OHE

WHITE PAPER

Are cost-effectiveness thresholds fit for purpose for real-world decision making?

Patricia Cubi-Molla
Maria Errea
Kyann Zhang
Martina Garau

ohe.org

FEBRUARY 2020

WHITE PAPER

Are cost-effectiveness thresholds fit for purpose for real-world decision making?

Patricia Cubi-Molla
Office of Health Economics, London

Maria Errea
Office of Health Economics, London

Kyann Zhang
Office of Health Economics, London

Martina Garau
Office of Health Economics, London

Corresponding Author:
Kyann Zhang
kzhang@ohe.org

For further information please contact:

Martina Garau
Director, OHE

Tel   +44 (0)207 747 8867
Email mgarau@ohe.org

OFFICE OF HEALTH ECONOMICS
CONSULTING

OHE

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 a rapid 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 or approval of the OHE's Editorial Panel or Research and Policy Committee, or its sponsors.

Funding and Acknowledgements

OFFICE OF HEALTH ECONOMICS
CONSULTING

Table of Contents

OFFICE OF HEALTH ECONOMICS
CONSULTING

OHE

Executive summary

Resources available to pay for increasingly sophisticated healthcare options are limited, hence payers must develop reasoned approaches to allocate scarce resources to the most valuable technologies. To this end, Health Technology Assessment (HTA) methodologies have been developed over the past half century to help guide policy choices about which health interventions to cover. Most notably, some countries' HTAs have adopted cost-effectiveness analysis to inform coverage decisions, where health gains are expressed as quality-adjusted life-years (QALYs).

On their own, cost-effectiveness analysis results cannot determine whether the gain from a new technology is 'worth' the cost; that requires a cut-off point that indicates the maximum amount of expenditure deemed appropriate (by society or by policymakers) for an additional health gain, the so-called cost-effectiveness threshold (CET).

CETs can be determined based on either (1) a demand-side willingness-to-pay (WTP) perspective, where the total healthcare budget remains – theoretically – flexible, or (2) a supply-side opportunity cost (OC) perspective where total budgets are fixed and including a new technology implies displacing existing ones. Current CET estimates based on these approaches have been questioned due to limitations of underlying data, as well as assumptions made which do not always match reality. Some recent academic literature based on OC has produced considerably lower values than those currently applied in practice. Whether viewed from the demand side or the supply side, using the 'wrong' threshold produces an *inefficient use of resources*.

On the supply side, the nature of medical innovation is rapidly evolving. Over the next decade, innovations that offer lifelong cures will become available, and the type and sources of evidence generated to show clinical benefits might depart from traditional clinical trials.

The role of cost-effectiveness analysis and CET is becoming more prominent than what was envisaged at their introduction when they were mainly tools for efficient resource allocation. Increasingly they are becoming a basis for regulating price, either directly (e.g. Australia and now potentially in Canada) or indirectly (e.g. in the UK).

As established HTA systems, such as those in Australia, Canada, and the UK, review their methods and their role on healthcare system efficiency, there is a clear need to open the debate on how best to identify a threshold that could support the system to meet its goals, given limited resources.

In addition, QALYs alone cannot encompass all important aspects of value that a therapy may offer. For example, society may be willing to pay more for treating life-threatening diseases such as cancer, or for curative therapies, than for treating a self-limiting or chronic disease that does not seriously compromise daily living.

To move the policy debate forward and support HTA agencies in their decision making, we recommend:

- For systems considering moving to an explicit CET to do so in stages. For example, start by adopting the 'latent' or 'empirical' CET (the observed threshold based on past decisions). This approach, which has been implemented by NICE, has the advantage of avoiding major shocks to the reimbursement system (which can significantly affect patient access to valuable treatments) and, at the same time, improving the predictability and 'fairness' of decision making. A 'soft' approach should also involve the introduction of a range, rather than a single-value, CET to allow

some flexibility for decision making committees. Regular reviews of the established CET should follow to capture new relevant evidence on decision making.

- In systems already applying an explicit threshold, to explore a number of aspects:

  - understanding the relationship between supply-side and demand-side estimates and exploring how the two values can converge in the long-term when the health budget can be changed, and healthcare systems spend should match society's preferences for health improvements; and

  - recognising the data barriers underlying current empirical estimates and being receptive to new approaches for defining the concept of a threshold and approaches to set its value, such as those considering resource allocation across different public sectors and those assessing the distribution of value generated by new medicines to payers, patients and developers in the long term.

The QALY as the sole measure of additional benefit may not capture all the dimensions that matter to patients and society. How to capture a multi-dimensional concept of value in decision making and understanding the implications for the threshold are still open research and policy questions.

In addition, new innovative therapies may have immature evidence at the time of launch; be used in patient populations with different characteristics and potentially different conditions, which will require multiple Incremental Cost Effectiveness Ratios (ICERs); be used in combination with other technologies (such as diagnostics or other medicines), which may lead to different value and ICERs depending on whether use is in combination or alone. These are issues relevant to any value assessments, although they are particularly acute in systems relying on the QALY where strict cut-off points (CETs) are used to guide funding decisions.

In conclusion, a rigid, single, QALY-based CET (a main factor proposed in HTA-conducting countries such as Canada) may not lead to efficient resource allocation. Setting the threshold at a level that fails to achieve this efficiency will have negative implications for patient access to new, valuable health technologies. Multiple thresholds reflecting value elements beyond the QALY may be needed. More research is essential to understanding how new spending is accommodated in the system and if society is willing to trade off some health gains for other benefits valued by patients and society.

1 Introduction

How much to pay for healthcare is a concern of both public and private payers worldwide. Resources available to pay for increasingly sophisticated healthcare options are limited. As budgets are unlikely to expand to cover all possibilities, payers must develop reasoned approaches to decide how available resources should be allocated to ensure that the most valuable technologies are funded.

Several Health Technology Assessment (HTA) methodologies have been developed over the past half-century to guide policy choices about which health interventions to cover[1]. Many countries now require that manufacturers complete and submit HTA studies - performed according to specific guidelines - as a precondition when considering a new medicine for reimbursement (Angelis, Lange and Kanavos, 2018). In the past, this required demonstrating that the product was at least as good as- if not superior to - standard care with existing medicines. More recently, there are additional requirements for studies showing that in addition to superior in therapeutic benefit, the product under consideration is also better in terms of its value for money when comparing its incremental benefit to its incremental cost. In a number of countries, these gains are expressed as quality-adjusted life-years (QALYs), a measure that considers both additional years of life and the quality of that life, i.e. closeness to full health. The use of QALYs allows comparisons of outcomes across different types of therapy - e.g. medicines can be compared to surgery or to, say, behavioural therapy - and disease areas - e.g. outcomes for cancer patients can be compared to those for cardiovascular or, say, mental health patients. See Appendix 1 for further details about the QALY as a measure of health gains.

Some payers have taken this a step further by establishing *how much* additional value a new therapy must provide as a precondition for funding or reimbursement. In practical terms, this usually requires the incremental cost *per unit of health* delivered by the new treatment to be below a threshold specified by the payer. When QALYs are used to characterize these *units of health,* an Incremental Cost-Effectiveness Ratio (ICER) is estimated. **A public payer might require the ICER of a new therapy to fall below a specific amount, a threshold**. In the UK, for example, the National Institute for Health and Care Excellence (NICE) requires that a new intervention fall under a threshold of £20,000 to £30,000 per QALY gained. Interventions above that range are unlikely to be recommended for funding by the NHS. Exceptions apply for end-of-life conditions or for certain ultra-rare diseases, for which a higher threshold is considered.

HTA and cost-effectiveness thresholds (CETs) were established to ensure public money is used to purchase interventions that provide the highest value (in terms of health gains) and therefore achieve efficiency. The practical challenge is setting the threshold in a way that makes sense from an economic and policy point of view, taking into consideration not only constraints on healthcare budgets, but also the needs of patients and incentives for continuing improvements in treatment.

The way by which current thresholds are determined by HTA agencies (or other relevant authority) has been found – in general – to lack a sense of clarity. Fresh debate on both the desirability of strict thresholds and how they are set is needed.

The key areas for debate are the following. First, **empirical evidence is scarce** for reliably determining the appropriate value of one extra unit of health gain (a QALY). Policymakers are in danger of relying on estimates that do not necessarily reflect the reality of healthcare systems. Therefore, understanding the underlying principles and assumptions behind these estimates of value

---

[1] *For a thorough summary of HTA methodologies and their relative strengths, see Kobelt (2013). See also a glossary with key terms in A.4 and information in A.1.*

is crucial. Some recent academic literature supports an approach for estimating thresholds that produces considerably lower values than those currently applied in practice. This originated with the publication of Claxton et al. (2015) and has subsequently gained international attention (Claxton et al., 2015). The key criticism of this approach is the sparsity of necessary empirical evidence. It has been noted by critics that the assumptions necessary to compensate for such a lack of data, lead to results that do not adequately reflect the way healthcare systems work in practice (Hernandez-Villafuerte, Zamora and Towse, 2018).

Second, **the nature of medical innovation is rapidly evolving**. Over the next decade, medical innovations that offer lifelong cures will become available. Society might value these curative therapies more highly than treatments that offer the same incremental gains over many years (Hampson, Towse and Henshall, 2017). HTA methodologies need to be adjusted to take this into consideration (Garrison et al., 2019; Danzon, Towse and Mestre-Ferrandiz, 2015).

Finally, and most importantly, **there are major reforms on the horizon** in countries where the QALY is the most commonly used measure of health benefits and where it drives the definition of CETs. In **Canada**, the Patented Medicines Prices Review Board (PMPRB) is in the formal consultation stage of establishing a **new nationwide system of list and net pricing**, where the maximum rebated price (net price) of patented medicines with yearly sales or annual treatment cost above a certain amount is set using market size and the pharmacoeconomic factors of treatment cost, incremental cost and incremental QALYs valued using a CET of CAD 60,000/QALY[2]. This figure is based on a study by Ochalek, Lomas and Claxton (2018), and has been debated due to uncertainty surrounding the estimates provided, the key limitation being that the relationship between health expenditures and health outcomes was not based on Canadian data (Patented Medicine Prices Review Board, 2019a).

In the **UK**, **NICE has started a method review for the evaluation of a wide range of technologies**, including medicines and diagnostics. Key topics considered include whether factors other than the QALY might be included in decision making (so-called 'modifiers'), and if so, which factors should be included. The review will also address specific challenges associated with evaluating next-generation treatments such as tumour-agnostic therapies and gene therapies which may potentially offer a cure[3].

The role of cost-effectiveness analysis (CEA) and CET is becoming more prominent as they are seen not just as tools for efficient resource allocation, but as a basis for **regulating price**, either directly (e.g. in Sweden, Australia and now potentially in Canada) or indirectly (e.g. in the UK) (Angelis, Lange and Kanavos, 2018; Deloitte, 2012). In this paper, we focus on three (high-income) countries (Australia, Canada, and the UK) that have similar HTA systems relying on the QALY framework.

The purpose of this paper is to inform the debate among healthcare policymakers and other stakeholders about the relative merits of current approaches to defining, estimating and applying CETs in healthcare decision making.

---

[2] *To summarize, prescription medicines new to the Canadian market will be divided into two categories based solely on market size. 'Patented medicines' with yearly sales expected to exceed either 50% of per capital GDP (in June 2019, this was roughly CAD 29,700 (CECI, 2019), or CAD 25 million, will be assigned a 'maximum rebated price' (MRP) calculated using market size and pharmacoeconomic factors of treatment cost, incremental cost and incremental QALYs valued using a CET of CAD 60,000/QALY. The MRP does not change over time unless certain conditions are met: a change in market size, a new indication, or an updated HTA assessment. Category II medicines will be priced using an international reference pricing approach; the list has been modified to remove countries with higher prices (the US and Switzerland). Such medicines may be reclassified as Category I based on increased revenue or a new indication. Products currently on the market are grandfathered (exempt), but the stated intention is to assess and price all products over time using this approach (Patented Medicines Prices Review Board, 2019b).*
[3] *https://www.nice.org.uk/news/article/nice-announces-details-of-health-technology-evaluation-methods-review*

2 Countries' use of CETs

Countries may apply implicit or explicit CETs. ***Explicit* thresholds are those formally stated** in official guidelines, and are generally accepted as key determinants in the decision making process (Baker et al., 2011). ***Implicit* thresholds** are those not specified in official documents but **are generally recognised and informally used** by decision makers. In the literature, implicit thresholds also refer to those that can be inferred retrospectively by analysing ICERs of previously recommended interventions (these can be defined as *empirical* thresholds) (Eichler et al., 2004; Baker et al., 2011).

Many countries lack an explicit CET and use alternative benchmarks based on real-life funding decisions or precedents. For example, the average annual cost of $50,000 for kidney haemodialysis has been used as a rule of thumb in the US for many years (Garrison et al., 2019).

The following table presents the CETs found in the literature in relation to countries considered in this paper (Australia, Canada, and the UK). Of the three, the UK[4] is the only one that uses an explicit threshold (as set out in the NICE Methods Guide). Two years after NICE was established in the UK, researchers noticed that the cut-off value for recommending a new technology for NHS coverage was frequently £30,000/QALY (Raftery, 2014; Towse et al., 2002). NICE then established an explicit threshold range of £20,000 to £30,000 per QALY gained, which has remained unchanged since 2004.

Australia and Canada, on the other hand, have relied on implicit CETs based largely on precedents and not described in any formal document or guideline.

---

[4] In this report we refer to NICE as the HTA agency operating in the UK for simplicity of reporting. However, we note that formally NICE's remit is only England and Wales. We note that there are HTA agencies responsible for the other UK countries, which follows similar cost-effectiveness modelling methods and CETs values.

Table 1: Applied thresholds in Australia, Canada and England

Country	HTA agency	Implicit vs explicit	Threshold (cost per QALY)	Reference
Australia	Pharmaceutical Benefits Advisory Committee (PBAC)*[5]	Implicit, values presented are based on empirical studies analysing past decisions	AUD50,000	Wang, Gum and Merlin (2018)
Canada	Canadian Agency for Drugs and Technologies in Health (CADTH)**	Implicit, values presented are based on empirical studies analysing past decisions	'Policy thresholds' as indicated in PMPRB (2019): CAD50,000 (non-oncology) CAD100,000 (oncology) 'Empirical' threshold as indicated in Skedgel, Wranik and Hu (2018) CAD140,000*** (oncology)	Patented Medicine Prices Review Board, (2019a); Skedgel, Wranik and Hu, (2018);
UK	National Institute for Health and Care Excellence (NICE)	Explicit as indicated in the NICE methods guide	£20,000 – 30,000 End-of-Life (EoL): £50,000**** Highly Specialised Technologies (HST): £100,000 - £300,000*****	National Institute for Health and Care Excellence, (2018) ******

* http://www.pbs.gov.au/info/industry/listing/participants/pbac
** https://cadth.ca/
*** https://cadth.ca/pcodr
**** https://www.nice.org.uk/process/pmg9/chapter/the-appraisal-of-the-evidence-and-structured-decision making
***** https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-highly-specialised-technologies-guidance
******https://www.nice.org.uk/

Most CETs, implicit and explicit, either lack a theoretical basis or are informed by limited empirical evidence. This situation has contributed to an on-going debate on how best to identify a threshold that could support the system to meet its goals, such as that of maximising health gains given the limited resources. Approaches for addressing this issue represent either the **supply side** (all the possible interventions that healthcare can offer) or **demand side** (seeking to identify the appropriate spending limit from the viewpoint of the society). Each theoretical approach involves a

---

[5] A number of medicines for life saving drugs for life threatening are also funded via the Life Saving Drugs Program in Australia. We do not report it in the table because it does not ft in the PBAC HTA process. See https://www1.health.gov.au/internet/main/publishing.nsf/Content/lsdp-criteria for more information

different set of methodologies for the estimation of the threshold, which is explained in the next sections.

3 Is a threshold needed?

The purpose of CEA is to determine whether a new technology provides value for money compared to the existing standard of care. The definition of 'value' determines what HTA methodology is most suitable.

Where value can be expressed in monetary terms, cost-benefit analyses (CBA) is the optimal approach. Net monetary benefit (NMB), the basis for determining value, is usually calculated as the difference between the benefits of each technology minus the differences in the costs. If the CBA of a new technology shows that its NMB is positive, then the new technology should provide good value for money.

In most HTA methodologies, however, the value associated with new health technologies cannot be easily translated into monetary terms. Placing a monetary value on human life can be considered morally questionable and not socially acceptable. Value may instead be expressed as 'utility', based on preference-based quality of life (QoL) measures. Expressing health outcomes as utilities has the advantage of producing results that allow comparison across a range of treatments - this is known as **cost-utility analysis**. The most commonly used unit measure for health gains is the QALY, which forms the basis for calculating ICERs. On their own, ICERs cannot determine whether the gain from a new technology is 'worth' the cost; that requires a threshold that indicates the maximum amount of expenditure deemed appropriate for one unit of health gain (Vallejo-Torres, García-Lorenzo and Serrano-Aguilar, 2018). Alternatively, an ad-hoc value judgement from the decision making committee as to whether or not the new technology represent good value for money may be used.

Thresholds may have other important, broader functions. In viewing NICE's threshold from the perspective of ethics and philosophy, Badano, John and Junghans (2017) assert that the continued use of the threshold plays an important role in the 'complex ecology' of politics: **the existence of a standard that remains stable over time has several positive effects**, including giving an appearance of fairness, stabilising expectations, and providing a basis for democratic discussions about allocation of resources. What is unclear is whether the same benefits are likely to be the same in the political and social culture of different countries and healthcare systems.

A criticism of applying any threshold is that it may not capture societal preferences accurately. If the purpose of a threshold is to identify the maximum amount that society is willing to pay for an additional unit of health gain - the demand side perspective - then using a threshold that is too high or too low means that resources are being misallocated. A threshold would be considered too high if it results in allocating more resources to health than society would prefer, thus reducing allocations to other areas such as education or defence. Setting a threshold that is too low would mean that too little is allocated to health and health gains would be foregone because treatments that should be considered cost effective would not be available.

There are similar concerns from the **supply-side perspective, which assumes that every new investment requires the reallocation of resources from somewhere else.** In the case of health, this would mean the displacement of an existing treatment or service. A threshold would be considered too high if resources invested in a new treatment provide fewer health gains than the treatment it displaces. If the threshold is too low, treatments that provide good value for money would be rejected, resulting in a health loss or lower health gain in societal terms. **Whether viewed from the**

demand side or the supply side, using the 'wrong' threshold produces an *inefficient use of resources*.

Recent studies have criticised the use of thresholds based on QALYs because they provide only a partial picture of everything that should be taken into consideration in healthcare decision making. Schwarzer et al. (2015) argue that the simple comparison of an ICER against a benchmark cannot capture all values important to a society, including intangibles such as distributional justice, equity, and dignity. Research has confirmed that the comparison of the estimated ICER to the threshold was the dominant consideration in NICE decisions between 1999 and 2002 (Devlin and Parkin, 2004) and that, more recently, most NICE decisions (82%) were driven by cost-effectiveness considerations (Dakin et al., 2015). However, Dakin et al. (2015) also found that, in reality the threshold at which interventions were recommended was higher than the stated £20,000 to £30,000 range. The authors noted that this might be due to other factors considered by committees, such as End-of Life-criteria.

Making decisions based primarily on CETs has been rejected in some countries. For example, Germany historically refused to use a single universal threshold because of 'concerns regarding solidarity, equity and fairness'. The German HTA body explicitly states that HTA should 'inform the decision maker about the efficiency of a given technology', but it should not attempt to 'judge … how much should be spent on it.' (Institute for Quality and Efficiency in Healthcare, 2009).

Even in countries where the use of a threshold is strongly supported, experience shows it cannot be applied in every instance. The NHS in England, for example, currently provides many services that cost more than £20,000 - £30,000 per QALY gained. One example of this is palliative care, which produces very few QALYs as its purpose is to provide comfort and emotional support rather than increase quantity or quality of life. In some cases, disinvestment should be considered to divert resources to more cost-effective uses.

The NHS also places a high priority on factors such as shorter waiting times and equality of service provision across the country. These are not judged against the cost-per-QALY framework and, if they were, would almost certainly be above the £20,000 - £30,000 range.

NICE itself has explored in the past, as part of the Value Based Assessment consultation, whether to define additional considerations in its decision making process that go beyond the maximisation of health gains (i.e. modifiers). This might include additional factors that might change the value of the QALYs gained, such as severity of illness, whereby health gains for worse-off population groups might be valued more highly by society even though such patients might benefit less from treatment than other patient populations (Nord, 2005). A possible way to reflect these considerations is to introduce weighted QALYs (Wailoo, Tsuchiya and McCabe, 2009). A similar principle has been applied by the HTA agencies in the Netherlands and Norway (Reckers-Droog, van Exel and Brouwer, 2018; Ottersen T et al., 2016). It might be expected that NICE will consider such approaches for the inclusion of QALY modifiers as part of its methods review. **This suggests that some policymakers are questioning whether a one-size-fits-all threshold is the best choice.**

4 How are thresholds determined?

The dominant approaches currently used to determine cost-effectiveness thresholds are based on opportunity cost (OC) (a supply-side perspective) or willingness-to-pay (WTP) (a demand-side viewpoint).

Which is more appropriate depends in large part on the actual and potential healthcare budget available. Each approach has its proponents, as well as its own set of limitations that affect whether and how healthcare allocation decisions can, or perhaps should, be made based on them.

4.1 Supply side: OC and league table approaches

The OC approach assumes the current health budget is fixed, i.e. it will neither increase nor decrease and any new technology introduced must displace a technology or service currently in use (see Claxton et al., 2015; Zhang, Ji and Li, 2017). This approach assumes that the objective of any authority making health decisions under a finite budget will be to realise as many QALYs as possible within the given budget. The new technology will provide new health gains, but health gains associated with the displaced technology will be lost. The OC of funding a new technology, then, will be equal to the health gains lost by defunding the displaced technology.

The league table approach also assumes a finite budget. All health interventions are ranked in order of cost-effectiveness. Funding begins with the intervention at the top of the table and works its way down in rank order until the budget is exhausted (Thokala et al., 2018; Cleemput et al., 2011; Smith and Richardson, 2005). With the OC approach it is possible to use a league table analysis to set the threshold at the level of the least cost-effective technology funded.

The league table approach can only be used when perfect divisibility holds, i.e. when the health intervention can be considered entirely separately from any other aspect of healthcare (Thokala et al., 2018) and when health programs can be reduced to different levels. This is an issue for the OC approach as well (Cleemput et al., 2011). In practice, the league table approach is virtually impossible to apply because it requires that the decision-maker have information on the cost-per-QALY gained of all potential health interventions (Ryen and Svensson, 2015).

4.1.1 How OC thresholds are determined

Several methods have been applied to estimate OC thresholds. Although they are based on a similar theoretical approach, they differ importantly in the specific estimation technique applied and the results obtained.

Some **qualitative research** has explored how decision-makers react in practice to changes in the budget. Karlsberg Schaffer et al. (2015, 2016) explored local disinvestment decisions in Scotland and Wales, using the annual budget scrutiny survey and interviews with a sample of NHS finance directors in both countries. The results showed that **existing services are displaced only as a last resort**: NHS managers are more likely to try to increase efficiency than reduce or stop services.

Another qualitative study explored whether and how economic evaluation was used by Primary Care Trusts (PCTs) to make local healthcare decisions in England. Information was gathered by observing decision making meetings, in-depth interviews with decision-makers and analysis of documents from the meetings. The results showed that, **although decision-makers were aware of scarcity issues and the concept of value for money, economic evaluations were generally not a consideration for decisions** (Eddama and Coast, 2009).

Quantitative analyses have attempted to reveal the relationship between changes in budget and the funding of health services. The most recent estimates of the CET (from the supply side) are based on this approach. A summary of the estimates is shown in Table 2.

Quantitative methods focus on **the effect of healthcare expenditure on health outcomes, which is usually indicated as 'elasticity'**. The estimation of elasticity typically involves observable health

outcomes that are not measured as QALYs (see column 'Measure of Health Outcomes' in Table 2), for which data are rarely available.

A second key challenge is capturing the effect of expenditure on outcomes, as opposed to the other way round. An increase in healthcare expenditure can also be a consequence - rather than the cause - of positive health outcomes (e.g. a population with a longer life expectancy will consume more healthcare). In econometric analyses, the use of an 'instrumental variable' (IV) is a widely accepted way of teasing out what is *cause* and what is *effect*. In our context, an IV instrument is a factor that only affects health outcomes via its effect on health expenditure, in order to purge the estimation of outcome elasticity to health expenditure from any reverse causation from outcomes to expenditure, or from measurement errors in health expenditure. Several indicators have been used as IVs to tackle this problem. The use of these instruments is frequently justified either by using statistical tests (such as socioeconomic or deprivation variables, as in Claxton et al. (2015) and Lomas, Martin and Claxton (2019), or they are supported on theoretical grounds (for example, on the `funding rule' that allocates health budget across local authorities as in Andrews et al., 2017; Claxton, Lomas and Martin, 2018).

Lomas, Martin and Claxton (2019) estimate the CET in England using socioeconomic and deprivation variables as IVs with good statistical properties. Their work builds on an extensive report published in 2015 (Claxton et al., 2015) using a method which linked health expenditure with mortality for different clinical areas and across local decision-makers. Claxton et al. (2015) calculate a CET of £12,936 per QALY gained for 2008. **Lomas et al. report a range of £5,000 to £15,000 per QALY gained for 2003-2012**. However, these estimations require questionable assumptions, which we highlight below.

▪ **Both papers use mortality rates as the main health outcome**, which leaves several clinical areas (e.g. maternity) without observable outcomes. The lack of data for mortality rates and expenditure, as well as the mismatch of such data by clinical area requires a first set of assumption and complementary datasets to translate the observable mortality outcome by disease and commissioner (standardised *years of life lost* rate) to absolute number of life years lost. Yet, questionable assumptions were applied to translate *reduced life years lost* into *QALY gains* (using the Global Burden of Disease study), to then extrapolate the results.

▪ **Critics also point out that the approach taken by Claxton, Lomas and Martin (2017) overlooks a number of important factors** which may be relevant when making funding decisions to be applied across the nation: crucial differences across locations, the importance of each clinical area to each decision making group (in terms of both number of patients and healthcare provision), and possible differences in priorities across disease areas.

▪ Even in the optimal situation where a single CET does acknowledge heterogeneity across locations and disease areas, the weighted average of clinical area-specific CETs (which provides the CET of £12,936 per QALY gained for 2008) is arguably not the correct indicator for the marginal effect. This is because an implicit assumption is needed to interpret the average as marginal productivity across locations, which is that all decision makers prioritise similarly by clinical area when displacing services to adopt a new technology. In reality, empirical studies of NHS Health Boards across the UK suggest this is not the case. These authorities- and presumably the local healthcare decision-makers in England (PCTs at the time of the analysis) - do not respond in the same way to the same funding challenges (Karlsberg Schaffer et al., 2015, 2016).

A recent study (not reported in Table 2 as it does not provide an estimate) explores the relationships between outcomes and expenditures from two perspectives:

- first, incorporating the possibility that some decision making groups at the local level are more efficient than others in some clinical areas, and recognising than some; commissioners could adopt new technologies without displacing current services;

- second, directly measuring the heterogeneity of the mortality elasticity to health expenditure in different locations (Hernandez-Villafuerte et al., 2019).

This study provides empirical evidence of production inefficiency - that is, the inability of some PCTs to achieve the best practice performance found in others. This means that estimates of the OC of introducing new technologies based on average performance (as in Claxton, Lomas and Martin, 2017) could be biased and subject to far greater variation than assumed. The authors conclude that scientific methods alone may not be sufficient for setting a cost-effectiveness criterion, and that political judgements need to be incorporated as well.

In Canada, **the most recent estimation of the CET from the supply side can be found in Ochalek, Lomas and Claxton (2018),** which was performed and discussed as part of the price reform proposals by the Patented Medicine Prices Review Board (PMPRB), 2019**.** The authors provide province-level estimates of health OCs. Health outcomes are expressed in terms of disability-adjusted life years (DALYs). The authors use different values for elasticities, mainly obtained from Bokhari, Gai and Gottret (2007), Claxton, Lomas and Martin (2017) and Andrews et al. (2017). None of these studies used data from Canada. **Results vary substantially depending on how the elasticities are estimated**. The authors conclude that the potential CET (expressed as cost-per-DALY averted) falls between CAD20,000 and CAD100,000. This result is consistent with those obtained using elasticities from Claxton, Lomas and Martin (2017), and Andrews et al., 2017. Notably, in the context of the Canadian reforms, **Ochalek, Lomas and Claxton (2018) recommend a cost-per-QALY threshold of CAD30,000 per QALY gained. However, this is not supported by evidence in the paper**, which was based on estimations expressed in DALYs. As a result, it was debated and critiqued by experts who were part of the PMPRB Working Group (Patented Medicine Prices Review Board, 2019a).

Edney et al. (2018a) provide an estimate of a CET for Australia. The authors also use an IV estimation methodology, which leads them to a **base case estimate of AUD28,033 per QALY gained** (ranging from AUD21,000 to AUD38,000). The authors use national data on health-related quality of life for their estimate. Edney et al. (2018b) show that the estimation of the elasticity of mortality outcomes to public health spending in Australia varies greatly across locations, with those with the highest mortality rates being more able to reduce mortality for a given increase in spending.

Although the UK and, more recently, Canada, have publicly embraced the concept of OC thresholds as part of their HTA process, in practice CET values applied for decision making are different from the empirical estimates based on that concept. The estimates presented in Table 2, representing 'empirical estimates' of the OC in each country's healthcare system, are lower than the figures in Table 1, which represent the 'decision thresholds'. The consequences of this discrepancy remain to be researched, but they crucially depend on the extent to which the estimated CETs represent, as they are intended to do, OCs and the marginal productivity of, say, the last one million pounds invested.

In the table in Appendix A.4, we included other recent OC estimates in countries that are outside our scope to provide a comprehensive picture of empirical evidence and methods in the field.

All methodologies for basing thresholds on OC suffer from similar limitations:

- first, and most importantly, **relying on past decisions does not necessarily produce a more efficient decision making process in the future**, as some studies have made clear (Vallejo-Torres et al., 2016);

- second, **the relationship between expenditure and health is not the same across all countries**. This means that a single methodology is not applicable across all countries. In addition, causality of the elasticity of health outcomes to health expenditure is very difficult to prove; even assuming the validity of the econometric model and IVs used with available data, there may be unobserved factors, such as behavioural factors affecting inertia, as well as trends in patient life expectancy and morbidity;

- finally, **data are insufficient**. Empirical studies so far have relied on mortality statistics and have not been able to estimate the impact of expenditure on quality of life alone. Claxton et al. (2015) attempted to compensate for this limitation by inferring the effect on quality of life as being proportionate to the estimated effect on mortality, which is a questionable assumption.

Table 2. Most recent country OC estimates

Reference	Country	Measures of Health Outcomes	Measures of expenditure/sources	Central estimate reported	Estimates range	
Edney et al. (2018a)	Australia	Years of Life Lost; morbidity; QALYs (SF-36)	National data from 2011/2012 capturing government health expenditure and healthcare need; Data on veterans	AUD28,033	AUD20,758	AUD38,000
Ochalek, Lomas and Claxton (2018)	Canada	Mortality, Years of Life Lost, Years Lived with Disability, DALYs	Elasticities form Bokhari, Gai and Gottret (2007): Military expenditure per capita of neighbouring countries; paved roads per square km; shock in 'donor funding'	n/r	CAD83,000	CAD249,000
			Elasticities from Claxton, Lomas and Martin (2017): English NHS expenditure; instrumental variables used reflecting socioeconomic deprivation	n/r	CAD16,000	CAD52,000
			Elasticities from Andrews et al. (2017): English NHS expenditure	n/r	CAD23,000	CAD76,000
			QALY. (Ochalek, Lomas and Claxton, 2018) citing the results obtained by Woods et al. (2016)	CAD28,089	CAD27,000	CAD34,000
Lomas, Martin and Claxton (2019)	UK	Year Life Lost Rate	NHS expenditure by geographically defined local health authorities, for each programme budgeting category (total: 23)	n/r	£5,000	£15,000
Claxton et al. (2015)	UK	Year Life Lost Rate	NHS expenditure by Primary Care Trust, for each programme budgeting category (total: 23)	£12,936	£2,018	£29,314

n/r: non-reported

4.1.2 How league tables are developed

Several generic league tables have been developed. WHO-CHOICE reports simple information on ICERs for a large number of interventions. Regional league tables are available for several diseases or risk factors.

The main advantage of league tables is that the last-ranked intervention funded (the intervention with the lowest ICER within a set of interventions) is thought to be a good approximation of willingness-to-pay for health benefits. League tables can also indicate the potential health benefits of cancelling an existing programme and reallocating resources to another programme.

To provide a reliable guide for decision-makers, the same outcome measure should be used for each intervention. In practice, although HTAs may apply a consistent methodology and rely on recommended data on resource consumption, current guidelines do not, and cannot, specify a single common outcome measure.

As a result of these challenges, a comprehensive and reliable league table of all possible health interventions in any given health system, is beyond the reach of today's decision-makers in most countries.

4.2 Demand side: WTP and GDP approaches

The WTP approach assumes the health budget is not finite but fluctuates with demand. In a single-payer system, this assumes the government knows the relative value that society would assign to all government services—health, education, defence and so on. A threshold would be set based on society's maximum willingness-to-pay for health, expressed perhaps as a cost-per-QALY. All new technologies at or below this threshold would be funded (Smith and Richardson, 2005; Mason, Baker and Donaldson, 2008).

The World Health Organisation (WHO) suggested that gross domestic product (GDP) might be a valid reference indicator for threshold setting, recommending a threshold between one and three times the GDP per capita per DALY (Robinson et al., 2017; Sachs, 2001). Although this approach does not have a clear theoretical basis, it has been applied in low- and middle-income countries. In the early 2000s, 10% of cost-effectiveness studies cite this threshold as a benchmark and by 2013–2015, this increased to 76% (Thokala et al., 2018). Some analysts have argued that thresholds at three times GDP per capita are considerably higher than those derived from the OC approach. In response to criticism, the **WHO has recognised the limitations of using GDP to set a threshold** (Bertram et al., 2016) and has discontinued its support of this approach.

4.2.1 How WTP thresholds are determined

Two primary methods are used to determine the maximum WTP value for an intervention (see Ryen and Svensson, 2015; Vallejo-Torres, García-Lorenzo and Serrano-Aguilar, 2018).

The first method surveys members of the public directly about their preference on health improvements. This direct elicitation method uses hypothetical questions to discover preferences. This might be a binary/dichotomous approach where respondents are asked whether their WTP is higher (or lower) than a specified amount for a given health improvement. Alternatively, it might ask respondents to choose between options that differ with respect to price/cost and level of health improvement (e.g. using discrete choice experiments) (Ryen and Svensson, 2015). The data collected can be used to estimate the WTP for the heath gain represented by an additional QALY (so-called 'WTP-Q').

The second method is based on the value of statistical life (VSL), which is equivalent to the monetary value of a prevented fatality (Viscusi, 2003). VSL can be estimated either by revealed preference or stated preference methods. Revealed preference methods assume that individuals reveal their preferences by their market behaviour. The information is obtained by identifying situations in which individuals, either implicitly or explicitly, perform actual trade-offs between wealth and mortality risks. In stated preference methods, a hypothetical market situation is presented and survey respondents are asked about their WTP for a given change in the level of mortality risk. In general, WTP for a QALY can be estimated from VSL studies by using a present-discounted-value formula (Ryen and Svensson, 2015).

Whichever method is used, **the WTP approach has inherent limitations**:

- WTP results differ depending on the technique used to elicit them. This has a discernible impact on final estimated values. A systematic literature review identified 24 papers with 383 WTP-Q estimates. The overall mean and median WTP-Q estimates found were €118,839 and €24,226, respectively (in 2010 Euros). **Estimates ranged from less than €1,000 to €4,800,000**, although 80% of all estimates were below €75,000. In addition, estimates of WTP using the stated preference approach were lower than those found using the VSL. WTP-Q also tended to be higher when a risk of premature death was included in the valuation scenario than if pure quality of life (QoL) changes were being valued (Ryen and Svensson, 2015);

- The WTP-Q method assumes that the valuation of health gains is linear with respect to changes in quality and duration of life. In other words, if an individual is willing to pay, say, £30,000 for one additional QALY, then the individual would be also willing to pay up to £60,000 for two additional QALYs. However, there is evidence in the literature that this assumption does not hold (Vallejo-Torres et al., 2016);

- WTP questionnaires imply a substitution of health – defined as QALYs – and wealth, which results from a trade-off between QoL and time to death/risk of death. For instance, say that an individual is willing to pay £30K for a QALY. This should be equivalent to say that the individual is willing to pay £30 for two additional years of life lived with utility 0.5, or four additional years of live lived with a QoL of 0.25, and similar partitions of quality and duration of life that are equivalent to one QALY. However, individuals are sometimes willing to make a trade off in one dimension but not in the other, and therefore the aggregation of trade-offs between elements of different kind frequently leads to inconsistencies and wide confidence intervals across the estimates.

We conclude that **current available methods produce a range of WTP estimates that is too wide to be useful for healthcare policy decision making**.

4.3 The procedural objectivity approach

Some recent publications by authors with expertise outside the fields of health or economics suggest that the basis for a threshold is less important than the existence of one. Badano, John and Junghans (2017), for example, assert that the most significant contribution of NICE's threshold is intangible: conveying a sense of fairness to coverage decisions, stabilising expectations about the prospects of funding for new health technologies, and providing a basis for discussing healthcare choices.

Although we did not find any obvious position of this approach from our review of HTA methods in the selected countries, the fact that some refer to a CET (either implicitly or explicitly) with no clear empirical basis could be – at least in part – the result of the above factors. That is, HTAs may apply

CETs to improve decision making processes by ensuring some level of consistency across decisions and predictability of their process, for the benefit of all the stakeholders.

4.4 Can we reconcile demand- and supply-side CETs?

The debate about whether supply-side or demand-side factors should determine CETs is likely to continue. Supply-side thresholds allow the quantification of the net health gains (or losses) that would result from the inclusion of a new intervention. Decisions made on the basis of supply-side CETs ensure that aggregate health is maximised within budget constraints. Estimates of demand-side CETs, such as WTP, tend to be higher than those based on OCs and raise issues of affordability in the short term, as realistically, not all technologies meeting a WTP threshold can be funded.

Theoretically, if health budgets can be adjusted in the long-run, WTP and OC approaches should produce the same threshold and ensure that healthcare system spending matches society's preferences for health improvements (Danzon et al., 2018). It is worth noting that if budget flexibility is allowed in the health sector, it can also apply to other areas of spending, thereby affecting aggregate public spending. Whether the resulting taxation measures are accepted by the public remains an empirical question. In practice, however, budgets are likely to remain constrained, with small increases each year, for the foreseeable future.

5 Can QALY-based approaches reward valuable future innovations?

Policy changes often have unforeseen and unintended consequences. For medical technology, a QALY-based approach to judge value for money may have effects well beyond negotiating prices or determining coverage. HTA recommendations and pricing decisions (together) may be interpreted as a signal about decisions on R&D efforts, which in turn can have a profound impact on innovation. For some types of interventions, HTA methodologies based on the QALY framework have struggled to accurately capture their value, a problem that is likely to become more serious as more innovative technologies come onto the market.

One topic that occupies a large part of the discussion is that of **gene therapies**, which is a technique that alters an individual's genes to treat – or cure – a disease. This may be achieved by replacing a disease-causing gene with a healthy copy of the same gene; deactivating a disease-causing gene; or introducing a new gene into the patient. This is an area identified to have experienced rapid growth worldwide (Ginn et al, 2018), and offers potential for one-time treatment targeting the fundamental cause of a disease, rather than long-term management of symptoms. This then may lead to substantial savings in costs over a patient's lifetime but is also associated with large cost upfront, at least in the near term.

A second trend that poses fundamental problems for HTA valuations is **combination therapy**, in which more than one intervention is used as part of a regimen to treat a disease to achieve greater health gains than is possible when components of the regimen are used separately. Combination therapy is familiar in the fields like HIV but increasingly applied in oncology. The challenge for HTA is how each of the components should be valued, an issue that is further complicated when each component is manufactured by different companies which will expect a fair price for their intervention part of the combination.

Outlined below are the most relevant issues posed by the current medical innovation for HTA methods. Some, such as increased uncertainty, are not confined to QALY-based approaches as they also apply to other methods for the assessment of value. Few of these issues are new—most have been raised and debated at length by HTA bodies-and most still await consensus on how they can be best addressed. What is new is the growing number of medical innovations that raise such issues and their growing importance in the policy debate.

- **Uncertainty about the eventual outcome.** Clinical trials are – by necessity - limited in population size and timeframe. This presents particularly serious challenges in the development of medicines for rare diseases (which have small populations) and for therapies that can be proven only in the long term to have provided a cure (timeframe). There are a number of aspects to uncertainty that need to be considered:

  - *Surrogate endpoints.* For a number of reasons (e.g. resource or time constraints), clinical trials may use surrogate outcomes such as progression-free survival in oncology. Trial results may report outcomes on certain biomarkers or other indicators, but not the ultimate outcome of mortality or disease progression. This can create uncertainty if such surrogate outcomes need to be related to final or 'hard' outcomes (such as overall survival) if this relationship has not been previously established (Marsden and Towse, 2017).

  - *Durability of long-term effects.* Given the relatively short timeframes of many clinical trials, it is difficult to predict long-term impact with a high degree of certainty. For non-curative treatments that need on-going administration, uncertainty arises from whether the health gains from future applications will be the same as those experienced in the trials. The extrapolation of short-term outcomes (including survival) is likely to be highly sensitive to the specifics of the HTA methodology used to model them.

  - *Extrapolation to wider populations.* Clinical trials of treatments for rare diseases are very small, hence there is uncertainty about whether and to what extent health gains can be generalised to a larger population. Under discussion is whether real-world evidence— including patient-reported outcomes—should be collected and considered as part of HTA assessments (Marsden and Towse, 2017).

- **When the choice of comparators is a challenge**. In measuring the cost-effectiveness of new interventions, HTA agencies seek to compare the new technology to existing, 'standard care'. This is challenging when interventions are developed for areas that have no existing alternatives (such as 'first-in-class' medicines), in dynamic classes where the standard of care may change several times in a short time interval, or where the 'next best' option is not obvious.

  In some cases, clinical trials may be conducted as single-arm trials (i.e. without a comparator). This may happen when the population size is small or when there are other factors (e.g. ethical concerns) that make this necessary. A number of methodologies have been developed to perform indirect comparisons based on available clinical trial data. However, their application and reliability remain contentious issues (Phillippo et al., 2018). Establishing reliable estimates of benefit becomes much more difficult without a comparator, unless the natural history of the disease is known and the patient population is homogenous (Marsden et al., 2017).

  - *Tumour agnostic treatments* (TAT). The difficulty of identifying a suitable comparator is particularly acute for TATs, an emerging form of cancer treatment that focus on the cause of cancer within a given patient rather than on the location of cancer within the body.

- TATs are studied using '*basket'* trials, which test the effect of one treatment on a single mutation in a variety of tumour types at the same time. This is in contrast to *umbrella* trials, which have multiple treatment arms within one trial. While such a trial design can be more efficient – relative to traditional clinical trials – as allowing new treatments to be tested and approved by regulatory bodies, the lack of a comparator arm and randomisation is a challenge for HTA evaluations (ISPOR, 2019; Cooper et al., 2020).

- Assessing ICERs associated with such treatments is expected to be challenging with current HTA approaches, as endpoints used in trials (such as the rate and duration of response), may lead to further uncertainty in the cost-effectiveness estimates (Cooper et al., 2020).

In October 2019, the TAT drug larotrectinib was not recommended by the CADTH pan-Canadian Oncology Drug review. The rationale behind this decision - uncertainty regarding the effectiveness of the treatment compared to available treatment options – reflects the challenges noted above (CADTH pan-Canadian Oncology Drug Review (pCODR), 2019).

We recognise that uncertainty is an inherent problem to most decision making around new interventions, therefore its presence should not stop committees assessing and making judgements around their value. However, a high degree of uncertainty can lead to the rejection of potentially valuable interventions. The likelihood of this happening will depend on the attitude of relevant decision-makers to uncertainty. (Barnsley et al., 2016).

- **Capturing full value**. One of the most challenging aspects of the HTA process is capturing the value of interventions that differ in important ways from existing technology.

  - *Value beyond the QALY*: Therapies that are potentially curative - especially where no treatment currently exists - generate value in ways not traditionally considered in HTA, most notably:

    - 'Value of hope', interventions that are potentially curative may hold more value for patients than interventions that require long-term management, even if the number of expected QALYs is the same over the same time horizon (Garrison et al., 2019).

    - 'Insurance value', where value is measured by willingness-to-pay, risk-averse individuals may be willing to pay a premium for the existence of interventions for ultra-rare conditions that are potentially catastrophic for both health and financial security (Garrison et al., 2019).

Published research on defining and measuring value elements not captured by the QALY have largely focused on aspects related to the burden of illness (usually expressed as 'severity') of the condition targeted by a new therapy (Reckers-Droog, van Exel and Brouwer, 2018), and productivity effects, including patient's or carers' ability to return to work as a result of using the new therapy an (Dixon and Round, 2019).

Some of these factors, such as severity, could be incorporated into the QALY framework by introducing 'weights' based on societal preferences and (as applied by some countries) adjusting the threshold accordingly. Other factors, such as productivity, points to the adoption of a societal perspective in evaluations (i.e. including interventions effects outside the health sector) which is applied only by a limited number of countries and it is still debated by health economists (Neumann and Kamal-Bahl, 2017). It has been pointed out that the consideration of new value elements might

have consequences on the value attributed to the QALY itself in systems with a fixed budget (or at least operating under this assumption) (Garrison et al., 2019). **These perspectives on value raise important questions about whether and how such tangential benefits, important to patients but not traditionally part of HTA, should enter into ICER calculations.**

▪ **Discounting rates for long-term therapies***.* A treatment that has lifelong benefits but also incurs a large upfront cost may be deemed not cost-effective if its costs and benefits are discounted at the same rate—for example, as in the NICE reference case, where both costs and effects are discounted at the rate of 3.5% per year. Whether very long-term benefits should be discounted at a lower rate than costs in such cases is recognised as an important question, but there is yet no consensus on how best to do this within existing HTA methodologies.

▪ **Separating component values.** Technologies that are used in combination present significant challenges for determining the value of each component. For combination therapies in oncology, for example, it is often not possible to generate discrete evidence on the relative effect of each treatment component on the overall outcomes. At present, HTA bodies generally consider combination therapies as one intervention, an approach that may lead to complications for pricing and reimbursement, particularly when the components are manufactured by different companies (Dankó, Blay and Garrison, 2019).

Companion diagnostic technologies present a similar dilemma. The use of a diagnostic technology may significantly enhance the value of a related treatment or even determine its success entirely. However, the value of the diagnostic has not been consistently accounted for in HTA systems. Although companies developing drugs may be motivated to invest in tests that contribute to the targeted use of their products, too few effective incentives are currently in place to encourage diagnostics firms to conduct the substantial independent research that can be used in HTA (Towse and Garrison, 2013). We note that in some HTA systems, this has been recognised and dedicated processes for diagnostics (like that created by NICE[6]) or for technologies used in conjunction (the so-called co-dependent technologies defined by PBAC[7]) have been established.

▪ **Value by indication.** Many technologies are approved for more than one indication. In some instances, manufacturers are aware of this potential during the R&D process. The initial approval may be followed-within a relatively short time-by submissions for approval in additional indications. In other cases, clinical use after the medicine reaches the market may lead to important indications that were not previously anticipated. Setting a single price (as suggested in the current Canadian draft guideline) for such technologies may result in inefficiencies in health systems. This has potential implications for manufacturer incentives and patient access. If the manufacturer anticipates that a product has the potential to be launched in different indications with different values, it may choose to launch only in the high-value indications. This may hinder access for patients in the lower-value (but still cost-effective) indications.

Australia has implemented a form of indication-based pricing, where medicines with multiple indications can have an indication-specific price linked to the cost-effectiveness of its different uses. The 'published price' that appears in the pharmaceutical benefits scheme (PBS) is the result of a weighted average of medicine utilisation for each indication[8].

Many payers recognise that the ICER or the value of a product can vary – sometimes substantially – by indication. It has also been shown – at least in principle – that pricing according to value in different uses has benefits for patient access and R&D incentives. However,

---

[6] The NICE Diagnostic Guidance Programme https://www.nice.org.uk/guidance/published?type=dg
[7] *https://pbac.pbs.gov.au/product-type-4-codependent-technologies.html*
[8] *http://www.pbs.gov.au/info/industry/pricing/pbs-items/fact-sheet-setting-an-approved-ex-manufacturer-price*

the application of indication-based pricing (such as that used in Australia) remains a topic of debate (Cole, Towse and Zamora, 2019).

The last two issues raised (separation of value and value by indication) are more relevant to price setting and do not invalidate either CEA or the use of CETs. However, their effects on HTA valuations and consequent pricing on the therapeutic direction of innovation should not be underestimated. The effect of pricing decisions on incentives for R&D across therapeutic areas including ones that may not have been anticipated when initial use was developed—has been documented in the literature (Cole, Towse and Zamora, 2019). **Today's HTA methods need to evolve to ensure they can appropriately assess the value of new technologies such as curative therapies, combination therapies or other major innovations that may result in high values for patients and society.**

6 Discussion and conclusions

CETs can be determined based either on (1) a demand-side WTP perspective, where the total healthcare budget is theoretically flexible, or **(2) a supply-side OC perspective** where total budgets are fixed and introducing a new technology implies displacing an existing one. The type and validity of data needed to estimate either type of threshold are difficult to generate, and current estimates rely on assumptions that do not always reflect reality. **The choice between these approaches depends on how health budgets are set and how societal preferences are captured**—directly through public input (WPT) or indirectly by assuming policymakers accurately reflect societal preferences (OC). Whether viewed from the demand side or the supply side, using the 'wrong' threshold can lead to an *inefficient use of resources*.

Although the UK and, more recently, Canada, have publicly embraced the concept of OC threshold in their HTAs, in practice the CET values applied for decision making differs from the empirical estimates based on that concept. The values shown in Table 2 representing 'empirical estimates' of the OC in the country's healthcare systems, are lower than the figures in Table 1, which represent the 'decision thresholds'. The consequences of this discrepancy remain to be researched, but they depend crucially on the extent to which the estimated CETs represent (as they are intended to) OCs and the marginal productivity of, say, the last one million pounds invested.

Countries that set a single national figure as a threshold are the exception rather than the rule. Instead, **most countries use a combination of approaches that emphasize therapeutic and social value** while remaining within healthcare budget constraints (Angelis, Lange and Kanavos, 2018). NICE in the UK has used a national (explicit) threshold, expressed as a range rather than a single figure, for the past two decades. However, this is the exception among high-income countries, and even NICE's upper bound is somewhat permeable. Moreover, NICE's threshold is primarily designed to support recommendations for funding in the NHS, not for price setting.

Expectations about CETs generally assume, unrealistically, that all decision makers will act based on 'objective' guides such as maximisation of QALYs. Although this may be possible within a small group of decision makers at the national level, implementation at the regional or local level is highly likely to consider factors not part of the standard QALY framework, or to find alternative strategies to accommodate new approvals (such as increase efficiency). Although the optimal solution would be to encourage appropriate local implementation and alignment between national and local criteria of decision making, in the short term it is important to recognise the existing divergence.

In practice, in most countries (QALY and non-QALY driven) CETs or other benchmarks are flexible to reflect differences in therapeutic areas. The underlying **intent is to echo the relative value that society assigns to treatments** for various diseases. For example, society may be willing to pay more

for treating immediately life-threatening diseases such as cancer, or for curative therapies, than for treating a self-limiting or chronic disease that do not seriously compromise daily living. This is important enough that a **growing number of public preference studies are being undertaken to inform decisions.** There is still, however, inconsistency in the results, with significant heterogeneity on preferences across a number of other factors or priority setting criteria in the conducted studies. This is an area where more evidence is needed to appropriately inform decision making.

Basing CETs on QALYs appears to be a clean and 'fair' approach to choosing among alternatives. Estimating health gains, however, can be challenging and involves assumptions that affect the accuracy and reliability of such exercises. In addition, **QALYs alone may not encompass all important aspects of value that a therapy may offer.**

The implications of creating single-figure, national CETs might affect access to new technologies in undesirable ways.

To move the policy debate forward and support HTA agencies in their decision making, we recommend:

- Systems that are considering **moving to an explicit CET should do so in stages**. For example, start by adopting the 'latent' or 'empirical' CET, which is the threshold observable based on past decisions. This approach, recommended in Badano, John and Junghans, (2017) and implemented by NICE, has the advantage of avoiding major shocks to the reimbursement system (which can significantly affect patient access to valuable treatments) and, at the same time, improving the predictability and 'fairness' of the decision making approach. A 'soft' approach should also involve the introduction of a range, rather than a single-value CET to allow some flexibility to decision making committees. This is not the direction that Canada is headed, where a fixed CET of CAD60,000 per QALY gained, with very limited room for exception, is likely to be established. This CET was informed by weak empirical evidence under the assumption of a supply-side threshold. The suggested figure is much lower than what has been observed in past decisions, which has been as high as CAD140,000 per QALY gained for oncology treatments.

- For systems embracing the concept of supply-side thresholds and/or already applying an explicit threshold, such as the UK, key priorities should include:

  - understanding the relationship between supply-side and demand-side estimates, and exploring how the two values can converge in the long-term when the health budget can be changed and healthcare systems spend should match society's preferences for health improvements, as predicted in Danzon et al. (2018);

  - recognizing the data barriers underlying current empirical estimates and being receptive to new approaches for defining the concept of a threshold and approaches to set its value.

    - It has been observed that the value of health gains and a 'life saved' are different in other government departments such as transport (see Public Health England, 2016). There is a need for a new framework that can help ensure the **efficient resource allocation of government expenditures across functional areas**. The consequences of continuing to apply different metrics and different benchmarks for decision making across public sectors deserve close examination and policy discussion.

    - Little work has been done to assess how CET levels affect the share of the total value captured by each stakeholder (i.e. payers, patients, developers etc.) relative to the use of the new technology in both the short- and the long-term. CET levels

can define the cut-off point for the reimbursement decision. They determine the maximum price at which the new technology is considered cost-effective and thus eligible for reimbursement. However, the final effective price for the new technology can differ from (and be lower than) this maximum cost-effective price in response to short-term factors (e.g. budget caps, price volume agreements, rebates) or mid- and long-term factors (e.g., in-class competition). Economic models on the functioning of pharmaceutical markets can help identify optimal levels of CETs and to improve understanding on the extent to which a single, explicit CET is the best option for decision makers (Berdud, Ferraro and Towse, Forthcoming).

The QALY as the sole measure of additional benefit may not capture all the dimensions that matter to patients and society. In addition, new innovative therapies may:

- have only immature evidence at the time of launch, producing a high level of uncertainty that makes the selection of the most 'likely' ICER a complex process, and the value assessment a dynamic process (not a one-off exercise that sets the price for the entire patent life of an intervention);

- be used in patient populations with different characteristics and potentially different conditions, which will require multiple ICERs;

- be used in combination with other technologies (such as diagnostics or other medicines), which again may lead to different value and ICERs depending on whether the use is in combination or alone.

In conclusion, a rigid, single, QALY-based CET (a main factor proposed by the PMPRB in Canada) may not lead to efficient resource allocation. Multiple thresholds reflecting value elements beyond the QALY may be needed. More research is needed to understand how new spending is accommodated in the system, and whether society is willing to trade off some health gains for other benefits valued by patients and society.

References

Andrews, M., Elamin, O., Hall, A.R., Kyriakoulis, K. and Sutton, M., 2017. Inference in the presence of redundant moment conditions and the impact of government health expenditure on health outcomes in England. *Econometric Reviews*, 36(1–3), pp.23–41. 10.1080/07474938.2016.1114205.

Angelis, A., Lange, A. and Kanavos, P., 2018. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *The European Journal of Health Economics*, 19(1), pp.123–152. 10.1007/s10198-017-0871-0.

van Baal, P., Morton, A., Meltzer, D. and Brouwer, W., 2019. Future unrelated medical costs need to be considered in cost effectiveness analysis. *The European Journal of Health Economics*, 20(1), pp.1–5. 10.1007/s10198-018-0976-0.

Badano, G., John, S. and Junghans, T., 2017. NICE's Cost-Effectiveness Threshold. p.19.

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. 10.1017/S1744133111000181.

Barnsley, P., Cubi-Molla, P., Fischer, A. and Towse, A., 2016. Uncertainty and Risk in HTA Decision Making. *OHE Research Paper 16/05*, p.32.

Berdud, M., Ferraro, J. and Towse, A., Forthcoming. *A theory on ICER pricing and optimal level of cost-effectiveness threshold*. OHE consulting Report.

Bertram, M.Y., Lauer, J.A., De Joncheere, K., Edejer, T., Hutubessy, R., Kieny, M.-P. and Hill, S.R., 2016. Cost–effectiveness thresholds: pros and cons. *Bulletin of the World Health Organization*, 94(12), pp.925–930. 10.2471/BLT.15.164418.

Bokhari, F.A.S., Gai, Y. and Gottret, P., 2007. Government health expenditures and health outcomes. *Health Economics*, 16(3), pp.257–273. 10.1002/hec.1157.

CADTH pan-Canadian Oncology Drug Review (pCODR), 2019. *pCODR Expert Review Committee (pERC)*.

CECI, 2019. *Canada GDP per capita*. [online] Available at: https://www.ceicdata.com/en/indicator/canada/gdp-per-capita [Accessed 11 Dec. 2019].

Claxton, K., Lomas, J. and Martin, S., 2017. *Estimating Expected Health Opportunity Costs in the NHS (Analysis of 2012/13 Expenditure Data)*. Available at: https://www.york.ac.uk/media/che/documents/Estimating_expected_health_opportunity_costs_in_the_NHS_201213.pdf .

Claxton, K., Lomas, J. and Martin, S., 2018. The impact of NHS expenditure on health outcomes in England: Alternative approaches to identification in all-cause and disease specific models of mortality. *Health Economics*, 27(6), pp.1017–1023. 10.1002/hec.3650.

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*, 19(14), pp.1–504. 10.3310/hta19140.

Cleemput, I., Neyt, M., Thiry, N., De Laet, C. and Leys, M., 2011. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *International Journal of Technology Assessment in Health Care*, 27(1), pp.71–76. 10.1017/S0266462310001194.

Cole, A., Towse, A. and Zamora, B., 2019. *Indication-Based Pricing (IBP) Discussion Paper*. Discussion Papers. [online] Office of Health Economics, p.17. Available at: https://www.ohe.org/system/files/private/publications/Rebrand%20IBP%20discussion%20paper_Final%20%2830%20September%202019%29.pdf [Accessed 22 Aug. 2019].

Cooper, S., Bouvy, J.C., Baker, L., Maignen, F., Jonsson, P., Clark, P., Palmer, S., Boysen, M. and Crabb, N., 2020. How should we assess the clinical and cost effectiveness of histology independent cancer drugs? *BMJ*, p.l6435. 10.1136/bmj.l6435.

OFFICE OF HEALTH ECONOMICS
CONSULTING

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P. and Parkin, D., 2015. The Influence of Cost-Effectiveness and Other Factors on Nice Decisions: FACTORS INFLUENCING NICE DECISIONS. *Health Economics*, 24(10), pp.1256–1271. 10.1002/hec.3086.

Dankó, D., Blay, J.-Y. and Garrison, L.P., 2019. Challenges in the value assessment, pricing and funding of targeted combination therapies in oncology. *Health Policy*, 123(12), pp.1230–1236. 10.1016/j.healthpol.2019.07.009.

Danzon, P., Towse, A. and Mestre-Ferrandiz, J., 2015. Value-Based Differential Pricing: Efficient Prices for Drugs in a Global Context: VALUE-BASED DIFFERENTIAL PRICING. *Health Economics*, 24(3), pp.294–301. 10.1002/hec.3021.

Danzon, P.M., Drummond, M.F., Towse, A. and Pauly, M.V., 2018. Objectives, Budgets, Thresholds, and Opportunity Costs—A Health Economics Approach: An ISPOR Special Task Force Report [4]. *Value in Health*, 21(2), pp.140–145. 10.1016/j.jval.2017.12.008.

Deloitte, 2012. *Issue Brief: Value-based pricing for pharmaceuticals: Implications of the shift from volume to value*. [online] Available at: https://www2.deloitte.com/us/en/pages/consulting/articles/value-based-pricing-for-pharmaceuticals.html .

Devlin, N. and Parkin, D., 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics*, 13(5), pp.437–452. 10/bf8kdj.

Dixon, P. and Round, J., 2019. Caring for Carers: Positive and Normative Challenges for Future Research on Carer Spillover Effects in Economic Evaluation. *Value in Health*, 22(5), pp.549–554. 10.1016/j.jval.2018.10.010.

Eddama, O. and Coast, J., 2009. Use of economic evaluation in local health care decision-making in England: A qualitative investigation. *Health Policy*, 89(3), pp.261–270. 10.1016/j.healthpol.2008.06.004.

Edney, L.C., Haji Ali Afzali, H., Cheng, T.C. and Karnon, J., 2018a. Estimating the Reference Incremental Cost-Effectiveness Ratio for the Australian Health System. *PharmacoEconomics*, 36(2), pp.239–252. 10.1007/s40273-017-0585-2.

Edney, L.C., Haji Ali Afzali, H., Cheng, T.C. and Karnon, J., 2018b. Mortality reductions from marginal increases in public spending on health. *Health Policy*, 122(8), pp.892–899. 10.1016/j.healthpol.2018.04.011.

Eichler, H.-G., Kong, S.X., Gerth, W.C., Mavros, P. and Jönsson, B., 2004. Use of Cost-Effectiveness Analysis in Health-Care Resource Allocation Decision-Making: How Are Cost-Effectiveness Thresholds Expected to Emerge? *Value in Health*, 7(5), pp.518–528. 10.1111/j.1524-4733.2004.75003.x.

European Federation of Pharmaceutical Industries and Associations, 2018. *Pipeline Review of innovative Therapies*.

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. 10.18553/jmcp.2019.18378.

Hampson, G., Towse, A. and Henshall, C., 2017. *How can HTA meet the needs of health system and government decision makers?* Briefings. [online] Office of Health Economics. Available at: https://ideas.repec.org/p/ohe/briefg/001836.html [Accessed 17 Jul. 2018].

Hernandez-Villafuerte, K., Zamora, B., Feng, Y., Parkin, D., Devlin, N., and Towse, A. (2019). Exploring Variations in the Opportunity Cost Cos tEffectiveness Threshold by Clinical Area: Results from a Feasibility Study in England. OHE Research Paper, London: Office of Health Economics. Available at: https://www.ohe.org/publications/exploring-variations-opportunity-cost-costeffectiveness-threshold-clinical-are

Hernandez-Villafuerte, K., Zamora, B. & Towse, A. (2018). R&D, Issues Surrounding the Estimation of the Opportunity Cost of Adopting a New Health Care Technology: Areas For Further Research. OHE Research Paper 18/07, London: Office of Health Economics. Available at: https://www.ohe.org/publications/issues-surroundingestimation-opportunity-cost-adopting-new-health-care-technology

Institute for Quality and Efficiency in Health Care, 2009. *General Methods for the Assessment of the Relation of Benefits to Costs*. [online] Available at: https://www.iqwig.de/download/General_Methods_for_the_Assessment_of_the_Relation_of_Benefits_to_Costs.pdf .

ISPOR, 2019. The Leading Global Conference for Health Economics and Outcomes Research. *ISPOR*.

Karlsberg Schaffer, S., Sussex, J., Devlin, N. and Walker, A., 2015. Local health care expenditure plans and their opportunity costs. *Health Policy*, 119(9), pp.1237–1244. 10.1016/j.healthpol.2015.07.007.

Karlsberg Schaffer, S., Sussex, J., Hughes, D. and Devlin, N., 2016. Opportunity costs and local health service spending decisions: a qualitative study from Wales. *BMC Health Services Research; London*, [online] 16. http://0-dx.doi.org.wam.city.ac.uk/10.1186/s12913-016-1354-1.

Kobelt, G., 2013. *Health Economics: An Introduction to Economic Evaluation*. [online] Office of Health Economics. Available at: https://ideas.repec.org/b/ohe/monogr/000004.html [Accessed 17 Jul. 2018].

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*, 22(9), pp.995–1002. 10.1016/j.jval.2019.04.1926.

Long, G., 2017. The Biopharmaceutical Pipeline: Innovative Therapies in Clinical Development. *Boston: Analysis Group Inc*, (Produced for Pharmaceuticals and Manufacturers of America).

Marsden, G. and Towse, A., 2017. *Exploring the Assessment and Appraisal of Regenerative Medicines and Cell Therapy Products: Is the NICE Approach Fit for Purpose?* Consulting Reports. [online] Office of Health Economics. Available at: https://ideas.repec.org/p/ohe/conrep/001802.html [Accessed 17 Jul. 2018].

Marsden, G., Towse, A., Pearson, S.D., Dreitlein, B. and Henshall, C., 2017. *Gene Therapy: Understanding the Science, Assessing the Evidence, and Paying for Value*. Research Papers. [online] Office of Health Economics. Available at: https://ideas.repec.org/p/ohe/respap/001811.html [Accessed 17 Jul. 2018].

Mason, H., Baker, R. and Donaldson, C., 2008. Willingness to pay for a QALY: past, present and future. *Expert Review of Pharmacoeconomics & Outcomes Research*, 8(6), pp.575–582. 10.1586/14737167.8.6.575.

National Institute for Health and Care Excellence, 2018. *Guide to the processes of technology appraisal*. [online] NICE. Available at: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/process .

Neumann, P.J. and Kamal-Bahl, S., 2017. Drugs and Medical Innovation. Should Value Frameworks Take A 'Societal Perspective'? *Health Affairs Blog*, p.9. 10.1377/hblog20170906.061833.

Nord, E., 2005. Concerns for the worse off: fair innings versus severity. *Social Science & Medicine*, 60(2), pp.257–263. 10.1016/j.socscimed.2004.05.003.

Ochalek, J., Lomas, J. and Claxton, K., 2018. *Assessing health opportunity costs for the Canadian health care systems*.

Ottersen T, Førde R, Kakad M, Kjellevold A, Melberg H, Ringard Å and Norheim, 2016. A new proposal for priority setting in Norway: Open and fair. [online] 120(3). Available at: https://www.sciencedirect.com/science/article/pii/S0168851016000269?via%3Dihub .

Patented Medicine Prices Review Board, 2019a. Working Group to Inform the Patented Medicine Prices Review Board (PMPRB) Steering Committee on Modernization of Price Review Process Guidelines. [online] Available at: http://www.pmprb-cepmb.gc.ca/view.asp?ccid=1449 [Accessed 11 Dec. 2019].

Patented Medicines Prices Review Board, 2019b. *PMRPB Guidelines 2019*. [online] Ottawa: PMRPB. Available at: https://www.canada.ca/content/dam/pmprb-cepmb/documents/consultations/draft-guidelines/draft-guidelines-en.pdf [Accessed 11 Dec. 2019].

Phillippo, D.M., Ades, A.E., Dias, S., Palmer, S., Abrams, K.R. and Welton, N.J., 2018. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Medical Decision Making*, 38(2), pp.200–211. 10.1177/0272989X17725740.

Public Health England, 2016. *NHS screening Programs in England*.

Raftery, J.P., 2014. NICE's Cost-Effectiveness Range: Should it be Lowered? *PharmacoEconomics*, 32(7), pp.613–615. 10.1007/s40273-014-0158-6.

Reckers-Droog, V.T., van Exel, N.J.A. and Brouwer, W.B.F., 2018. Looking back and moving forward: On the application of proportional shortfall in healthcare priority setting in the Netherlands. *Health Policy*, 122(6), pp.621–629. 10.1016/j.healthpol.2018.04.001.

Robinson, L.A., Hammitt, J.K., Chang, A.Y. and Resch, S., 2017. Understanding and improving the one and three times GDP per capita cost-effectiveness thresholds. *Health Policy and Planning*, 32(1), pp.141–145. 10.1093/heapol/czw096.

Ryen, L. and Svensson, M., 2015. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature: THE WILLINGNESS TO PAY FOR A QALY. *Health Economics*, 24(10), pp.1289–1301. 10.1002/hec.3085.

Sachs, J. ed., 2001. *Macroeconomics and health: investing in health for economic development ; report of the Commission on Macroeconomics and Health*. Geneva: World Health Organization.

Schwarzer, R., Rochau, U., Saverno, K., Jahn, B., Bornschein, B., Muehlberger, N., Flatscher-Thoeni, M., Schnell-Inderst, P., Sroczynski, G., Lackner, M., Schall, I., Hebborn, A., Pugner, K., Fehervary, A., Brixner, D. and Siebert, U., 2015. Systematic overview of cost–effectiveness thresholds in ten countries across four continents. *Journal of Comparative Effectiveness Research*, 4(5), pp.485–504. 10.2217/cer.15.38.

Siverskog, J. and Henriksson, M., 2019. Estimating the marginal cost of a life year in Sweden's public healthcare sector. *The European Journal of Health Economics*, 20(5), pp.751–762. 10.1007/s10198-019-01039-0.

Skedgel, C., Wranik, D. and Hu, M., 2018. The Relative Importance of Clinical, Economic, Patient Values and Feasibility Criteria in Cancer Drug Reimbursement in Canada: A Revealed Preferences Analysis of Recommendations of the Pan-Canadian Oncology Drug Review 2011–2017. *PharmacoEconomics*, 36(4), pp.467–475. 10.1007/s40273-018-0610-0.

Smith, R.D. and Richardson, J., 2005. Can we estimate the 'social' value of a QALY? *Health Policy*, 74(1), pp.77–84. 10.1016/j.healthpol.2004.12.009.

Stadhouders, N., Koolman, X., Dijk, C., Jeurissen, P. and Adang, E., 2019. The marginal benefits of healthcare spending in the Netherlands: Estimating cost-effectiveness thresholds using a translog production function. *Health Economics*, 28(11), pp.1331–1344. 10.1002/hec.3946.

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. 10.1007/s40273-017-0606-1.

Towse, A. and Garrison, L.P., 2013. Economic Incentives for Evidence Generation: Promoting an Efficient Path to Personalized Medicine. *Value in Health*, 16(6), pp.S39–S43. 10.1016/j.jval.2013.06.003.

Towse, A., Pritchard, C., Devlin, N.J., Kings Fund and Office of Health Economics eds., 2002. *Cost-effectiveness thresholds: economic and ethical issues*. London.

Vallejo-Torres, L. et al., 2016. On the estimation of the Cost-Effectiveness Threshold: Why, What, How? *Value in Health*, 19, pp.558–566.

Vallejo-Torres, L., García-Lorenzo, B. and Serrano-Aguilar, P., 2018. Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Economics*, 27(4), pp.746–761. 10.1002/hec.3633.

Viscusi, A., 2003. The Value of Statistical Life: A critical review of market estimates throuhout the world. *Working paper 9487*.

Wailoo, A., Tsuchiya, A. and McCabe, C., 2009. Weighting Must Wait: Incorporating Equity Concerns into Cost-Effectiveness Analysis May Take Longer than Expected. *PharmacoEconomics*, 27(12), pp.983–989. 10.2165/11314100-000000000-00000.

Wang, S., Gum, D. and Merlin, T., 2018. Comparing the ICERs in Medicine Reimbursement Submissions to NICE and PBAC—Does the Presence of an Explicit Threshold Affect the ICER Proposed? *Value in Health*, 21(8), pp.938–943. 10.1016/j.jval.2018.01.017.

Woods, B., Revill, P., Sculpher, M. and Claxton, K., 2016. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. *Value in Health*, 19(8), pp.929–935. 10.1016/j.jval.2016.02.017.

Wubulihasimu, P., Gheorghe, M., Slobbe, L., Polder, J. and van Baal, P., 2015. Trends in Dutch hospital spending by age and disease 1994–2010. *Health Policy*, 119(3), pp.316–323. 10.1016/j.healthpol.2014.11.009.

Zhang, N., Ji, L.-J. and Li, Y., 2017. Cultural Differences in Opportunity Cost Consideration. *Frontiers in Psychology*, [online] 8. 10.3389/fpsyg.2017.00045.

Appendix

A.1 What is the QALY?

Quality-Adjusted Life Years (QALYs)

Quality-Adjusted Life Years (QALYs) are used in health economics to measure the benefit of different treatments across all (or at least most) possible disease areas.

QALYs are calculated as the length of life weighted by health-related quality-of-life (also called 'utility' or 'tariff'):

QALY = length of survival x utility

One year of life lived in the best imaginable health state (full health) has a QALY 'value' of 1.0.

A reduction in either health-related quality of life ('utility') or the length of life reduces the QALY value:

- One year lived in a state that is only 75% as good as perfect health has a QALY value of 0.75
- Nine months of a year lived in perfect health also has a value of 0.75

As an illustration: imagine the quality of life for an individual can be captured by a number between 0 and 1, where 1 equals full health and 0 equals death. Let's assume that a patient has been living with a chronic disease for four years with a quality of life of 0.6 (i.e. at 60% of full health). Without treatment, the patient's quality of life for ten years would be that represented in blue in the graph below. If we apply the formula above, the patient would live a total of 14 years which is equivalent to 8.4 QALYs (8.4 years lived in full health). If the patient starts treatment in year 4, the quality of life for the patient from year 4 onwards is represented by the green area. With treatment, the patient's quality of life improves and life expectancy increases by a year. Note that the treatment provides the patient with 1.6 additional QALYs, which is the area represented in green.
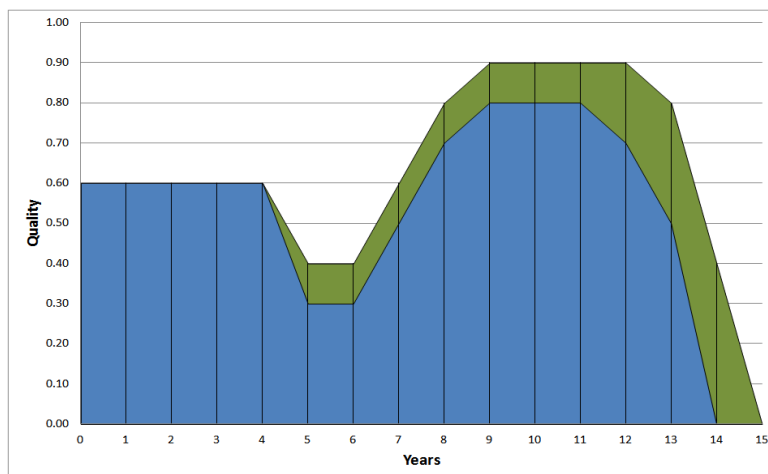
A.2 Method of sections 2, 3, and 4

OHE first established a selection of the relevant themes that were required to explore the focus of the paper. The following themes were selected:

- HTA vs C/E threshold – is there any need?

- Conceptual approach to the C/E threshold
    a. Concept
    b. OCs
    c. League Table
    d. WTP
    e. GDP approach (WHO)
    f. Other alternatives
    g. Discussion of the several methods

- Moving from theory to practice – estimates
    a. OCs
    b. League table
    c. WTP
    d. GDP
    e. Other alternatives
    f. Discussion of the several estimates of thresholds

- The use of threshold in healthcare decision making
    a. Australia
    b. Canada
    c. UK
    d. Other
    e. Discussion/Comparison

Once the themes were agreed, OHE brought together a collection of papers relevant to cost-effectiveness thresholds based on past work as well as additional conversations with colleagues with expertise in the area. This formed the initial pool of papers. The OHE team explored the papers in the collection and identified the themes that were not well reported. Complementary (targeted) searches were implemented to fill the theme gaps and include most recently published research and policy reports. Our approach for the literature review is described in the diagram in Figure A1.

OHE

Figure A1. Literature review method

```
┌─────────────────────┐
│   Four themes       │
│   selection         │
└──────────┬──────────┘
           │
           ▼
┌─────────────────────┐
│  OHE's paper        │
│  collection on      │
│  threshold          │
│  (N = 84)           │
└──────────┬──────────┘
           │
  ┌────────┼────────────────────┐
  ▼        ▼                     ▼
┌──────────┐ ┌──────────────┐ ┌──────────────┐
│Core      │ │Supplementary │ │Excluded      │
│papers    │ │papers        │ │Papers        │
│(N=18)    │ │(N=20)        │ │(N=46)        │
└────┬─────┘ └──────┬───────┘ └──────────────┘
     │              │
     ▼              ▼
┌──────────┐ ┌──────────────────┐
│Core      │ │Theme-specific    │
│papers    │ │PubMed search     │
│allocated │ │(N=1)             │
│to themes │ │Snowball citations│
│(N=18)    │ │(N=12)            │
└────┬─────┘ │Targeted searches │
     │       │(N=9)             │
     │       └──────┬───────────┘
     ▼              ▼
┌──────────┐ ┌──────────────────┐
│Final list│ │Final list of     │
│of Core   │ │Supplementary     │
│papers    │ │papers            │
│(N=18)    │ │(N=42)            │
└────┬─────┘ └──────┬───────────┘
     │              │
     └──────┬───────┘
            ▼
   ┌──────────────────┐
   │Total number of   │
   │papers reviewed   │
   │(N=60)            │
   └──────────────────┘
```
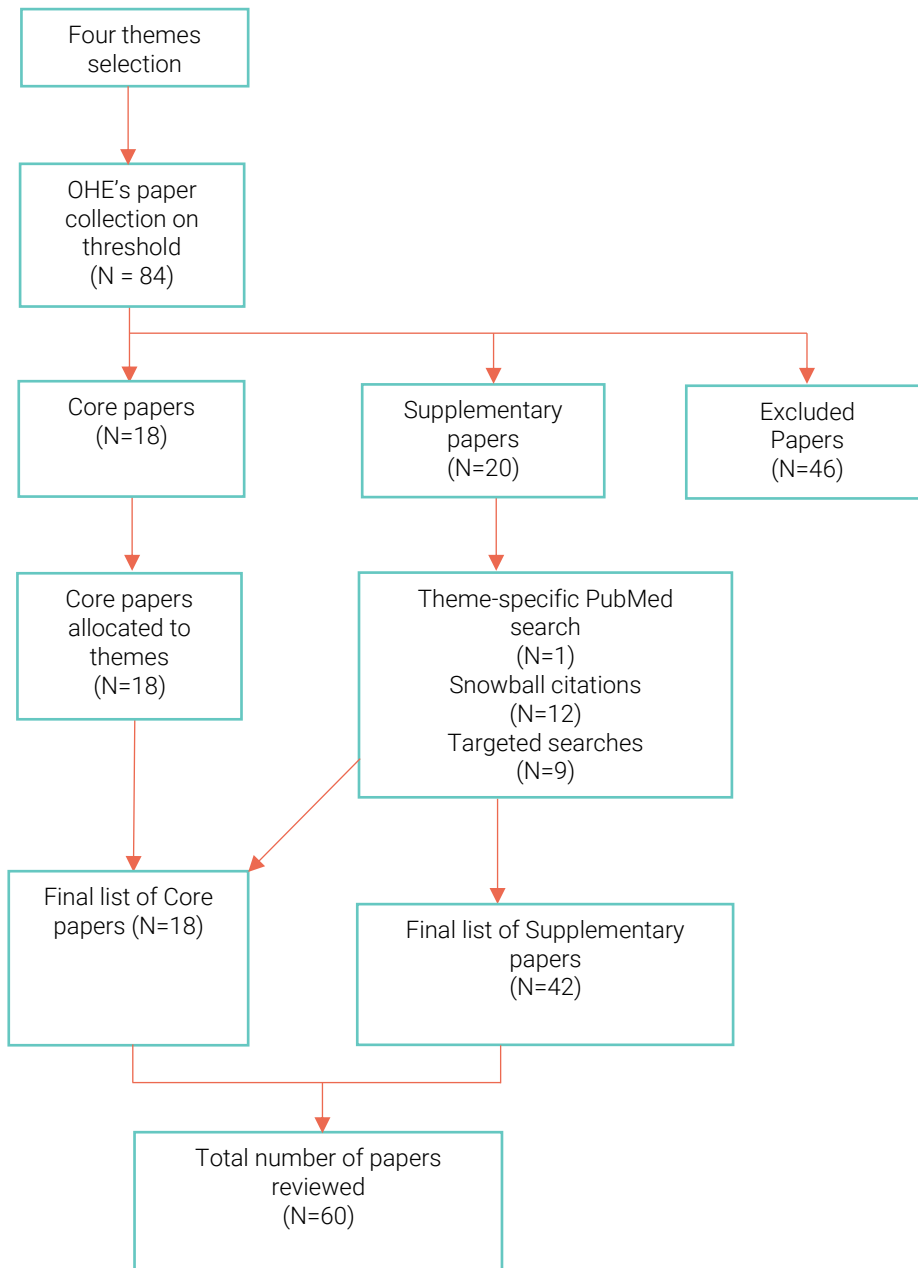
A.3 Methods for section 5

To understand trends in health technology innovation, OHE consulted two key documents from PhRMA (Long, 2017) and EFPIA (European Federation of Pharmaceutical Industries and Associations, 2018) regarding the innovation pipeline. The selection of these two reports was based on the assumption that trade associations commission and produce the current analyses on the biopharmaceutical industry research & development (R&D) pipeline. An alternative approach could be to obtain IQVIA pipeline data or clinical trials published information to compile a long list of interventions expected to come to market in the near future. However, deriving what can be defined as "innovative" based solely on the name of molecules was likely to be a resource-intensive exercise with limited added benefits compared to using analyses already available in the public domain.

It is noted that while these studies highlight a number of technologies that are currently in development and expected to gain prominence – over the next 5 to 10 years – there does not appear to be a systematic manner by which they are categorised. For example, the EFPIA pipeline review identifies both lists individual technologies and promising medicines and disease areas for which are large number of innovations are in the pipeline. It is difficult, therefore, to select a particular methodology to group innovations – e.g. by type of technology or by disease area – and it is inevitable that there will be areas of overlap. Our initial attempt to create a list of innovative technologies is available in a supplementary Excel file (available on request).

Based on the initial list of expected innovations, we identified relevant literature from Value in Health issues and from OHE publication series. These were reviewed to identify the most prominent issues faced by the HTA process as it currently stands. This was supplemented by a search through ISPOR panels over the past two years as a means of ensuring we captured the latest insights on the debate. The issues identified were mapped to the areas or technologies listed in the Excel file.

A.4 Most recent country OC estimates

(for countries not included in this paper)

A study in Spain (Vallejo-Torres, García-Lorenzo and Serrano-Aguilar, 2018) explore the effect of changes in the national healthcare budget on health outcomes in 17 regional health services. A fixed-effect model with an IV approach is used. The authors measure health outcomes in terms of quality-adjusted life expectance (which is closer to the notion of QALY than the mortality rates used in Claxton et al. (2015)). The results show a **CET in the range of €22,000 to €25,000 in Spain**.

Siverskog and Henriksson (2019) estimate **the CET applicable to Sweden** using data on remaining life expectancy. The authors estimate a threshold in terms of cost-per-life year. Using some health-related quality of life adjustments, the threshold **may be as high as €45,000 per QALY gained**.

Two recent studies explore the estimation of a CET **in The Netherlands** following a similar quantitative approach, although using a different type of data. van Baal et al. (2019) focus on hospital spending related to cardiovascular disease (CVD), and identify a **CET of €41,000**. Stadhouders et al. (2019) estimate a CET from hospital healthcare expenditure across disease areas, providing an **estimate of €73,600** (ranging €53,000 - €94,000). The authors derive health outcomes from quality of life data, which allows the direct estimation of a cost-per-QALY threshold.

Reference	Country	Measures of Health Outcomes	Measures of expenditure / sources	Central estimate reported	Estimates range	
Vallejo-Torres, García-Lorenzo and Serrano-Aguilar (2018)	Spain	Quality-Adjusted Life Expectancy	Data for 17 regions of the SNHS over the period 2008-2012. European Commission data on growth rates for health spending per capita.	€23,158	€22,000	€25,000
Siverskog and Henriksson(2019)	Sweden	Remaining life expectancy (from Statistics Sweden)	Regional council healthcare expenditure per capita for 20 regions between 2003 and 2016	SEK370,000 (or €39K)/life year	€19,000/QALY	€45,000/ QALY

| van Baal et al. (2019) | The Netherlands | Mortality | Annual per capita hospital expenditures and mortality rates for CVD, from Cost of illness studies (Wubulihasimu et al., 2015) . | €41,000 | €25,900 | €110,400 |
| Stadhouders et al. (2019) | The Netherlands | QALYs | Annual per capita hospital expenditures and mortality rates | €73,600 | €53,000 | €94,000 |

Glossary

Cost-benefit analysis (CBA) – A comparison of alternative interventions in which costs and outcomes are quantified in common monetary units

Cost-effectiveness analysis (CEA) – A type of evaluation the produces a ratio of benefit to risk of an intervention, in a standard clinical setting, using outcomes measuring issues of importance to patients (e.g. perform daily activities, longer life, etc.)

Cost-effectiveness threshold (CET) – A monetary cut-off used by payers to determine if a health intervention provides enough gain to warrant inclusion in the healthcare system. Usually expressed as the maximum cost per QALY gained, a CET may be implicit or explicit, and applied to all interventions or only to some

Cost-utility analysis (CUA) – A form of cost-effectiveness analysis that measures therapeutic consequences in terms of their utility, usually to the patient; most often based on QALYs

Demand side cost-effectiveness threshold – assumes the health budget it not finite, but fluctuates with demand. In a single-payer system, this assumes the government knows the relative value that society would assign to all government services, not just healthcare. A threshold is set based on society's maximum willingness to pay for health, possibly expressed as a cost per QALY gained. All new technologies below this threshold would be funded

Health technology assessment (HTA) – An evidence-based, multidisciplinary process intended to support healthcare decision making by assessing the properties and effects of one or more new or existing health technologies in comparison with a current standard. It may address the direct, intended consequences of technologies as well as their indirect, unintended consequences. Its main purpose is to inform technology-related policy decisions in healthcare. HTA is conducted by interdisciplinary groups using explicit analytical frameworks drawing from a variety of methods

Incremental cost-effectiveness ratio (ICER) – The additional cost of an intervention as compared with the standard of care intervention divided by the difference in effect or patient outcome between the interventions, e.g. additional cost per QALY gained

OC – The amount that could be spent on alternative healthcare strategies if the health technology in question was not used

Quality adjusted life year (QALY) – The outcome of a treatment measured as the number of years of life saved, adjusted for their utility (quality). QALYs can provide a common unit for comparing cost-utility across different interventions and health problems

Supply side cost-effectiveness thresholds assume that the health budget is fixed, i.e. will neither increase nor decrease. OC approaches to CETs require that use of any new technology must displace one currently in use

Willingness to pay – The maximum amount that a person is willing to pay to (a) achieve a particular good health state or outcome, or to increase its probability or (b) avoid a specific bad-health state or outcome, or decrease its probability

OHE

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

ohe.org