the market, this outcome would undermine future investment in R&D. Therefore, the resulting amount of innovation produced in the long-term would be suboptimal. In such a situation, the payer would maximise the short-run surplus at the expense of future health benefits.

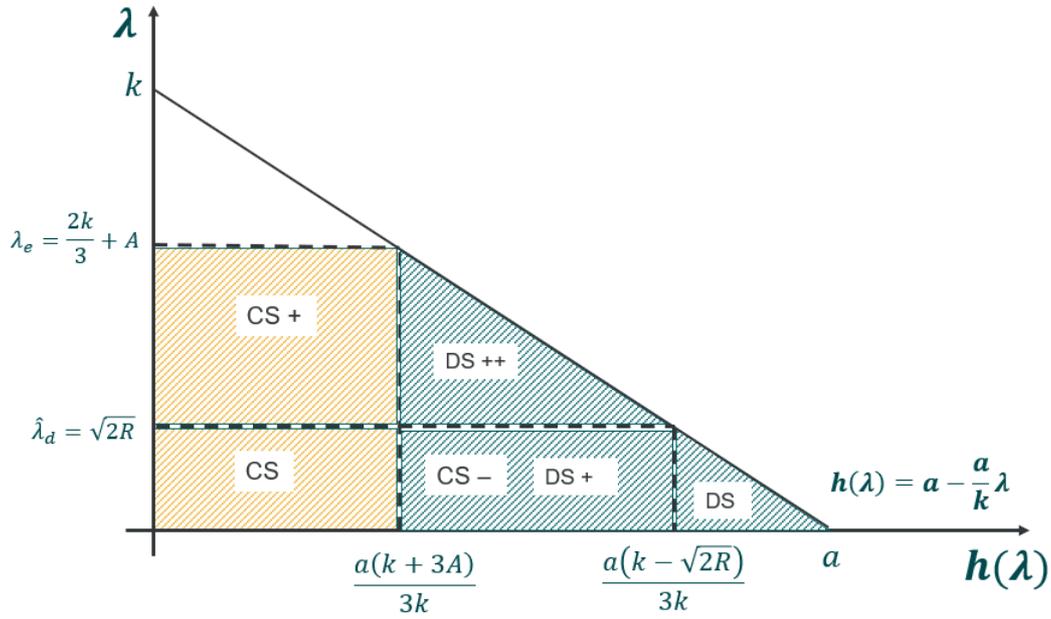
To allow industry to make enough return to cover the investment in R&D, a threshold of  $\hat{\lambda}_d$  – as proposed in equation [32] above – is required. At this CET level, the payer maximizes NPB subject to the condition of developers having a non-negative surplus. Once again, an equal distribution of the surplus generated by new medicines requires a higher threshold ( $\lambda_e$  in Figure 8). When the R&D cost is introduced to the analysis, the surplus equating threshold places closer the maximum WTP of the payer<sup>34</sup> – the one set by the nature at stage 1 or  $k$  as assumed for our analysis.



**Figure 9: NPB MAXIMISING PAYER WITH R&D COST**

Figure 10 shows the distribution of total surplus at the threshold levels  $\hat{\lambda}_d$  and  $\lambda_e$ . With the former, only the payer gets a positive surplus while the developers obtain zero; with the latter, the total surplus generated is equally distributed between the payer and developers.

<sup>34</sup> Solving the problem of equal surpluses when accounting for industry's R&D cost produces the following second-degree equation:  $3a\lambda^2 - 2ka\lambda - 2kR$ . Solving only for the positive value of  $\lambda$  we have that  $\lambda = \frac{k}{3} + \frac{\sqrt{4a^2k^2 + 12aRk}}{6a}$ . If we assume that the second addend within the square root is equal to zero, then we have that  $\lambda = \frac{2k}{3}$ . We know that the second addend within the square root is strictly positive – assume it is an amount that makes a difference of  $A$  in the result. Then the final value of the threshold that equates surplus of both players is  $\lambda_e = \frac{2k}{3} + A$



**Figure 10: SURPLUS DISTRIBUTION AND DYNAMIC EFFICIENCY WITH R&D COST**

At threshold level  $\hat{\lambda}_d$ , the consumer surplus is shown in Figure 10 as the areas CS and CS-, while the producer surplus is shown as the area DS. DS is sufficient to cover R&D cost but does not secure a positive rate of return. It only guarantees zero profit level after recouping R&D investment. If the threshold is set where there is an equal share of total surplus between consumers and developers, consumers increase the surplus accrued by CS+, and reduces it by CS-, a smaller amount than CS+, while developers increase the surplus by DS+ (= CS-) and DS++.

In summary, the threshold level that maximizes payer surplus is *statically* efficient but may not be *dynamically* efficient when the developer's R&D costs are large enough. As a result, if securing an optimal amount of innovation in the long-term is part of the payer's (or policy makers) objective, then the CET needs to be fixed closer to the maximum WTP,  $k$ .

The question remains then: at what point are R&D costs large enough to render the maximisation of the NPB not dynamically efficient? The answer is when the dynamic reserve ICER is above the NPB maximizing CET, i.e. where  $\hat{\lambda}_d > \lambda^*$ . Where  $\hat{\lambda}_d < \lambda^*$  dynamic efficiency would be ensured, and the relevant level of R&D cost would be the one that satisfies the condition:

$$\hat{\lambda}_d = \lambda_p^* \quad [33]$$

Substituting [25] and [30] into [33], and rearranging, we have:

$$R = \frac{\beta k^2}{2(1+\beta)^2} \quad [34]$$

Assuming  $\beta = 1$ , the level of R&D costs that makes the NPB-maximising CET dynamically inefficient is,

$$R > \frac{k^2}{8} \quad [35]$$

It is important to note here that for a payer to take into account the dynamically efficient threshold when setting the CET, we have to assume: the payer has some knowledge or can somehow infer the developers' reserve ICERs distribution and (ii) the payer knows that more innovation would be needed in the future and it depends on the current investment level in R&D. Only if these assumptions hold, the payer will have incentives to increase the CET above its health maximisation point to the dynamically efficient level (or higher depending how much innovation would like to see produced in the future).

### 3.3 Restrictions based on non-value-based criteria

Another potential way to alter the ICER pricing is when payers (or policy makers) impose non-value-based restrictions to the drug use. Such measures have important economic implications. Sometimes, expenditure ceilings with negotiated rebates, budget caps (or budget impact thresholds) or patient access restrictions, are put in place. These are used not only to facilitate access to treatments that are not cost-effective, but also to control the expenditure on treatments that are cost effective but have a large financial impact (or where there are issues with affordability). For example, when the government imposes austerity measures and needs to cut spending on new medicines, especially those that – while highly cost effective – are very expensive and/or used in a large population of patients.

Mechanisms to control expenditure and restrictions on patient numbers can be interpreted as an effective use of the bargaining power by the payer. Assume, for instance, that the policy maker fixes a budget limit or a pharmaceutical expenditure ceiling  $\Phi$  for technologies  $i = 1, \dots, I$ . If the imposed limit is lower than the budget or expenditure corresponding to the threshold level  $\lambda_e$  – and therefore also smaller than the amount corresponding to  $k$  – this would imply that if priced at  $p_i^e(\lambda_e)$ , the expenditure on new technologies  $i = 1, \dots, I$  will exceed the budget constraint. That is:

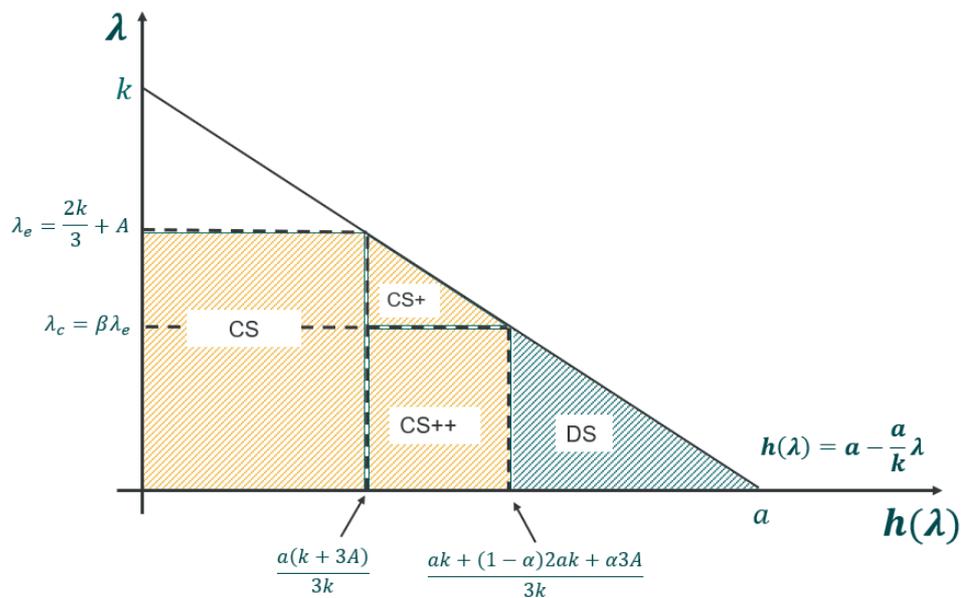
$$\sum_i p_i^e(\lambda_e) q_i > \Phi \quad [36]$$

In order to meet the new budgetary limit, the payer implicitly reduces the effective threshold until a level  $\lambda^c < \lambda_e$ , while maintaining  $\lambda_e$  as the threshold to explicitly declare new technologies cost effective. The value of  $\lambda^c$  is fixed so that following condition is met:

$$\Phi \geq \sum_i p_i^e(\lambda^c) q_i \quad [37]$$

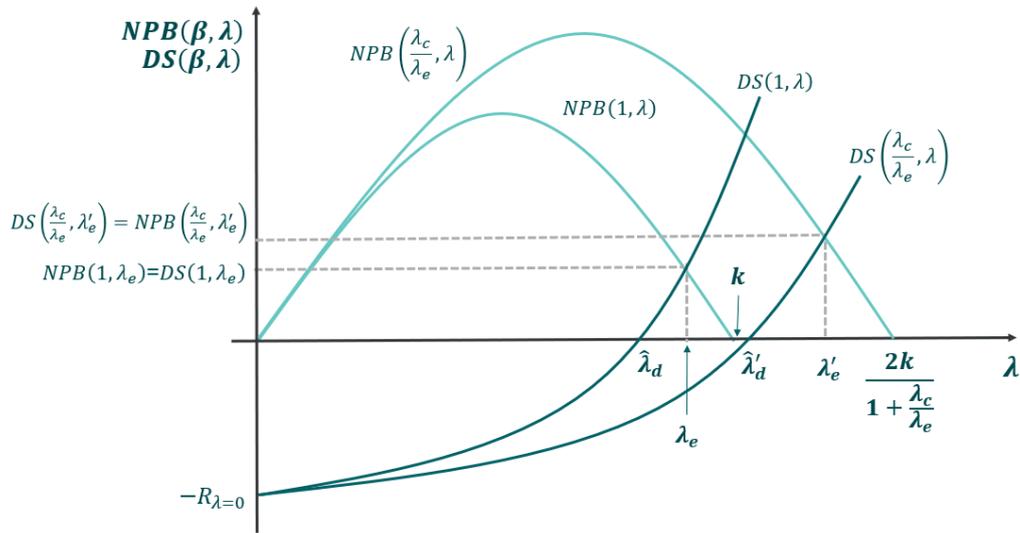
All medicines with reserve ICERs lower than (or equal to) the explicit CET  $\lambda_e$  are incorporated into the system. However, their final price is determined according to the implicit threshold  $\lambda^c$ . Thus, the payer can extract part of the developers' surplus using their monopsonistic bargaining power.

Figure 11 shows how budget constraints (or expenditure ceilings, or patient restrictions), allow the payer to extract surplus from the developer while allowing access to medicines with reserve ICERS above the level of  $\lambda^c$ .



**Figure 11: SURPLUS DISTRIBUTION WITH EFFECTIVE ICERS ALTERED BY POLICY**

Using these restrictions, the payer reduces the effective ICER to a proportion ( $\beta \in (0,1)$ ) of the CET. That is, the payer exploits their bargaining power of  $(1 - \beta)$ . Areas C+ and C++ in Figure 11 are extracted from the developers' surplus and transferred to the payer.



**Figure 12: CONSUMER AND PRODUCER SURPLUS WITH PAYER IMPOSED RESTRICTIONS**

We modelled the ability of the payer to impose restrictions on the use of new health technologies – or the money that can be spent on acquiring them – as an increase of their bargaining power. We assume that the increase in the payer bargaining power,  $(1 - \beta)$ , is proportional to the impact of the imposed budget restriction on the effective ICER, that is  $\beta = \frac{\lambda_c}{\lambda_e}$ . The increase in the payer’s bargaining power implies that a share  $(1 - \beta)$  of the developers’ surplus is transferred to the payer. This is represented by areas CS+ and CS++ in Figure 11. Figure 12 illustrates this transfer as a downward rotation of the DS from  $DS(1, \lambda)$  to  $DS\left(1, \frac{\lambda_c}{\lambda_e}\right)$ . The area between the two curves represents the surplus extracted by the payer, which is also represented by the upward shift of the CTC curve from  $NPB(1, \lambda)$  to  $NPB\left(1, \frac{\lambda_c}{\lambda_e}\right)$ .

An implication of the increased bargaining power of the payer is that their maximum WTP per unit of health becomes  $k' = \frac{2k}{1 + \frac{\lambda_c}{\lambda_e}}$ , which is unambiguously higher than  $k$ . This is because at the level  $k$ , the population still obtains a positive NPB via the price decrease resulting from the restrictive measure. Therefore, under certain circumstances, it may be efficient to increase the CET to a level above  $k$ . Figure 12 represents such a case where the threshold that equates the payer and developer surplus after the restriction ( $\lambda'_e$  in the figure), is higher than  $k$ .

Again, the social surplus-maximising result  $\lambda^* = k$  is not incentive compatible and therefore it does not guarantee the dynamic efficiency i.e. optimal investment in R&D for the future. Thus, the minimum CET required to ensure non-negative surplus to developers (for the example in Figure 12) should be fixed at  $\lambda = \hat{\lambda}'_d$ . After the policy-induced change in the effective ICER, developers require a higher CET to cover all cost of R&D, obtain a non-negative surplus, and send the right signal to investors to incentivise continued funding of new projects. This CET is higher than  $k$ , and at this level all net positive surplus is captured by the payer.

Finally, the maximum WTP of the health system for  $\beta = \frac{\lambda_c}{\lambda_e}$  is  $\lambda = \frac{2k}{1+\frac{\lambda_c}{\lambda_e}}$ . The CET could be increased up to this level without implying a net health loss for the system. At this level all surplus will be accrued by developers. A protection term using an optimal patent length is required to allow consumers or payers to extract part of the total social surplus generated by the new medicines after patent expiration (Danzon, Towse and Mestre-Ferrandiz, 2015; Danzon, Towse and Mulcahy, 2011).

This result implies that the payer can use the bargaining power as a monopsonist and policy decision maker to affect the distribution of the surplus in such a way that high-cost medicines that are nevertheless cost-effective can be incorporated into the system and provide higher NPV. This is because part of the developers' profit is transferred to consumers via restrictions or regulation. This transfer of economic value can also be interpreted as a developer-funded budget increase that is used to increase health. In response, the threshold level may need to be increased to prevent a part of the industry from obtaining a potential negative surplus and low return on R&D investment, which will in turn disincentivise new investments and lead to a suboptimal level of innovation in the future (dynamic inefficiency).

### 3.4 Competition in pharmaceutical markets for patented products

Another factor that has a positive influence on the payer's bargaining power is the existence of on-patent (or in-class) therapeutic competition. On-patent competition implies that the payer has other medicines that can be used as substitutive treatment options for patients with the same disease or health condition. Competitors fight for market share via prices or the use of non-price competition. The payer – as a single buyer monopsonist – can establish the nature of this competition. At times, free market interaction is allowed, but mechanisms such as tenders, price-volume agreements and/or reference pricing are often also used.

At stage 5, the final price of the new medicine is set through a bargaining process between the producer and the payer. The developer and the payer both know that there are substitutes for this new medicine. The payer uses the increased bargaining power due to existing market competition to agree on a final price that is below the CET price. The magnitude of the difference between the corresponding effective ICER and the CET depends on how the competition affects the bargaining power of each player.

At stage 4, the developer of the new medicine observes the CET and, if it is above their reserve ICER, sets an ICER equal (or lower) than the CET.

At stage 3, the payer fixes the CET following the nature's determination of  $k$  in stage 1 and once developers have invested in R&D in stage 2. Then, the payer – anticipating the developer's decision – can strategically set the CET to maximise their own share of the total surplus or, if needed, increase the CET up to the dynamically efficient CET. Alternatively, the payer may choose to split the surplus evenly, or proposing a CET that will distribute the surplus somewhere in between these two possible allocations. For either setting the CET at the dynamically efficient CET or the equal market surplus split, some form of payer's knowledge of developers' reserve ICERs distribution is assumed.

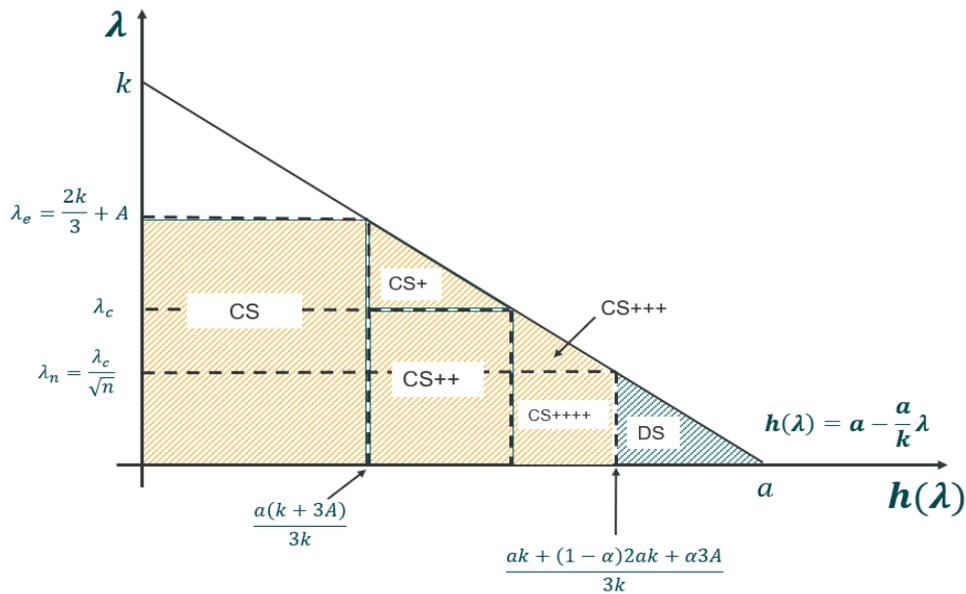
At stage 2, the developer – following the value of the maximum WTP set by the nature in stage 1 – decides whether or not to invest in R&D<sup>35</sup>.

We incorporate competition into our model as a change in bargaining power. The bargaining power of both players – defined as  $\beta$  – is now defined by the following function:

$$\beta = \frac{\lambda_c}{\lambda_e \sqrt{n}} \quad [38]$$

Where  $n$  is the number of competitors (other potentially substitute medicines),  $\lambda_c$  is the effective ICER when restrictions are imposed, and  $\lambda_e$  is the level of the CET that equates the surplus of both players<sup>36</sup>. With bargaining power now defined by [38], where the number of substitute medicines is two or more ( $n \geq 2$ ), the effective ICER is now even lower than  $\lambda_c$ . Figure 13 shows how – in addition to the effect of the policies discussed in the previous section – the incorporation of competition affects the distribution of the total surplus.

As a response to a reduced level of developer bargaining power and increased level of payer bargaining power, areas C+++ and C++++ are transferred to the payer in addition to C+ and C++ as. The level of net health gains is now higher<sup>37</sup> because the effective ICER is lower. As we have shown in the previous section, the increase of payer bargaining power leads to a share  $(1 - \beta)$  of the developer surplus is transferred to the payer.



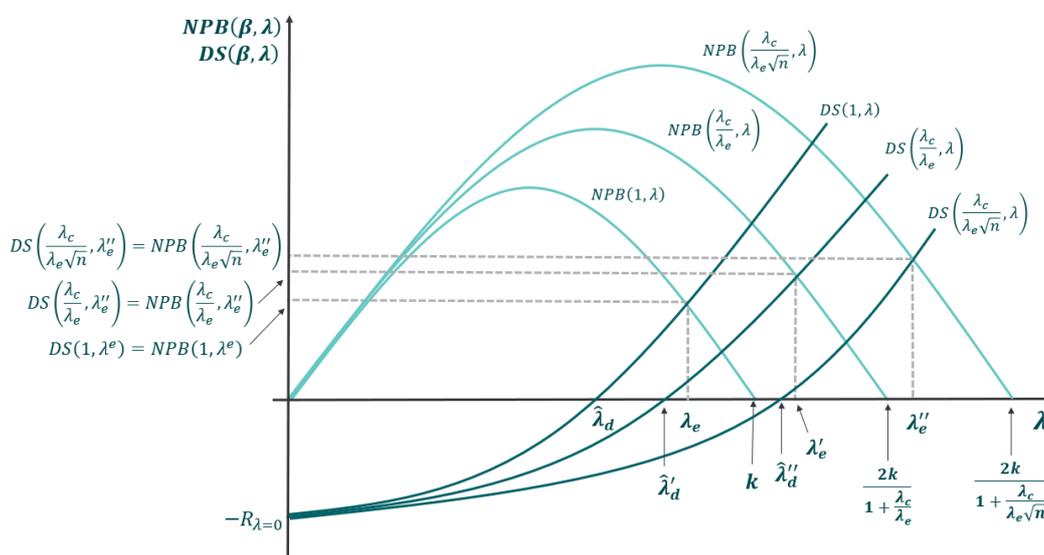
**Figure 13: SURPLUS DISTRIBUTION WITH EFFECTIVE ICERS ALTERED BY POLICY AND COMPETITION**

<sup>35</sup> An appropriate specification of the R&D decision and pharmaceutical market outcomes should be dynamic. The key reason is that R&D decisions are endogenous to market functioning and regulation, and vice-versa.

<sup>36</sup> With no restrictions,  $\lambda_c = \lambda_e$ , and no competition  $n = 1$  equation [38] represents the case of Pandey et al. (2018)

<sup>37</sup> It is the value  $\frac{ak+(1-\beta)2ak+\beta3A}{3k}$  in Figure 13.

Figure 14 shows – in addition to the previous case – the implications of on-patent competition on the distribution of the total surplus, the CET, the dynamic reserve ICER of developers, and the maximum WTP of the payer.



**Figure 14: SURPLUS DISTRIBUTION WITH EFFECTIVE ICERS ALTERED BY POLICY AND COMPETITION**

With competition, both the NPB and DS curves shift to the right. The payer can capture a larger share of the surplus because its bargaining power  $\left(1 - \frac{\lambda_c}{\lambda_e \sqrt{n}}\right)$  is now higher than it was previously  $\left(1 - \frac{\lambda_c}{\lambda_e}\right)$ . For a large enough impact of the market competition and pharmaceutical regulations in the distribution of the bargaining power, the CET might be set above the supply-side threshold  $k$  without causing the health foregone to exceed the health gained. This case is represented in Figure 14, where  $k < \hat{\lambda}_d''$  and therefore a CET set at  $\lambda = k$  is not incentive compatible.

The surplus a payer can extract from developers – through use of its bargaining power and lower final prices – can compensate (or even overcome) the excess price that a CET above  $k$  involves. It is also important to note that a CET higher than the supply-side threshold is still mutually beneficial in some circumstances – e.g. at certain levels of R&D cost and/or the number of competitors. However, if these conditions are met, the payer’s ability to extract part of the surplus from developers enables the payer to increase their own maximum WTP, as this extracted surplus (via lower prices at procurement stage) can be used as an extra source of funding for new medicines.

Increased bargaining power for the payer implies an increase of the developer’s dynamic reserve ICER from  $\hat{\lambda}_d$  to  $\hat{\lambda}_d''$ , as shown by the rotation of the DS curve (Figure 14). The CET that equates the surplus of the two players increases from  $\lambda_e$  to  $\lambda_e''$ .

If bargaining power is not taken into consideration, the payer will determine the CET level following the  $NPB(1, \lambda)$  and  $DS(1, \lambda)$  curves in Figure 14. Even if the payer decides to fix the CET in a way that splits the surplus evenly (i.e., at  $\lambda_e$ ), the outcome can still be

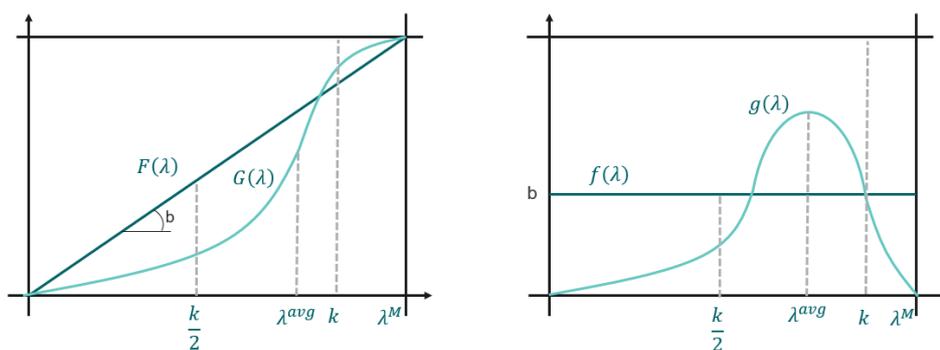
profitable for the payer but lead to dynamic inefficiency ( $\hat{\lambda}_d'' > \lambda_e$ ) as shown by the curves  $DS\left(\frac{\lambda_c}{\lambda_e\sqrt{n}}, \lambda\right)$  and  $NPB\left(\frac{\lambda_c}{\lambda_e\sqrt{n}}, \lambda\right)$ .

Finally, it is important to note that in a perfectly atomised supply, the bargaining power of developers approaches zero. The payer would fix the CET (and price) at a level equal to the marginal cost of production of the marginal developer and the condition for the dynamic efficiency would not meet. While such a situation is possible, it not very likely to happen in the market for new patented medicines where atomised markets are not plausible to be observed.

### 3.5 Distribution of reserve ICERs and the value of health displaced

Taking the assumption that reserve ICERs are non-uniformly distributed within the range  $[0, k]$  may have interesting implications on our results. We may assume that the best existing technology for developing new drugs is common knowledge, and thus accessible for all developers. Alternatively, we may assume that the scientific challenge is therapy area-specific rather than firm-specific, thus the reserve ICERs of all developers concentrate around certain values of the threshold (specially within therapy areas). This may also be the case for same-class technologies developed within a similar time frame.

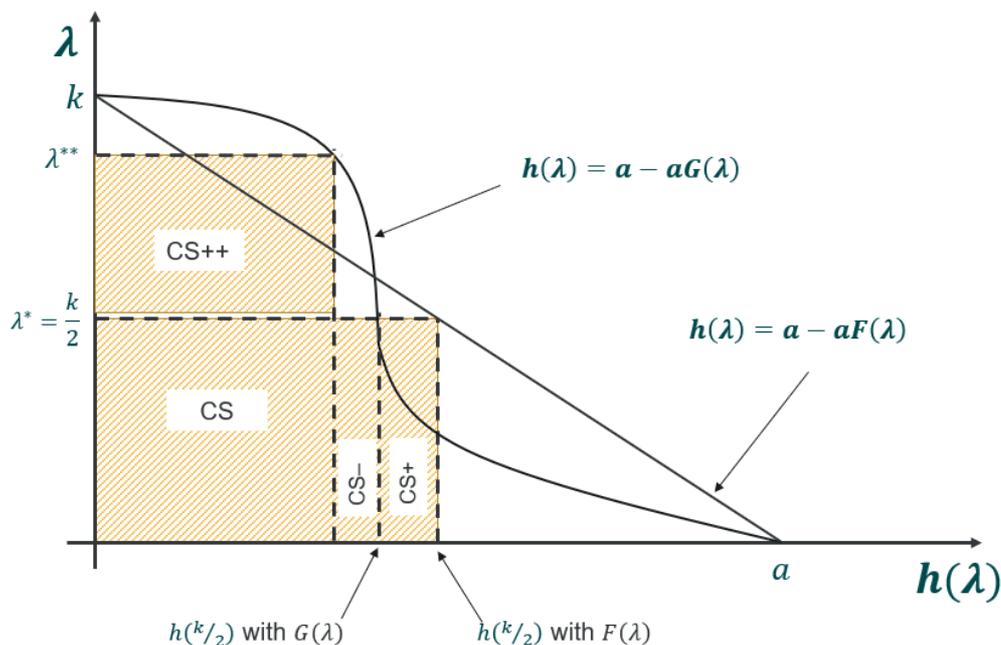
Let us assume the reserve ICERs follow a distribution function with mean  $\mu = \lambda^{avg}$  ( $\lambda^{avg} > \lambda^*$ ), with an associated Gaussian density function. Under this assumption, the CTC would be biased to the right. The graph on the left-hand side of Figure 15 compares the uniform distribution with the theoretical normal distribution. The right-hand side graph compares the density functions of both distributions.



**Figure 15: DISTRIBUTION AND DENSITY FUNCTIONS OF DEVELOPERS RESERVE ICERS**

The distribution function  $G(\lambda)$  represents the case where most developers concentrate around a given value of reserve ICER equal to  $\lambda^{avg}$ . For threshold values that are low, the number of potential entrants is also low. New entrants increase as the threshold increases. As the value of the threshold approaches  $\lambda^{avg}$ , the number of new entrants as a result of an increase in the threshold of the same magnitude increases rapidly. This means that the additional health gain obtained from the market entry of new developers in response to of the rise in the threshold increases more rapidly than the health foregone that results from

the additional cost of all medicines being priced at a higher threshold. Thus, the threshold level that maximises NPB will increase. Figure 16 shows how the  $h(\lambda)$  function changes as a response to the change in the distribution. This can be interpreted as a change in the elasticity of  $h(\lambda)$ , which implies that the maximisation of the payer surplus occurs for a different value of the threshold, The example below shows the case where this threshold is at a higher value.



**Figure 16: DISTRIBUTION AND DENSITY FUNCTIONS OF DEVELOPERS RESERVE ICERS**

With the reserve ICERs following a distribution function  $G(\lambda)$ , the shape of the  $h(\lambda)$  function is represented in Figure 16. Here, the value that maximises the surplus of the consumer is  $\lambda^{**}$ . With the uniform distribution and  $\lambda^*$ , the payer (consumer) surplus is the sum of the three areas  $CS$ ,  $CS^-$  and  $CS^+$ . At the threshold level as the non-uniform distribution, the payer (consumer) surplus is the sum of the two areas  $CS$  and  $CS^-$ . Therefore, the payer could increase the total surplus by increasing the threshold up to  $\lambda^{**}$ , the point where  $CS^-$  is lost, but  $CS^{++}$  is gained.

With the distribution of reserve ICERs defined by  $G(\lambda)$ , most of the relevant technologies produce health benefits at a higher threshold level. As a result, the CET that maximises the payer's surplus increases up to  $\lambda^{**}$ . Therefore,  $\lambda^{**}$  becomes the value of the threshold where the supply effect is equal to the demand effect.

The new shape of the net health gain function  $h(\lambda)$  also has implications for both the NPB and the DS functions. Figure 17 shows how these new functions compare to the ones characterised under the assumption of a payer with bargaining power  $(1 - \frac{\lambda_c}{\lambda_e \sqrt{n}})$ .

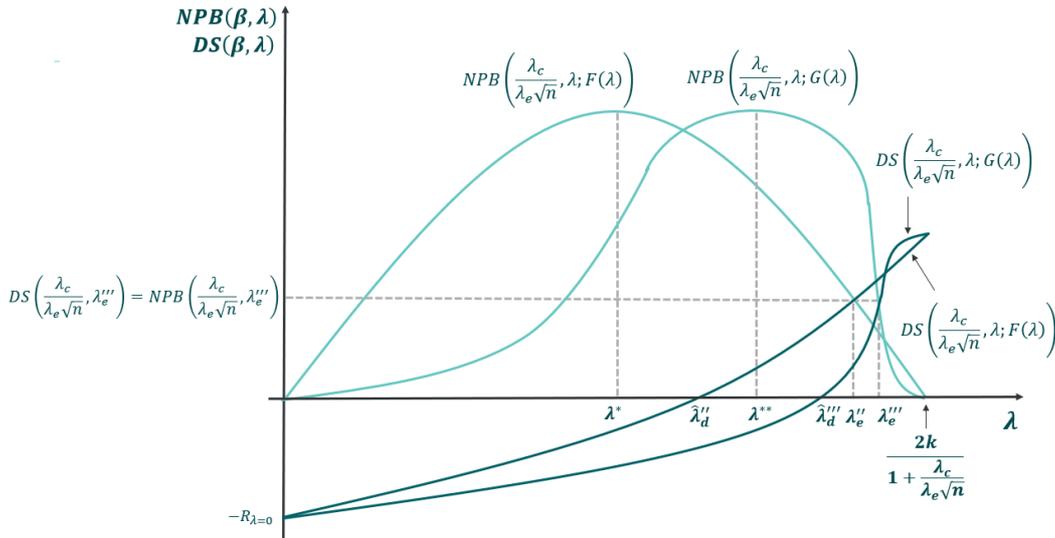
With the reserve ICERs of most of the developers concentrated around  $\lambda^{avg}$ , the NPB curve shifts to the right, and becomes biased around the threshold value  $\lambda^{**}$ , where NPB reaches its maximum. It then decreases rapidly with values of the threshold higher than  $\lambda^{**}$ . This decreasing section of the NPB is the part of the curve where the demand effect dominates. The shape of the DS curve is also changed. It increases slowly for lower values of  $\lambda$ , and increases rapidly when it approaches  $\lambda^{**}$ . The dynamic reserve ICER  $\hat{\lambda}_d'''$  is now at a higher level. A first implication of the example drawn in Figure 17 is that  $\hat{\lambda}_d''' > \lambda^{**}$ , which means the log-run dynamic efficiency is not achieved at  $\lambda^{**}$ . Therefore, a payer who seeks to incentivise an optimal amount of innovation in the long run will need to increase the CET to a level equal or higher than  $\hat{\lambda}_d'''$ .

At the new CET, values of the threshold that are now incentive-compatible are in the range

$$\left[ \hat{\lambda}_d''', \frac{2k}{1 + \frac{\lambda_c}{\lambda_e \sqrt{n}}} \right].$$

A narrower interval than that corresponding to the uniform distribution

$$\left[ \hat{\lambda}_d'', \frac{2k}{1 + \frac{\lambda_c}{\lambda_e \sqrt{n}}} \right].$$



**Figure 17: SURPLUS DISTRIBUTION NON-UNIFORM RESERVE ICERS DISTRIBUTION AND PAYER BARGAINING POWER**

With the distribution  $G(\lambda)$ , the CET that equates the surplus of both players or is  $\lambda_e'''$ , is higher than the CET that equated the surplus of both players with uniform distribution or  $\lambda_e''$ . For some specific values of  $n$  and  $\lambda_c$ , this CET is also higher than  $k^{38}$ .

### 3.6 Non-binding flexible health budgets in the long run

The assumption of fixed budgets can be reasonable for the short run. However, systems and policy makers may increase the health budget over the medium and long run. In many markets, the budgets are increased every year (Cubí-Molla et al., 2020). In some cases, the

<sup>38</sup>  $\lambda_e''' > k$  if and only if  $\lambda_c > \lambda_e \sqrt{n}$  which implies that  $n < \frac{\lambda_c^2}{\lambda_e^2}$ .

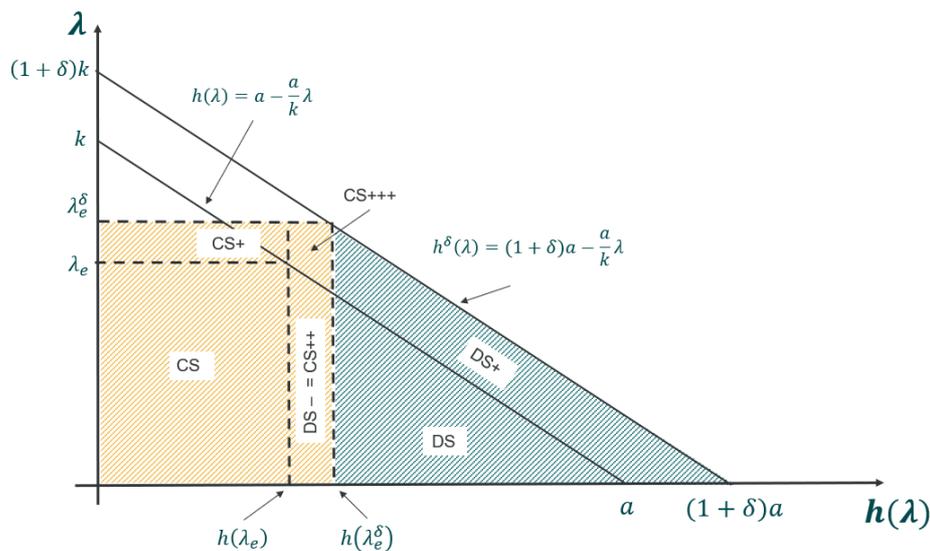
budget may even be increased in the short run to deal with a public health priority or emergency, or if a high-cost product that is also highly effective is approved for the use by regulatory authorities in large treatment populations (e.g., in the cases of Portugal and Italy adopting direct-acting antivirals for the hepatitis-C as noted in Berdud et al. (2018)).

An increase in the budget shifts the  $h(\lambda)$  function to the right. The health foregone in exchange of the health added by the additional budget is zero. This implies that: (i) total net health gain increases at every level of the threshold, and (ii) the value of the supply-side threshold – or the health system’s maximum WTP – increases. In other words, the medium and long-term maximum WTP of the health system increases by a positive factor proportional to the budget increase. In a dynamic setting, this means that some cost-effective medicines adopted in the current period at a given ICER below the CET will produce a larger surplus in the medium and long run when the CET is augmented. Alternatively, some medicines with ICERs slightly above the CET which were previously rejected for not being cost effective could be considered cost effective if a medium or long-run perspective is adopted.

Let us assume that the accumulated medium-term growth of the health budget – appropriately discounted to the present – is  $\delta$  (with  $\delta \in (0, 1)$ ). Then, during the present period, for a new medicine that is being appraised with consideration for the impact on the health system in the medium-term, the payer should account for this increase in the budget by increasing their maximum WTP by the same proportion. The implication for the medium-term budget growth is a change in the function  $h(\lambda)$ , now represented as:

$$h^\delta(\lambda) = \left( (1 + \delta)a - \frac{a\lambda}{\kappa} \right) \quad [39]$$

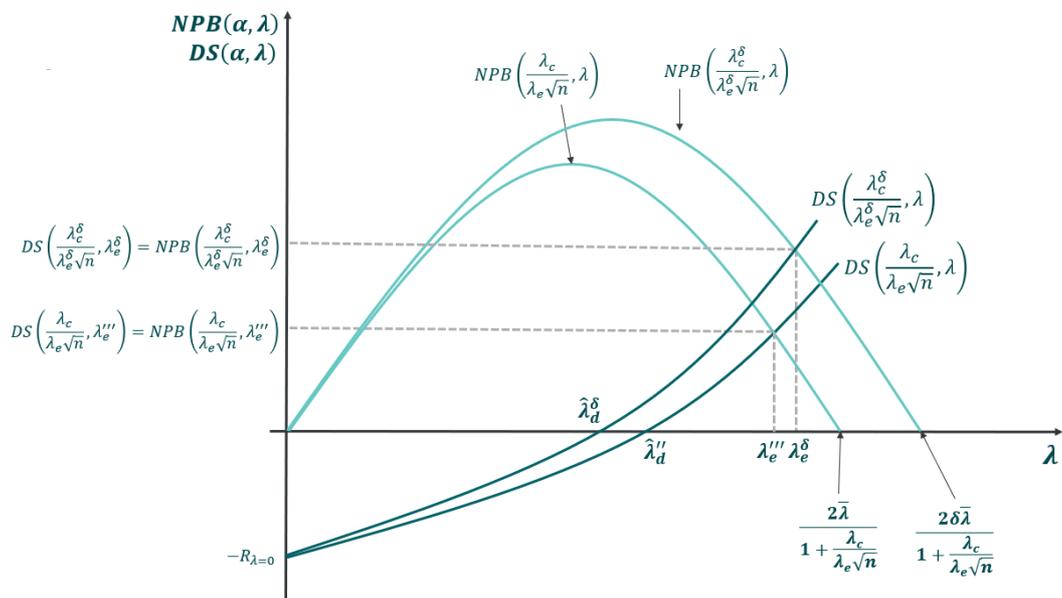
Figure 18 shows how the function  $h(\lambda)$  shifts to the right as a result of the increase in the health budget. The value of the supply-side CET (the maximum WTP of the payer) increases from  $k$  to  $(1 + \delta)k$ , while the maximum NPB for the payer (when  $\lambda = 0$ ) increases to  $(1 + \delta)a$ . The value of the threshold that equates the surplus of both players before the budget increase  $\lambda_e$ , increases to  $\lambda_e^\delta$  as a response to the shift in  $h(\lambda)$ . The surplus of both players, the payer and developers, increase at  $\lambda_e^\delta$  when compared to the surplus levels corresponding to threshold value of  $\lambda_e$  (with no budget increase).



**Figure 18: THE  $h(\lambda)$  FUNCTION AND SURPLUS DISTRIBUTION WITH FLEXIBLE BUDGETS**

At the new threshold level  $\lambda_e^\delta$ , the consumer surplus increases by CS+ and CS++, while the developers surplus increases by DS+ but decreases by DS-. Figure 18 shows how the incorporation of a flexible budget in the medium term affects the CTC and DS curves, the consumer and producer' surplus, and the threshold level.

An increase of a given proportion in the budget implies that the health system's maximum WTP increases by the same proportion. In our example, the new supply-side CET – with a flexible budget – is  $(1 + \delta)k$  involves that for all other scenarios discussed in previous sections of this paper (i.e., the R&D cost, policy and regulation, competition, distribution of reserve ICERs), all socially optimal and/or surplus equating threshold levels discussed increase in the same proportion. It is important to note that flexible budgets potentially increase the total surplus and can therefore make the two players to be better off in the final solution as Figure 19 shows. This is an essential difference from the scenario involving a change in the bargaining power, where the gains obtained by the payer come at expense of the developers.



**Figure 19: SURPLUS DISTRIBUTION AND THRESHOLD LEVELS WITH FLEXIBLE BUDGETS**

## 4 Concluding remarks

In this paper, we proposed a general theoretical model of CET and ICER pricing. We used Pandey et al., (2018) to set the baseline for our generalisation. We incorporated a staged decision sequence for the players (i.e., payers and developers), and added a formal structure to incorporate four key elements into the baseline model. These are:

- Incorporating sunk R&D costs in the characterisation of supply;
- Incorporating the NBS as a way to fix the net price of new medicines and the final effective ICER through a bargaining process and bargaining power distribution between the payer and developers;
- Assuming a non-uniform distribution of reserve ICERs of the developers; and
- Allowing flexible health budgets in the medium-term.

These four elements have economic implications that affect the level of the dynamic reserve ICER of the developers, the level of the optimal CET, the distribution of the total surplus between consumers and producers, the maximum WTP of the payer, and final net price and its corresponding effective ICER.

The incorporation of sunk R&D costs shifts the developers' surplus curve downwards. On this new curve, developers receive negative surplus (or negative returns) for a range of positive threshold values. Developers will still supply already-developed new medicines at a loss because the size of this loss is smaller than the sunk R&D costs. This market allocation can be considered *statically efficient* as it maximises access to medical innovation in the short term. However, it is not *dynamically efficient* as the signal of negative returns will disincentivise investors, who will respond by reducing the amount of resource devoted to pharmaceutical R&D. If the level of R&D is reduced, the amount of future pharmaceutical innovation will end up being suboptimal and thus inefficient. Our analysis shows that the CET should be set – at a minimum – at the level of what we have identified to be the *dynamic reserve ICER*, which is placed at a positive value higher than the short-run reserve ICER.

An implication of the dynamic reserve ICER is that the set of CETs that maximise short-run health gains subject to this dynamic efficiency condition will be in a narrow interval close to the maximum WTP of the payer. In such cases, some short-term population benefit should be traded off via a higher CET to get more long-term population benefit from an optimal level of long-term pharmaceutical R&D investment. This implication is even more relevant for multi-country settings, where the decision maker of each country has an incentive to free-ride on R&D and innovation by setting CETs below the long term optimum for their population in one of two situations: if the R&D investment, necessary to market a new medicine, has already been carried out; or if other countries commit to set the CET equal or above such dynamic reserve ICER. If many countries set the CET strategically in the short-run to free-ride, the final outcome will be CETs set below the dynamic reserve ICER and the developers' supplying at a loss when sunk costs are taken into account. However, in a repeat game, developers will stop investing in new medicines for the future.

Thus, the amount of new medicines developed would be too low in the long-run, affecting negatively overall long-run health gains.

The incorporation of NBS and bargaining power enables the analysis of the use of non-value-based policies (such as budget caps, price-volume agreements, clawbacks, rebates and other forms of discounts), or the impact of in-class competition between new medicines still on patent. This is modelled as higher bargaining power on the part of the (monopsonist) payer, and the principal effect is that a share of the developers surplus is extracted by the payer at the procurement and contracting stage, which follows the price and reimbursement recommendation stage. This transfer of an economic benefit between the two players happens via lower effective ICERs and prices. The main implication is that a payer with higher bargaining power has the ability to negotiate a lower final price per QALY gained than the price that would correspond to the CET. This means that the final short-term net population benefit is higher. Thus, the maximum WTP – or the maximum acceptable level of the CET prior to this exercise of bargaining power – increases above the supply-side threshold. In other words, consumers still obtain a net benefit when setting a threshold above the opportunity cost of the health system. At the same time, for any value of the threshold, the surplus captured by the developer falls by the same proportion of the payer's increase in bargaining power. The main implication is that the dynamic reserve ICER of developers also increases in expectation of this use of bargaining power, thereby reducing the set of dynamically efficient CETs. The incorporation of the bargaining process and the bargaining power distribution then has a twofold implication. First, if CETs are not adjusted to reflect bargaining power, there will be medicines that are denied access because they are considered not cost effective, even though they still produce net health gains after their net prices are agreed, and second, developers whose reserve ICERs are close to the level of the notional (pre-bargaining) CET will face potentially negative returns on their investment on R&D after the negotiation of the net price (effective ICER).

Assuming that the developers' reserve ICERs concentrate around a given narrow range of the threshold level, the distribution of new entrants to the market is biased around this range. This means that the payer and developers obtain most of the health benefits and economic value when the value of the CET is set within this interval. When the CET is set below this range, fewer developers of new medicines will supply their technologies, which means society does not benefit from the full potential of medical innovation. Above this range, the total surplus is maximised when the CET is set where there is zero net population benefit – the health system's opportunity cost – or even above the supply-side CET when the payer has sufficient bargaining power. But fixing the CET at the maximum WTP of the health system means that all the surplus is obtained by the industry. Even when the CET is fixed at a level where the surplus of the payer and the developers are equal, such CET is higher than the CET that equates both players' surplus with uniform distribution of the reserve ICERs. The dynamic reserve ICER also increases. For the payer, the range of possible values of the CET is a narrower interval situated closer to the maximum WTP. Together with the effect of bargaining power and R&D costs, and under certain conditions, a CET above the supply-side threshold would make sense to efficiently allocate resources in the long term.

Finally, our model explored the impact of flexible budgets. We incorporated this as a shift in the function measuring the net health gain. This interpretation rests on the fact that the extra budget does not finance any health intervention and it can therefore be used to finance new treatments without displacing any current interventions. This is equivalent to an increase in the health system's opportunity cost, while the additional budget remains

unused the same amount of health is produced at higher cost. The increase of the health budget in the medium-term means also increases funding for medicines assessed in the present as a share of their cost will be borne by the system in the future. Along with previously discussed changes, the incorporation of flexible budgets also increases the threshold level that splits the surplus equally between both sides of the market. However, unlike the three previous changes, this reduces the dynamic reserve ICER of the developer. This means that the range of CETs that are incentive-compatible for both payer and developers is wider, and the lower bound is lower.

A critical omission from the work in this paper is the incorporation of potential changes in the value of the new medicine during the medicine's life cycle. This requires dynamic modelling over at least two periods: pre- and post- patent expiry. The life-cycle value evolution has been shown to have a significant impact on the size of the surplus generated, and the distribution of this surplus between players (Berdud et al., 2019). Factors such as generic / biosimilar competition, development of new forms or improved presentations of the same medicine, its approval for new indications, or the entry of new in-class competitors in the future may reduce the future ICER of a medicine under review in the current period. This implies that the optimal CET for the payer should be set in the context of a two-period model. The question of what proportion of this long-term surplus generation should be given to each party – the elasticity of the intertemporal value and surplus trade-off – is an outstanding question. The proposed dynamic analysis is an area of development from the present work and will be considered for future research.

The study of the optimal threshold level from the social welfare perspective also requires a dynamic approach. In the static approach presented in this paper, we establish that the socially optimal threshold is the maximum WTP of the payer – or the health system opportunity cost. However, this threshold level gives all the surplus to the industry. In their work, Danzon and colleagues (Danzon et al., 2015; Danzon et al., 2011) show that this is the socially optimal result under an optimal patent length, after which consumers are able to capture some of the surplus through generic / biosimilar competition. The dynamic modelling is therefore a potential future research area that may clarify how the patent length can be determined to generate a fair (or reasonable) share of the total surplus generated by innovative medicines.

Overall, the results depend on the marginal productivity (or maximum WTP) of the health system, as well as on the payoff function of the payer (NPB) and the payoff function of developers (DS). However, as this analysis shows, these functions can be significantly affected by bargaining power on the payer's side, the size of R&D costs, the distributions of the reserve ICERs, the presence of flexible health care budgets in the medium-term, and other dynamic considerations affecting the life-cycle value of new medicines. Generalizing the model of Pandey et al., (2018) by incorporating these effects shows that the maximum WTP of the payer (consumers) can overcome the health system's opportunity cost under some circumstances, and also moves the set of possible incentive-compatible CETs closer to such maximum WTP. From the results presented in this paper, we can conclude that it is not evident that a certain fixed and explicit level of the CET is optimal for society. Adjustable levels of the CET that account for changes discussed here must be taken into consideration by decision makers.

## References

- Baker, R., Chilton, S., Donaldson, C., Jones-Lee, M., Lancsar, E., Mason, H., Metcalf, H., Pennington, M. and Wildman, J., 2011. Searchers vs surveyors in estimating the monetary value of a QALY: resolving a nasty dilemma for NICE. *Health Economics, Policy and Law*, 6(4), pp.435–447.
- Berdud, M., Garau, M., Neri, M., O'Neill, P., Sampson, C. and Towse, A., 2018. *R&D, Competition and Diffusion of Innovation in the EU: The Case of Hepatitis C*.
- Berdud, M., Wallin-Bernhardsson, N., Zamora, B., Lindgren, P. and Towse, A., 2019. *The Case of Risperidone: Assessing the Life-cycle Value of Second-Generation Antipsychotics in Sweden and the UK*.
- Binmore, K., Rubinstein, A. and Wolinsky, A., 1986. The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, pp.176–188.
- Brouwer, W., van Baal, P., van Exel, J. and Versteegh, M., 2019. *When is it too expensive? Cost-effectiveness thresholds and health care decision-making*. Springer.
- Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P.C. and Sculpher, M., 2015. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health technology assessment (Winchester, England)*, 19(14), p.1.
- Cubi-Molla, P., Errea, M., Zhang, K. and Garau, Ma., 2020. *Are cost-effectiveness thresholds fit for purpose for real-world decision making?* [OHE Consulting Report] Office of Health Economics. Available at: <https://www.ohe.org/publications/are-cost-effectiveness-thresholds-fit-purpose-real-world-decision-making> .
- 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), pp.294–301.
- Danzon, P.M., Towse, A. and Mulcahy, A.W., 2011. Setting cost-effectiveness thresholds as a means to achieve appropriate drug prices in rich and poor countries. *Health Affairs*, 30(8), pp.1529–1538.
- Dubois, P., Gandhi, A. and Vasserman, S., 2019. *Bargaining and international reference pricing in the pharmaceutical industry*. Working paper, Harvard University.
- Garcia Mariñoso, B., Jelovac, I. and Olivella, P., 2011. External referencing and pharmaceutical price negotiation. *Health Economics*, 20(6), pp.737–756.
- Garrison, L.P., Jackson, T., Paul, D. and Kenston, M., 2019. Value-based pricing for emerging gene therapies: the economic case for a higher cost-effectiveness threshold. *Journal of managed care & specialty pharmacy*, 25(7), pp.793–799.
- Grennan, M., 2013. Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review*, 103(1), pp.145–77.

Jelovac, I., 2015. On the relationship between the negotiated price of pharmaceuticals and the patients' co-payment. *Economics Bull 2015*, 35, pp.481–93.

Lakdawalla, D.N., Doshi, J.A., Garrison Jr, L.P., Phelps, C.E., Basu, A. and Danzon, P.M., 2018. Defining elements of value in health care—a health economics approach: an ISPOR Special Task Force report [3]. *Value in Health*, 21(2), pp.131–139.

Lomas, J., Martin, S. and Claxton, K., 2019. Estimating the Marginal Productivity of the English National Health Service From 2003 to 2012. *Value in Health*.

Mason, H., Jones-Lee, M. and Donaldson, C., 2009. Modelling the monetary value of a QALY: a new approach based on UK data. *Health economics*, 18(8), pp.933–950.

Pandey, H., Paulden, M and McCabe, C., 2018. *Theoretical models of the cost-effectiveness threshold, value assessment, and health care system sustainability*. Edmonton (AB): Institute of Health Economics.

Sculpher, M., Claxton, K. and Pearson, S.D., 2017. Developing a value framework: the need to reflect the opportunity costs of funding decisions. *Value in Health*, 20(2), pp.234–239.

Smith, R.D. and Richardson, J., 2005. Can we estimate the 'social' value of a QALY?: Four core issues to resolve. *Health Policy*, 74(1), pp.77–84.

Thokala, P., Ochalek, J., Leech, A.A. and Tong, T., 2018. Cost-effectiveness thresholds: the past, the present and the future. *Pharmacoeconomics*, 36(5), pp.509–522.

Vallejo-Torres, L., García-Lorenzo, B., Castilla, I., Valcárcel-Nazco, C., García-Pérez, L., Linertová, R., Polentinos-Castro, E. and Serrano-Aguilar, P., 2016. On the estimation of the cost-effectiveness threshold: why, what, how? *Value in Health*, 19(5), pp.558–566.

# Appendix 1

Nash bargaining approach requires to define the Nash product between the maximum achievable market interaction payoff of each of the players involved. We depart from the payoff of the producer which is defined by the benefit obtained in case that an agreement is reached net of the outside option, or the payoff obtained in case that the agreement is not reached. The benefit function can be written as per equation [A1]:

$$\Pi(p_i) = p_i q_i - C_i(q_i) \quad [A1]$$

Where the cost  $C_i(q_i)$ , is given by equation [A2],

$$C_i(q_i) = R_i + K_i + c q_i \quad [A2]$$

Assuming that fixed costs are sunk in the short term, the outside option for the developer would be given by equation [A3],

$$\Theta^d = -R_i - K_i \quad [A3]$$

Then the payoff function (the bargaining function) is then written by combining equations [A1-A3],

$$\Pi(p_i) - \Theta^d = p_i q_i - c q_i \quad [A4]$$

For a rational player to accept a possible agreement requires of obtaining a non-negative payoff, which using equation [A4] means that,

$$p_i q_i \geq c q_i \quad [A5]$$

From equation [A5] we can determine the minimum acceptable price for the developer in the bargaining, which is equal to the marginal cost of production,

$$\underline{p} = c \quad [A6]$$

Let's assume that the new technology  $i$  has the same health care cost than the existing treatment  $j$  that is going to be replaced. Thus, we can define the ICER of a new technology brought to the market by a developer as follows,

$$\lambda_i = \frac{p_i - p_j}{h_i - h_j} \quad [A7]$$

Then we can write the price of technology  $i$  as a function of the ICER,

$$p_i(\lambda_i) = \lambda_i (h_i - h_j) + p_j \quad [A8]$$

Substituting the minimum acceptable price  $\underline{p}$  into [A6] we can formulate the minimum acceptable ICER for the developer,

$$\underline{\lambda} = \frac{\underline{p} - p_j}{h_i - h_j} \quad [A9]$$

And rearranging we can write,

$$\underline{p}(\underline{\lambda}) = \underline{\lambda} (h_i - h_j) + p_j \quad [A10]$$

Now, combining equations [A4], [A6], [A8] and [A10] we have that,

$$\Pi(p_i(\lambda_i)) - \Theta^d = (\lambda_i (h_i - h_j) + p_j) q_i - (\underline{\lambda} (h_i - h_j) + p_j) q_i \quad [A11]$$

Rearranging [A11] we finally obtain,

$$\Pi(p_i(\lambda_i)) - \Theta^d = q_i(h_i - h_j)(\lambda_i - \underline{\lambda}) \quad [A12]$$

We continue now to derive the payoff function of the payer. Let the payoff health benefit be the difference between the maximum WTP of the payer (we assume it is the CET) and how much is finally paid for the new medicine. Such a function can be written as follows,

$$B_i(p_i) = kh_iq_i - (p_iq_i + T_i(q_i)) \quad [A13]$$

Where  $k$  is the system's maximum WTP – in our model this is the marginal cost of a unit of health (supply-side definition). Additionally,  $T_i(q_i)$  represents the direct health care cost associated to the use of the new medicine  $i$ .

Let define now the outside option for the payer, or the value of the health benefit provided by the existing alternative treatments for the same patient population. It is expressed by the following equation,

$$\Theta^p = B_j(p_j) = kh_jq_j - (p_jq_j + T_j(q_j)) \quad [A14]$$

Assuming that health care associated costs are equal for the two medicines  $T_i(q_i) = T_j(q_j)$ , we can now write the payer bargaining payoff function as,

$$B_i(p_i) - \Theta^p = q_ik(h_i - h_j) - (p_i - p_j) \quad [A15]$$

Substituting [A8] into [A15] and rearranging, we have that,

$$B_i(p_i(\lambda_i)) - \Theta^p = q_i(h_i - h_j)(k - \lambda_i) \quad [A16]$$

Equation [A16] measures the payer's net surplus from introducing the new technology at a maximum WTP of  $k$ . However, as defined in the timing of this game, the payer can set an explicit CET to inform the funding and reimbursement decisions at a level below  $k$ . That level is  $\bar{\lambda}$ , and implies there is no market for ICER values such that  $\lambda_i > \bar{\lambda}$  as these technologies are not introduced and there is no bargaining stage for them. Then, the actual bargaining set is defined by the range  $\lambda_i \in (\bar{\lambda}, \underline{\lambda})$  hence equation [A16] can be rewritten as follows for the application of the NBS:

$$B'_i(p_i(\lambda_i)) - \Theta^p = q_i(h_i - h_j)(\bar{\lambda} - \lambda_i) \quad [A17]$$

It is important to note that when  $\bar{\lambda} = k$  equations [A16] and [A17] are equivalent.

We now define the Nash bargaining product as,

$$\text{argMax}_{\lambda_i} (\Pi(p_i(\lambda_i)) - \Theta^d)^\beta (B'_i(p_i(\lambda_i)) - \Theta^p)^{1-\beta} \quad [A18]$$

Where  $\beta \in [0,1]$  reflects the bargaining power of the developer. Then, substituting equations [A12] and [A17] into [A18] we can write,

$$\text{argMax}_{\lambda_i} (q_i(h_i - h_j)(\lambda_i - \underline{\lambda}))^\beta (q_i(h_i - h_j)(\bar{\lambda} - \lambda_i))^{1-\beta} \quad [A19]$$

Rearranging we have that,

$$\text{argMax}_{\lambda_i} q_i(h_i - h_j)(\lambda_i - \underline{\lambda})^\beta (\bar{\lambda} - \lambda_i)^{1-\beta} \quad [A19]$$

Assuming that  $q_i$ ,  $h_i$  and  $h_j$  are given and constant, the problem stated at [A19] is equivalent to the one below,

$$\text{argMax}_{\lambda_i} (\lambda_i - \underline{\lambda})^\beta (\bar{\lambda} - \lambda_i)^{1-\beta}.$$

## Appendix 2

Table A.2.1 shows definitions of notation (lambdas) used in the paper and definitions for the relevant notation related to the lambdas.

**Table A.2.1: DEFINITION OF KEY NOTATION AND SYMBOLS**

Symbol	Definition	Related symbols
$\lambda$	Any value that the cost-effectiveness threshold can take measured as Incremental Cost-Effectiveness Ratio (ICER)	
$\lambda^M$	The maximum value of the developers' reserve ICER. The value at which all potential developers make non-negative profits by launching the new medicine to the market	
$\underline{\lambda}^m$	The minimum value of the developers' reserve ICER. The value at which the first developer(s) entering the market make non-negative profits by launching the new medicine	
$\bar{\lambda}$	The maximum WTP of the payers. In our model this value is a choice variable for the payer who can choose for it a lower value than the supply-side threshold	$k$ : the supply-side threshold or the cost of opportunity of the health system
$\underline{\lambda}$	The static short-term reserve ICER of the developers. It is the minimum acceptable ICER of each developer (or the minimum price at which a developer is willing to sell the new medicine). It does not guarantee that developers cover the R&D investment necessary to develop the new medicine	$\underline{p}$ : the minimum price at which a developer will be willing to sell the new medicine
$\hat{\lambda}$	The dynamic long-term reserve ICER of the developers. It is the minimum acceptable ICER of each developer that ensures all R&D investment is covered	$\hat{p}$ : the minimum price at which a developer will be willing to sell the new medicine and will cover all the R&D investment
$\lambda_i^*$	The individual ICER that solves the Nash bargaining problem between the developer and the payer	$p_i^*$ : the price level corresponding to the individual ICER solving the Nash bargaining solution
$\lambda_p^*$	The CET value that that maximises the payer's surplus of equivalently, as defined in this work and in Pandey et al. (2018) the net population benefit	

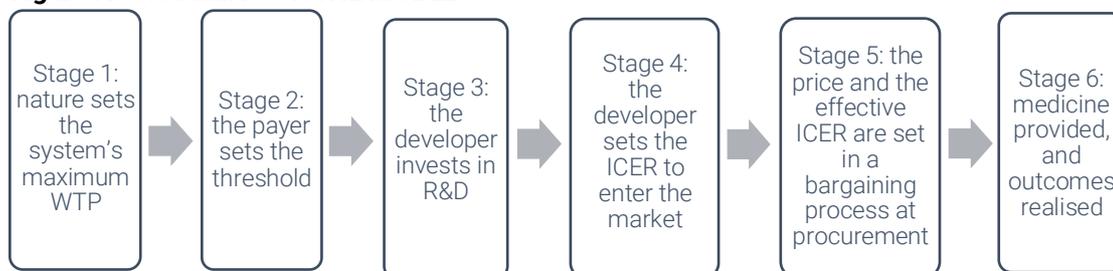
Symbol	Definition	Related symbols
$\hat{\lambda}_D$	The minimum CET value that guarantees a sufficient price to cover the R&D expenditure. We call this level of the threshold, the dynamic efficient CET or the dynamic reserve ICER of the developer.	
$\lambda^*$	The value of the CET that maximises the social welfare. Results show that the social welfare, defined as the sum of the payer's and developers' surplus, is maximised when all the surplus is given to the developers. This is exactly when $\lambda^* = k$ .	$k$ : the supply-side threshold or the cost of opportunity of the health system
$\lambda_e$	The level of the CET that equates both, payer's and developers', surpluses	
$\hat{\lambda}_d$	The minimum level of the CET that guarantees the dynamic efficiency by ensuring non-negative profits to all developers entering the market. It is the dynamic reserve ICER of the industry. At this level some entrants make positive profits, and none makes a loss	$\hat{\lambda}_D$ : this is the individual developer dynamic reserve ICER.
$\lambda_c$	The effective CET level resulting from imposing cost-containment measures like budget caps, budget impact thresholds, etc.	
$\lambda^{avg}$	The value of the CET around which majority of the developers' reserve ICERS concentrate when a non-uniform distribution function for developers' reserve ICERS is assumed to model the supply curve.	

## Appendix 3

The model presented in this work considers the developers' decision to invest in R&D as a consequence of the level of the maximum WTP of the system. Such a maximum WTP level of the system is assumed to be the supply-side CET  $k$ . Thus, although we can derive the dynamic efficient value of the CET that the payer should set considering the amount of R&D investment that industry would have already devoted as a rational response to  $k$ , there is no economic incentive for a rational payer to behave in such a way. Put in other words, a surplus (or health gain) maximising payer would not care about the amount of R&D invested because it cannot be disinvested. Consequently, the payer would try to push down prices by lowering the CET until the point where the maximum health gain is achieved with the use of the existing resources and health technologies (including the new ones for which R&D investment has already taken place). Our model can discuss that this allocation is inefficient because we can measure the loss of the industry if the payer's health gain maximising threshold is below the industry's dynamic efficient ICER. Therefore, we can make judgements about potential inefficiencies resulting from a repeated game, if developers can adapt their expectations of the true actual threshold and then invest, as a consequence, a suboptimal amount of resources in R&D.

In this appendix we present a way to address the analytical constraint imposed by the assumed exogeneity of the threshold and R&D investment. The main change necessary to implement is a switch between stages 2 and 3 of the timing of the model presented in section 2. Figure A.3.1 represents the timing of the model with stages 2 and 3 switched.

**Figure A.3.1: TIMING OF THE MODEL**



The new timing in Figure A.3.1 then is explained as follows:

**Stage 1:** Nature fixes the level of the system's maximum WTP i.e. the value of  $k$ , the system's opportunity cost of the marginal health gain .

**Stage 2:** the payer (health care decision maker) commits to a CET level  $\bar{\lambda}$  to inform reimbursement recommendations and used at pre-bargaining decision making. The value of  $\bar{\lambda}$  will be equal or lower to the system's maximum WTP set by nature in stage 1.

**Stage 3:** the developer, knowing  $\bar{\lambda}$  (the CET established by the payer determined in stage 2) and its own reserve ICER, decides to (or not to) invest in R&D to develop a new medicine

**Stage 4:** the developer sets the price and the corresponding ICER using both the clinical and the health cost evidence in order to get reimbursement status.

**Stage 5:** the price of the new technology and the agreed effective ICER are set in a bargaining process at the procurement stage considering the bargaining power of both

players, i.e. the payer and the developer. It may therefore be below the price implied by the threshold  $\bar{\lambda}$  committed by the payer at stage 2.

**Stage 6:** medicine is sold by the developer to the payer and provided to patients. Outcomes are realized.

With the new timing proposed in this Appendix the decision of developers as to how much to invest in R&D becomes a function of the value of the CET committed by the payer. This has implications:

- i. A health gain maximising rational payer still maximises the consumer surplus at  $\lambda_p^* = \frac{k}{1+\beta}$ . The value of the NPB maximising threshold depends on the bargaining power though, producing values of the consumers' surplus maximising threshold within the range  $[\frac{k}{2}, k]$ . It is important to note that if the payer has all the bargaining power the consumers' surplus maximising threshold is at the value of the supply-side CET (payer's maximum WTP). With all the bargaining power ( $\beta = 0$ ), the payer extracts all the developers' surplus and so provide access to all technologies with reserve ICERs equal to the system opportunity cost becomes optimal strategy.
- ii. Developers take their decisions considering the information about the CET of the payer  $\lambda$ , implying that the value of  $k$  is now irrelevant to determining the dynamic reserve ICER of the industry. We obtain the new dynamic reserve ICER by introducing equation [21] into equation [27] and rearranging to obtain,

$$DS = \frac{\beta a \lambda^2}{2k} - \frac{\lambda}{k} R \quad [A28]$$

Equating [A28] to zero we obtain the new reserve ICER of the industry, or the value of the threshold that produces a non-negative surplus for the whole industry, which is,

$$\hat{\lambda}_d = \frac{2R}{\beta a}$$

- iii. The industry's surplus function is equal to zero at  $\lambda = 0$  and  $\lambda = \hat{\lambda}_d$ . It has a minimum at the following value of the threshold,

$$\operatorname{argmin}_{\lambda} \frac{\beta a \lambda^2}{2k} - \frac{\lambda}{k} R$$

Applying the first order condition we obtain:

$$\lambda^- = \frac{R}{\beta a}$$

The industry surplus curve when developers' decision to (not to) invest in R&D after knowing the payer's threshold is represented in Figure A.3.1 for different values of developers bargaining power.

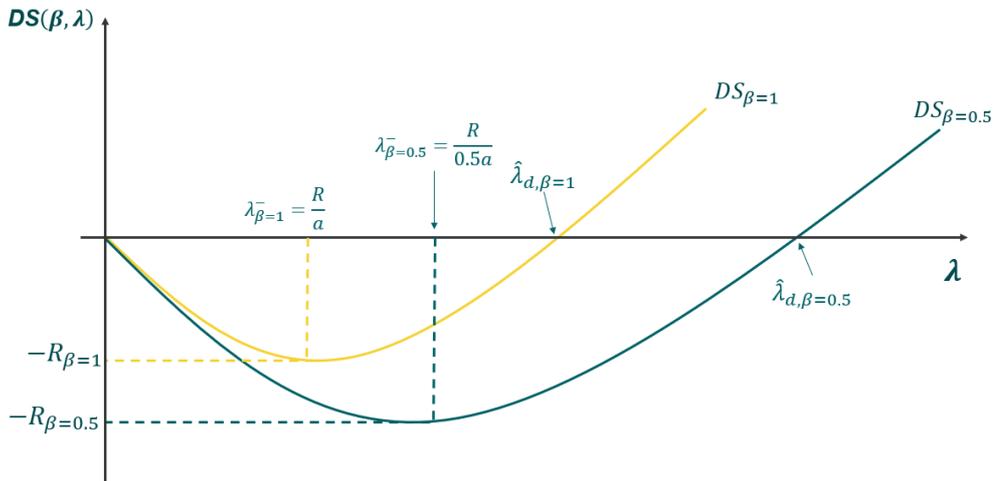
- iv. Because the decision of the developers about the investment in R&D depends on the value to the payer's threshold, the social surplus maximising threshold also changes. The new value of the threshold that maximises social surplus is now obtained by solving the following objective function

$$\operatorname{argmax}_{\lambda} (1 - \beta) \lambda \left( a - \frac{a}{2k} \lambda \right) + \beta \lambda \left( a - \frac{a \lambda}{k} \right) + \frac{\beta a \lambda^2}{2k} - \frac{\lambda}{k} R$$

Applying the first order condition we have that,

$$\lambda^* = k - \frac{R}{a} \quad [A34]$$

The social surplus maximising threshold is lower by an amount of industry's total R&D investment divided by the health gain that payer would obtain by getting all new medicines for free (i.e., the value of the parameter  $a$  of the function  $h(\lambda)$ ).



**Figure A.3.1: SURPLUS FUNCTION OF THE INDUSTRY WITH ENDOGENOUS RESERVE ICERs**

Table A.3.1 compares results obtained with the timing of this appendix with results obtained with the timing proposed in section 2 of the paper.

**Table A.3.1: COMPARATIVE OF RELEVANT RESULTS**

RESULT	SECTION 2	APPENDIX 3
Consumers' surplus maximising threshold	$\lambda_p^* = \frac{k}{1+\beta}$	$\lambda_p^* = \frac{k}{1+\beta}$
Developers' dynamic reserve ICER	$\hat{\lambda}_d = \sqrt{\frac{2kR}{\beta a}}$	$\hat{\lambda}_d = \frac{2R}{\beta a}$
Social surplus maximising threshold	$\lambda^* = k$	$\lambda^* = k - \frac{R}{a}$

Results presented in this section are preliminary and need further interpretation. They are part of follow up further research that is still ongoing. Although changes made on the model in this Appendix partially address the endogenous nature of thresholds and R&D investment, the authors judgement is that, to model such endogeneity properly a dynamic multiperiod model is needed. This might be a repeated version of the model presented in the main body of this paper where developers have adaptive expectations of the CET based on the history. Such specification would allow them to account for all long-term losses and gains accurately.



### About us

Founded in 1962 by the Association of the British Pharmaceutical Society, the Office of Health Economics (OHE) is not only the world's oldest health economics research group, but also one of the most prestigious and influential.

OHE provides market-leading insights and in-depth analyses into health economics & health policy. Our pioneering work informs health care and pharmaceutical decision-making across the globe, enabling clients to think differently and to find alternative solutions to the industry's most complex problems.

Our mission is to guide and inform the healthcare industry through today's era of unprecedented change and evolution. We are dedicated to helping policy makers and the pharmaceutical industry make better decisions that ultimately benefit patients, the industry and society as a whole.

*OHE. For better healthcare decisions.*

### Areas of expertise

- Evaluation of health care policy
- The economics of health care systems
- Health technology assessment (HTA) methodology and approaches
- HTA's impact on decision making, health care spending and the delivery of care
- Pricing and reimbursement for biologics and pharmaceuticals, including value-based pricing, risk sharing and biosimilars market competition
- The costs of treating, or failing to treat, specific diseases and conditions
- Drivers of, and incentives for, the uptake of pharmaceuticals and prescription medicines
- Competition and incentives for improving the quality and efficiency of health care
- Incentives, disincentives, regulation and the costs of R&D for pharmaceuticals and innovation in medicine
- Capturing preferences using patient-reported outcomes measures (PROMs) and time trade-off (TTO) methodology
- Roles of the private and charity sectors in health care and research
- Health and health care statistics