



Supply-Side Cost-Effectiveness Thresholds: Questions for Evidence-Based Policy

Chris Sampson¹ · Bernarda Zamora² · Sam Watson³ · John Cairns⁴ · Kalipso Chalkidou² · Patricia Cubi-Molla¹ · Nancy Devlin⁵ · Borja García-Lorenzo^{6,7} · Dyfrig A. Hughes⁸ · Ashley A. Leech⁹ · Adrian Towse¹

Accepted: 29 March 2022 / Published online: 7 June 2022
© The Author(s) 2022

Abstract

There is growing interest in cost-effectiveness thresholds as a tool to inform resource allocation decisions in health care. Studies from several countries have sought to estimate health system opportunity costs, which supply-side cost-effectiveness thresholds are intended to represent. In this paper, we consider the role of empirical estimates of supply-side thresholds in policy-making. Recent studies estimate the cost per unit of health based on average displacement or outcome elasticity. We distinguish the types of point estimates reported in empirical work, including marginal productivity, average displacement, and outcome elasticity. Using this classification, we summarise the limitations of current approaches to threshold estimation in terms of theory, methods, and data. We highlight the questions that arise from alternative interpretations of thresholds and provide recommendations to policymakers seeking to use a supply-side threshold where the evidence base is emerging or incomplete. We recommend that: (1) policymakers must clearly define the scope of the application of a threshold, and the theoretical basis for empirical estimates should be consistent with that scope; (2) a process for the assessment of new evidence and for determining changes in the threshold to be applied in policy-making should be created; (3) decision-making processes should retain flexibility in the application of a threshold; and (4) policymakers should provide support for decision-makers relating to the use of thresholds and the implementation of decisions stemming from their application.

1 Introduction

New health care technologies are commonly both cost-increasing and health-improving, rather than cost-decreasing

✉ Chris Sampson
csampson@ohe.org

¹ Office of Health Economics, London, UK

² Imperial College London, London, UK

³ University of Birmingham, Birmingham, UK

⁴ London School of Hygiene and Tropical Medicine, London, UK

⁵ University of Melbourne, Melbourne, Australia

⁶ Kronikune Institute for Health Services Research, Basque Country, Spain

⁷ Assessment of Innovations and New Technologies Unit, Hospital Clínic Barcelona, University of Barcelona, Barcelona, Spain

⁸ Centre for Health Economics and Medicines Evaluation, Bangor University, Bangor, UK

⁹ Vanderbilt University, Nashville, USA

or clinically ineffective. In such cases, an incremental cost-effectiveness ratio (ICER) threshold may be used as the basis for judging whether a technology represents value for money. The use of a threshold approach may be consistent with the objective of maximising health improvements from given health care budgets if the threshold is correctly specified. Health technology assessment (HTA) processes commonly adopt the quality-adjusted life year (QALY) as a measure of health, which reflects differences in life expectancy and quality of life. A cost-effectiveness threshold (CET) can be defined in these terms. However, a threshold approach is generalisable to outcomes other than QALYs.

If we accept the notion that a monetary value (i.e. a threshold) must be attached to health outcomes—to interpret cost-effectiveness evidence and translate this evidence into investment and disinvestment decisions—then both positive and normative questions regarding the choice of threshold arise. Explicit thresholds are not commonly used in decision-making, and those that do exist have been specified without reference to evidence, perhaps based on precedents [1, 2]. More recently, applied research on the relationship between health care spending and outcomes has sought to provide a

Key Points for Decision Makers

Empirical estimates of opportunity cost may inform cost-effectiveness thresholds used in policy-making, but there are limitations in the evidence base.

Cost-effectiveness thresholds are not synonymous with opportunity costs, and adoption of any metric—such as average displacement—implies a range of assumptions about the nature of service provision and the objectives of technology assessment.

Policymakers can make appropriate use of imperfect evidence on opportunity costs by establishing transparent and flexible processes for the assessment and use of this evidence.

basis in evidence for CETs. This research has in part arisen from the recognition that CETs used in policy may not correspond to the opportunity cost of decisions, while the true opportunity cost of every decision cannot be observed. This growing body of research has created the potential to use evidence-based thresholds to guide policy decisions [3, 4].

Two fundamentally different approaches have historically been used to provide an evidence base for the selection of CETs, each based on different principles and implying alternative empirical strategies. In one approach, the threshold could be determined by the value that society places on QALY gains. This can be estimated empirically using willingness-to-pay (WTP) experiments and is often referred to as the ‘demand-side’ approach, grounded in welfare economics, whereby the ‘demanders’ have preferences over health and other goods to allocate resources within their budget constraint [5, 6]. An alternative approach determines the threshold by identifying the opportunity cost of implementing cost-increasing technologies in the presence of a fixed budget. This approach is known as a ‘supply-side’ threshold.¹ In this paper, we focus on recent evidence centred on the optimal allocation of fixed budgets, in which case supply-side thresholds provide a practical and intuitive basis for decision-making.

Internationally, policymakers require guidance on applying appropriate decision rules in HTA and resource allocation [7]. The aim of this paper is to consider how the emerging evidence on health system opportunity costs should be used to set and use a CET in policy and decision making.

¹ We note that the ‘supply-side’ and ‘demand-side’ terminology is problematic, particularly where ‘supply-side’ thresholds are used to indicate society’s demand for new technologies within a given budget constraint. However, we use them in this paper as they are readily understood by researchers.

We present a narrative review of the literature on the nature and use of supply-side thresholds and outline several recommendations for policymakers. The review and the recommendations arose from a series of discussions, first between the lead author and each co-author, and then as a group. Relevant articles were primarily identified on the basis of the authors’ collective knowledge, and supported by a snowballing strategy.

The remainder of this paper is presented in six sections. The next three sections (Sects. 2–4) introduce some basic principles and further background to the research. We first summarise the observed impact of estimates on policy, then describe how—we believe—thresholds may be ‘evidence-based’, and finally consider some different ways in which a ‘supply-side threshold’ may be characterised. After this, Section 5 considers the current evidence base with respect to theory, methods, and data. Building on this, Section 6 outlines our recommendations to policymakers, before Sect. 7 concludes.

2 What Impact have Estimates had on Policy?

The first concerted attempt to estimate a supply-side CET, without relying on cost-effectiveness estimates for individual technologies, was by Claxton et al. [8] in the context of the National Health Service (NHS) in England, which recommended a central estimate of £12,936 per QALY. The researchers developed econometric models to estimate the relationship between differences in expenditure and differences in mortality. The methods have been influential, with studies conducted in Australia, Spain, and other countries adopting similar methods [3].

The UK Department of Health and Social Care (DHSC) explicitly cites Claxton et al. [8] as the basis for recommending the opportunity cost of a QALY as £15,000 in impact assessments (see, for example, [9]). The empirical estimate for England has not resulted in the adoption of a new threshold by the National Institute for Health and Care Excellence (NICE). The 2019 voluntary scheme for branded medicines pricing and access (VPAS), agreed between the Association of the British Pharmaceutical Industry (ABPI) and DHSC, maintains the existing threshold range used by NICE at £20,000–£30,000 per QALY gained [10]. A proposal to adopt a threshold of £15,000 per QALY in decision-making about vaccines was rejected by the government [11].

In other countries for which empirical CET estimates (based on similar methods to those used for the England estimate) have been available for several years, such as Spain and Australia [12, 13], their influence has been limited because policymakers do not specify an explicit CET. In Spain, a figure of €30,000 per QALY has been widely

cited based on a review of economic evaluations, rather than on an analysis of expenditure and outcomes [14]. In 2012, the government introduced a legal stipulation stating that health technology financing decisions had to be based on scientific evidence from cost-effectiveness analysis and economic evaluations. No threshold value was adopted formally because of the perceived lack of a theoretical and scientific basis, and controversy around its estimation and what the threshold should represent. In this context, the Ministry of Health commissioned the Spanish Network of HTA agencies to prepare methodological reports on the definition and estimation of Spain's CET [15–17]. The Ministry of Health has not formally adopted the published estimate of €22,000–€25,000 per QALY [12]. In Australia, the Pharmaceutical Benefits Advisory Committee (PBAC) is believed to apply a threshold range of \$45,000 to \$60,000 per QALY [18], but this is not officially stated, and is not based on the published empirical estimate of \$28,033 [13].

For most countries, empirical estimates from country-specific data are not available. However, researchers have used estimates from England to generate threshold estimates for other countries [19]. This approach has had some limited influence in Canada, where the Patented Medicines Prices Review Board (PMPRB) recommended a CET of \$60,000 per QALY, which was twice the \$30,000 evidence-based recommendation [20, 21], citing earlier work by Woods et al. [19] that applied estimates from Claxton et al. [8] to the Canadian context.

Published CET estimates have been adopted in cost-effectiveness research [22] and have influenced debate about thresholds, if not those used in decision-making. In the following section, we consider how evidence on thresholds might be used to inform policy.

3 How can Thresholds be Evidence Based?

It is common for policy-making to be based on heuristics, which expedite processes and outcomes [23] and may be informed by evidence to a greater or lesser extent. This is true for the use of CETs, for which heuristic values or ranges, such as £20,000–£30,000 per QALY in the UK and \$50,000 or \$100,000 per QALY in the USA, have long been discussed [24–26]. Evidence-based policy-making, which would be less reliant on heuristics, remains the exception in public administration, though long favoured by the research community [27, 28].

Throughout this article, we distinguish between ‘policymakers’ and ‘decision-makers’. Policymakers represent those who specify a CET based on their interpretation of what it ought to represent. Decision-makers represent those who must make decisions with reference to the threshold and local evidence, but who play no role in the specification of

the CET. The *policy threshold*—as some researchers have described it—may differ from the most accurate empirical estimate of opportunity cost [4]. This gives rise to two potentially divergent interpretations of what a CET represents and makes it important to identify a third party in threshold identification and use—researchers, who generate and interpret evidence to inform the specification of a threshold by policymakers. While we assert that it is preferable for CETs to be evidence based, other stakeholders may be more or less favourable to evidence-based CETs, depending on the implications of their adoption.

Policymakers face the challenge of using quantitative analyses as an input to their decisions, while recognising the limitations of the evidence, yet not disregarding it for its imperfection [29]. There are different ways in which evidence can be used. Lavis et al. [30] have argued that evidence may be used in *instrumental*, *conceptual*, or *symbolic* ways.

Instrumental use of an empirically estimated threshold may include direct adoption. Alternatively, if an empirical CET estimate were used as a justification to increase or decrease an explicit threshold used in policy, this would also constitute an instrumental use of the evidence by policymakers.

Conceptual use of the evidence could involve a more informal or indirect recognition in policy. For instance, if empirical estimates are used to influence implicit thresholds—say, through consideration by health technology appraisal committees—this would constitute a conceptual use of the evidence.

Symbolic use of evidence is a political strategy, which might involve post hoc justifications for decisions made without reference to the evidence. For instance, a decision-maker may justify a negative reimbursement decision on the basis of an ICER exceeding a threshold, where this threshold was not prespecified. We do not consider this a basis for evidence-based policy.

Most evidence-based policy does not involve the identification of a single point estimate on which to base decisions, not least to allow leeway to accommodate the unique factors and context of each decision. Furthermore, decisions are not generally determined by the efficiency of prevailing service provision and, by extension, historic decisions. CETs, on the other hand, may be used to determine—and be determined by—the efficiency of health care. Consequently, to our knowledge, there are no approximate analogies in policy to the adoption of an empirically estimated CET. Some policy-making uses evidence on the value of a statistical life (VSL) [31], which is similar in some respects to a CET. However, the estimation and use of supply-side thresholds assumes the presence of a fixed budget, whereas VSL estimates do not, and the budget constraints are implicit. For these reasons, the

translation of evidence into policy in the context of CETs is uniquely challenging.

4 What is a Supply-Side Threshold?

The basic principle of a supply-side threshold is to represent the benefits that a health system can currently achieve from the reallocation of a fixed budget. If spending more on one thing means spending less on another, then new technologies will result in the displacement of current service provision. New technologies that are not cost-effective, when judged against the threshold, would displace more beneficial expenditure from existing programmes, with an overall negative impact on population health.

In an ideal world (i.e. a stylised theoretical conceptualisation), policymakers could observe the cost-effectiveness of all possible health care programmes. New technologies could displace the least cost-effective programmes currently provided. The cost-effectiveness of these displaced programmes could be specified as a *shadow price* and adopted as a threshold. In this case, the threshold would represent *health opportunity cost*—the value of the health gain foregone from the next best use of the resources involved in adopting a cost-increasing technology [32, 33]. The following discussion of threshold estimates is limited to the domain of health in the context of health maximisation within a fixed budget, recognising that a decision-maker may value health gains differently. Though we refer to *opportunity costs*, it is important to note that we are not considering the full (non-health) opportunity costs of resource allocation decisions in health.

It is infeasible to observe the cost-effectiveness of all programmes and services within a health system and observe that which is displaced. To our knowledge, there are no examples of successful attempts to do so at scale, though there is one well-known unsuccessful example (i.e. in Oregon [34]). In lieu of this information, policymakers may seek to identify a CET that is relevant across the whole health care system.

The *marginal product* of health care expenditure can be represented by the change in health outcomes arising from a unit change in expenditure ‘at the margin’. Alternatively, *average displacement* could represent the average change in health outcomes under observed budget contractions or expansions. In this case, new technologies that are cost-effective relative to this threshold would, *on average*, improve the efficiency of health care expenditure. Recent examples of threshold estimation (e.g. [12, 13, 35, 36]) have sought to identify marginal productivity. Depending on our interpretation of the methods and our satisfaction with the inherent assumptions, we may alternatively interpret these studies as suited to the identification of average

displacement. While these two approaches are distinguishable in theory, any given analysis may include characteristics that partially satisfy both.

A threshold estimate may also be characterised as an *outcome elasticity* when a proportional association between inputs and outputs is identified (as may be imposed by a statistical model). Outcome elasticity estimates can identify different thresholds depending on potential health gains in different contexts (see, for example, [37]) and may not formally identify the causal effect of expenditure on outcomes.

Table 1 outlines these four related interpretations of a supply-side CET. Ideally, an empirical study would adopt one of these concepts. However, in the assessment of current evidence, these four interpretations are neither mutually exclusive nor collectively exhaustive. While we label concept A as ‘shadow price of health’, any concept could be equivalent under the stated assumptions (according to the theoretical interpretation described in Table 1). Only concept A will provide an accurate estimate of opportunity cost if these assumptions cannot be met. An important question is how well concepts B–D approximate the true opportunity cost when the assumptions fail or, equivalently, how biased they are. We discuss aspects of published empirical estimates in more detail in Sect. 5. We do not classify each study according to the concepts outlined in Table 1 since each empirical study exhibits characteristics appropriate to multiple conceptions.

Theoretical conceptions of supply-side thresholds provide a basis for their estimation and interpretation, determining the evidence and assumptions required. Researchers have proposed a variety of different approaches, each with different implications for the resulting threshold. With perfect information, all investment decisions would be based on a specific opportunity cost, but, as Table 1 illustrates, there is a trade-off between realistic assumptions and data requirements. A key question to consider is whether each concept can and should be used to inform a supply-side threshold, given the assumptions required and how it will be used. We discuss these matters further in Sect. 5.

5 What is the Basis of Current Evidence?

The evidentiary requirements for identifying a threshold relate to (1) the theoretical interpretation of ‘cost-effectiveness threshold’, (2) the data sources for the generation of the evidence, and (3) the methods used to generate estimates. In Table 1, we provide an illustration of the differences in evidence that might be associated with different interpretations of the CET.

Concepts B, C, and D (in Table 1) represent imperfect estimates of the opportunity cost of a decision and are inferior to concept A as a basis for informing decision-making.

Table 1 Different conceptions of supply-side cost-effectiveness thresholds

Concept	Theoretical interpretation	Evidence requirement	Data sources	Methods
A Shadow price of health	<i>Cost per unit of health gain</i> The next best opportunity foregone (measured by the health gain that would be sacrificed) as a result of specific health system investment decisions about specific technologies	Cost-effectiveness of all current and potential programmes of expenditure, accounting for budget impact and the timing of expenditures	Economic evaluations of all individual technologies funded and unfunded	Local league tables. Programming budgeting and marginal analysis
B Marginal product	<i>Change in health per unit change in expenditure</i> Change in output caused by a single unit change in inputs at the margin, according to a specified production function, typically allowing for diminishing returns Equivalent to the inverse of A with perfect divisibility of treatments, independence, marginal budget impacts, and perfect information	Levels of input (health spending/capital/labour) and output (mortality/life years/QALYs), with exogenous variation across observations	Expenditure and outcomes data with variation at the programme-level for different sites (commissioners or geographies) or individual patients and time points	Linear programming. Estimation of the coefficients of a production function (regression methods or stochastic frontier analysis) relating health spending as an input to health outputs/outcomes
C Average displacement	<i>Change in health per change in expenditure on average</i> Causal effect on outcomes of previous changes to expenditure equivalent to the net cost of the new technology Equivalent to B if there is a constant marginal product and the health system has allocative efficiency and optimal displacement	Levels of input (health spending/capital/labour) and output (mortality/life years/QALYs), with variation across observations	As for B , plus information on control variables and instruments as necessary	Experimental or quasi-experimental methods to identify causality (e.g. instrumental variable estimator to allow for endogeneity of health care spending)
D Outcome elasticity	<i>Percentage change in health per 1% change in expenditure</i> Average proportional association between budget changes and health output Equivalent to C in relative terms if assumptions about causal effect are identified	As for C Deriving the absolute average displacement for a given relative displacement requires data on total potential amount of health gain in a given population (burden of mortality and morbidity)	As for B	Linear (or log linear) regression of health outcomes on health spending

QALY quality-adjusted life year

However, the evidentiary requirements for thresholds that guarantee improvements in efficiency for every decision are far greater than those that may ensure improvements in efficiency on average (e.g. concept C). It is unlikely that any decision will be made based on evidence that is sufficient to precisely estimate a concept A threshold. Indeed, if the true opportunity cost of every decision could be estimated, a threshold approach would not be necessary, as a new technology would be directly assessed against that which would be displaced. The purpose of an evidence-based supply-side CET is to support decisions that *approximate* optimal decision-making under such conditions.

Several reviews of the evidence have already been conducted [1, 3, 8, 38–40]. In this section, we consider the evidence base generally, from a global perspective, relating it to the different interpretations of thresholds specified in Table 1. In the following sections, in relation to theory, methodology, and data, we assess the issues that might limit the informativeness of evidence for policy-making. In particular, we assert that the evidence base generally provides estimates of outcome elasticity (concept D) or average displacement (concept C), and that the limitations of the evidence as a basis for decision-making are poorly understood.

5.1 Are Estimates Theoretically Robust?

The use of CETs, in general, has been criticised on practical and theoretical grounds [41–43]. We assert that empirical estimates of thresholds are, in principle, a useful input to decision-making. We focus on the theoretical nuances that help to determine how they ought to be used.

The appropriateness of the theoretical basis underlying an empirical estimate of a threshold depends on its application [44]. Here, we assume that policymakers wish to use a threshold that ensures efficiency in the use of a fixed health care budget, and that technologies and their prices are exogenously determined. In this case, the true opportunity cost of the decision would be the appropriate basis for decision-making, and empirically based thresholds should seek to approximate this.

The most influential attempt to estimate a threshold in recent years was published by Claxton et al. [8], building on earlier work by Martin et al. [45, 46]. This work specified ‘the expected health effects (in terms of length and [quality of life]) of the average displacement’ as the relevant statistic for NICE’s remit. Thus, the theoretical basis for this estimation is specified, corresponding to a concept C estimate in Table 1. However, the methodology proceeds with an intention to identify the effect of marginal changes in expenditure, which corresponds more closely to concept B. Other work using similar methods has been described in similarly inconsistent ways, tending to frame key results as elasticities [12, 13, 47]. Some studies, also using similar methods, have

described their analyses as estimating marginal productivity or marginal returns to expenditure [35, 36, 48, 49]. Other studies state that they seek to identify opportunity costs, but present elasticities between expenditures and outcomes [37, 50, 51], with little consideration of the extent to which this represents a compelling theoretical basis for a CET.

In general, the framing of the relevant threshold estimate is determined by the specification of the econometric model in each study. The term ‘opportunity cost’ is routinely used in research and in material for wider audiences [52, 53], as a broader concept that highlights the importance of trade-offs. However, this tendency to overlook nuance in empirical work that seeks to identify opportunity cost may mislead policymakers and other stakeholders.

To equate marginal product and average displacement is to infer changes at the margin from differences on average. Where marginal product is not constant, average estimates may differ substantially from the margin. Average displacement can, in theory, support the correct decision *on average* and can inform high-level decisions that affect population health. However, it may result in many suboptimal decisions and systematically disadvantage outcomes in certain contexts (e.g. where displacement is less efficient).

There are clear differences between the objectives underlying purchasers’ prioritisation decisions and the assumptions inherent in a threshold based on QALY maximisation. Research exploring investment and disinvestment decisions has revealed that payers tend to have more complex objectives [54–56], such as reducing inequalities or waiting times. This inconsistency undermines the extent to which empirical CET estimates represent opportunity cost. In practice, there are many sources of variation that affect local production functions, including varying valuations of health gain [57]. This limits the extent to which average measures of opportunity cost capture the realities of displacement in local settings.

The focus of empirical work to date has been on the estimation of a single threshold for use in policy. Little attention has been given to the theoretical basis for single thresholds compared with the use of multiple thresholds or a threshold range, despite the latter approach being adopted in some contexts, including appraisals by NICE and the Institute for Clinical and Economic Review [2]. Some researchers have recently argued that such estimates should be used in appraisals to represent opportunity costs, but not to define a policy threshold range, because such threshold ranges exist in recognition of a variety of other criteria beyond efficiency [58].

Recent research has begun to address the reasons why a threshold might be adjusted according to the parameters of a particular decision [59]. The prevailing supply-side threshold relating to a given decision assumes a marginal budget impact, such that a non-marginal budget impact

implies a lower threshold [60]. In a supply-side threshold decision context, Claxton et al. [61] and Gravelle et al. [62] have argued that the discount rate applied to expected future health effects could be lower than that applied to expected future costs. This research highlights that thresholds are not necessarily fixed over time and that optimal CETs rely on assumptions about other parameters of a decision.

The true opportunity cost of expenditure will differ for every decision, as different technologies will affect different populations and clinical areas, and thus displace technologies delivered with varying levels of efficiency [63]. There may also be systematic heterogeneity across different decisions. In this case, the use of a single national threshold as an overarching decision rule may introduce bias, favouring those areas where health care is less productive. For example, regional variation in opportunity costs may be unavoidable due to differing decentralised health budgets, costs, epidemiology, or quality of care.

Some HTA agencies have exhibited a preference for some QALY gains over others, by using differential CETs. For example, NICE has attached a greater value to QALY gains for 'end of life' treatments, implying a threshold of £50,000/QALY, and uses a threshold of £100,000/QALY for highly specialised technologies [64]. Similarly, Zorginstituut Nederland (ZIN) accepts higher ICERs for treatments according to the severity of conditions [65]. If we accept that there will be special cases, then the use of empirical estimates for the base case implies the need for empirical estimates for special cases. To date, empirical research to identify CETs has not recognised the use of multiple CETs in decision-making, or the role of equity considerations and other factors that impact the value given to health gain by decision-makers.

The use of a threshold in decision-making is necessarily political (in the broadest sense), as it implies a certain objective for public health care funding overall, with the consequence that the allocation of public budgets may or may not align with the public's preferences for individual decisions. Yet, public choice theory has been given little or no attention in the estimation of empirical supply-side thresholds. This lack of consideration makes it difficult to identify how empirical thresholds will be used in practice and, therefore, difficult to judge the appropriateness of the theoretical basis for their estimation. As described in the following section, current methods rely on historical correlation for the estimation of a threshold. The extent to which this is a satisfactory basis for decision-making remains unexplored by researchers.

Empirical estimates to date have relied on the identification of elasticities to provide a single threshold in terms of QALYs. Independent of whether these analyses succeed in their aims, it is not clear whether such estimates provide the

information that policymakers need or expect. Some HTA agencies use multiple thresholds and threshold ranges [2], signalling that strict adoption of a single threshold is not a policy-making objective. In this case, there is a theoretical divide between evidence-based supply-side thresholds and thresholds used in practice. This divide may be bridged with clearer articulation of policymakers' objectives beyond health maximisation and further empirical work on that basis.

5.2 Are Estimates Based on Robust Methodology?

League tables are based on ranking (and funding) a set of interventions from the most to least cost-effective, with the threshold defined as the cost-per-outcome of the lowest-ranked intervention that can be funded from the budget [66]. League tables, in principle, enable the inference of opportunity cost, based on knowledge of costs and outcomes of all current and potential programmes of expenditure.

Linear programming is a constrained optimisation method, based on a single budget constraint, and constitutes an early conceptualisation of threshold estimation. In this case, the threshold represents the inverse of the shadow price of the budget constraint and is defined as the magnitude of the improvement in outcomes that would result from a one-unit increase in the budget [32], meaning that it represents a concept B estimate in Table 1.

Both league tables and linear programming require full (a priori) knowledge of the costs and benefits of all interventions and have therefore not been used to identify empirical thresholds. Thokala et al. [3] demonstrated the equivalence of linear programming and league table methods under a set of strong assumptions including perfect divisibility of treatments, independence, marginal budget impacts, and perfect information. Epstein et al. [67] demonstrated that budget constraint optimisation can be extended from a linear programming problem with a single budget, to include further constraints and assume indivisibilities and non-linearities, with the consequence that (concept A and B) thresholds are shown to differ depending on the budget, the discount rate, and equity considerations.

In the absence of complete data on the cost-effectiveness of all individual programmes, or investment and disinvestment decisions under budget changes, current evidence relies on aggregate information on spending and health outputs in the health system.

When relying on aggregate data (for concept B, C, or D estimates), the target of inference is the causal effect of health spending on health outcomes. In an idealised experiment, this effect could be estimated by randomly allocating clusters of health care purchasers (e.g. NHS Commissioners)

to either a 'treatment group' with increased (or decreased) budget or a 'control group' and observing outputs over time. Obviously, this experiment would be unethical and impractical. The methodological challenge is to replicate this experiment using observational data.

Changes in health care expenditure are deterministic; variations between providers, jurisdictions, or regions are typically determined by need, particularly for systems with centrally allocated budgets. In any study that naively compares spending and outcomes in these circumstances, the effect of spending is endogenous due to reverse causality. An exception may arise if budget expansions or contractions result from policy changes and political shifts, such as a change in government. These changes may approximate random change with respect to underlying health shifts. However, if these budget shifts occur simultaneously across all sites, there will be a lack of contemporaneous control. Even in this experimental scenario, the effect of budget changes will not be separable from concurrent secular shifts in health outcomes.

Some studies have adopted an instrumental variable approach to estimate the effects of health care expenditure on outcomes [8, 12, 13, 47]. These studies rely on the existence of a valid instrument for health spending that is both strongly correlated with health spending and only affects health outcomes through its effect on health spending (i.e. the exclusion restriction). The existence of such an instrument is questionable, as health systems generally try to avoid variation in health spending that is unrelated to the health needs of the population. Nevertheless, the exclusion restriction has been claimed to hold for instruments based on socio-economic variables [35, 63] and is supported by the decomposition of health expenditure according to the 'funding rule' [68]. Variation in the instrument can simulate random variation in health expenditure and estimate a local average treatment effect that represents the causal effect of health expenditure on health outcomes. In this case, it is necessary to assume that the local average treatment effect is generalisable to other causes of change in the budget; otherwise, the interpretation of the estimates changes again.

Health outcomes at an aggregate level are dynamic processes with future values of the series dependent on past values. Furthermore, not all the effects of health care expenditure are likely to be realised immediately; many manifest over several years [69]. This creates another challenge for the accurate identification of causal effects. Yet, threshold

studies have relied on static models, even where data are available over time [35]. Researchers have found that it is not possible to use time series analyses to estimate a threshold based on marginal changes in expenditure and outcomes, due to reverse causality [47]. Thus, it has not been possible to use variation in expenditure over time to estimate its impact on health outcomes. Analyses remain reliant on regional variation within cross-sectional data.

A recent systematic review and meta-analysis of studies investigating the relationship between health care expenditure and population health outcomes identified 65 studies on the topic [70]. Most of these studies employed panel or longitudinal data on regions or countries and estimated static models. We have identified examples of dynamic panel-data models used in this context [71–73]. Both Crémieux et al. [73] and Guindon and Contoyannis [71] examined the impact of pharmaceutical spending on life expectancy and infant mortality in Canada using panel data from Canadian provinces, allowing for autocorrelated error terms. Ivaschenko [72] estimated the effect of health care expenditure on life expectancy using a panel of Russian provinces, modelling an autoregressive AR(1) process. Recent work using a distributed lag model highlights the potential importance of such methodological complexities [74]. Some authors have advocated that static models can be used to estimate long-run equilibrium relationships between two dynamic processes by imposing certain assumptions on model parameters [75, 76]. However, the estimators of such static models may be significantly biased [77, 78], especially with small sample sizes as many of the aforementioned studies have.

There are additional relationships in the data that may need to be included in a model to provide reliable estimates. For example, spatial correlation between health care centres, sites, or regions is likely to be present since health outcomes are similar in proximal locations. Ignoring this correlation will lead to inefficiencies and potential bias depending on the modelling assumptions. For instance, if people in higher-spending areas have poorer health behaviours, the effect of spending may be underestimated.

Heterogeneity is an important consideration in this context, and there is no consensus on how heterogeneity ought to be accounted for in the estimation of a threshold. One approach that has been used is weighting. For instance, Claxton et al. [8] present estimates weighted by clinical area for purchasers in England. Heterogeneity is observed across different specialisms, and observations may be weighted by

mortality rates [63]. Vallejo-Torres et al. [12] argue that differences in the threshold due to scale—associated with population size—should not be considered. If larger regions are associated with greater marginal productivity (i.e. a lower threshold), weights depending on population size in regression analyses will drive down the threshold, leading to sub-optimal decision-making in smaller regions.

Little research, if any, has dealt with these various modelling questions in relation to CETs. Studies that estimate CETs tend not to compare results from fundamentally different modelling assumptions. While some researchers have considered structural uncertainty (e.g. [49]), it is difficult to say how much uncertainty remains regarding the values of the parameters of interest. Published standard errors only reflect sampling variation for a specific model.

Empirical estimation of a supply-side CET demands identification of causal effects. Idealised experimental conditions will never be satisfied, and the endogeneity of budgets makes reliable estimation challenging. The individual challenges are not necessarily unique to identifying thresholds, but these methodological uncertainties call for careful consideration by policymakers seeking to adopt an empirical estimate in policy.

5.3 Are Estimates Based on Appropriate Data?

All empirical estimates of supply-side thresholds rely on data relating to health care expenditure and health outcomes. These may be observed directly or indirectly, and at different levels of aggregation. Yet, there are no data designed for the purpose of estimating supply-side CETs and, therefore, limitations in the data are inevitable. In addition to specific challenges with expenditure and outcome data, the association of the two can also be problematic because of a mismatch in data coverage across clinical areas.

At the minimum level of aggregation, it may be possible to observe patient-level data from insurance claims (for instance). At the maximum level of aggregation, time series of national data on expenditure may be available. More aggregated data prevents analysts from fully accounting for heterogeneity and controlling for individuals' health care needs. The more disaggregated the data are, the greater the risk of missing health expenditure that is incurred at the national level.

For expenditure, Claxton et al. [8] used programme-level (clinical area) budget data, while analyses in Australia

and Spain were not able to do this and relied on overall regional expenditure [12, 13]. Expenditure data cannot be disaggregated in many settings and limitations in the quality of expenditure data are poorly understood. In settings where budget deficits are common (e.g. in England [79]), it is important that estimates should rely on realised expenditure rather than budget allocation.

High-quality data on morbidity outcomes associated with health care are lacking. This is arguably the most significant challenge to the identification of appropriate data for the estimation of opportunity costs. Threshold estimates tend to rely on mortality data (e.g. for England [8], Australia [13], and South Africa [36]), which can provide limited information about heterogeneity in the population. For instance, mortality data often do not include statistics by age and health care provider. In some clinical areas (e.g. maternity and neonatal care in high-income countries), mortality does not capture the most meaningful health outcomes.

Threshold estimates for Spain, Australia, and the Netherlands have relied on an indirect estimation of average health-related quality of life values. This involves using age- or sex-specific estimates, observed in surveys, to weight life years derived from mortality data [12, 13, 49]. In England, there is a lack of quality of life data that can be used to estimate QALYs, so the analysts relied on a set of assumptions—relating to disease burden and the comparability of mortality effects to morbidity effects—that may be inaccurate [80].

Problems with identifying the relationship between expenditure and outcomes may be further compounded when the availability of data relies on health care expenditure. This is especially relevant in low- and middle-income countries (LMICs), where reliable outcomes data do not generally exist. In this context, mortality or disability-adjusted life year (DALY) estimates from sources like the Global Burden of Disease studies are used (e.g. [51]). The quality of these estimates may partly depend on the availability of certain health care and public health technologies, which depend on health care spending.

The availability of data affects the empirical model, the possibilities to correct for the endogeneity of expenditure, and the approach to attributing health effects to changes in health expenditure as distinct from other confounding factors and unobservable heterogeneity. Therefore, numerous limitations in the data used to estimate supply-side thresholds are inevitable, the implications of which remain unknown.

6 How Should Policymakers Use Imperfect Evidence?

We have provided reasons to be cautious in adopting empirically estimated supply-side CETs. The appropriate theoretical and empirical basis for their estimation is unclear, the methods employed to date may not be capable of accurately estimating causal associations, and the data available may not be adequate. However, the question remains as to what ought to be the basis for a threshold.

Policymakers should strive to make instrumental use of evidence to inform CETs, so long as the evidence is sufficient. It is important to distinguish between statistical estimates of the impact of expenditure on outcomes and a CET that is used in policy. The identification of a threshold for use in policy necessarily depends on a normative judgement rather than a purely empirical estimate.

In the absence of any other prioritisation criteria, having no threshold risks an unfair and inefficient allocation of resources and overpaying for some new technologies [44]. Some attention has been given to the cost of setting a threshold too high, in terms of the displacement of more cost-effective care [81], though there is no evidence to demonstrate the extent of this. Little attention has been given to the costs associated with setting a threshold too low, which might include a restriction of access to cost-effective new technologies and a stifling of innovation.

Policymakers may perceive that any change to the threshold due to uncertain or misinterpreted evidence may result in harm, especially if a current threshold is politically acceptable. A certain standard of evidence should therefore be required to affect such change. To this end, we provide four key recommendations to policymakers seeking to choose a threshold: (1) define the decision scope, (2) develop an evidence assessment process, (3) maintain flexibility, and (4) support decision-makers.

6.1 Define the Decision Scope

The extent to which evidence is sufficient depends on the ways in which the evidence will be applied. Despite their duty of transparency [82], policymakers in general have not provided an account of their intended and actual use of CETs.

The key point on which clarity is lacking is in the scope of thresholds—that is, to which decisions they should apply and how they will be used. Other researchers have described the importance of scope in relation to the definition of (opportunity) costs [44]. Policymakers should refine the scope of CETs and communicate this to researchers and the public. In Box 1, we provide a set of questions that should be considered in defining the scope of decisions where thresholds will be applied.

Box 1: Questions to define the decision scope

Questions to define the decision scope

Several key questions must be answered in defining the decision scope for a threshold. These questions should determine the approach to estimating an empirical threshold.

- On which decision-makers' perspective(s) will the threshold be based?
 - What is the budget?
 - Who is the budget holder(s)?
 - How do we characterise the budget holder's aims?
 - What outcomes are they seeking to achieve?
 - How should outcomes be measured?
 - Is there a single or multiple objectives?
 - What is the objective (e.g. maximisation)?
- To which technologies and services will the threshold be applied?
 - Can the value of these technologies be captured by a single outcome (e.g. quality-adjusted life years)?
 - Why should the application of the threshold be restricted in this way?
- What type of decisions will the threshold inform?
 - Policy decisions
 - Policy evaluation
 - Market approval
 - Price negotiation
 - Investment and disinvestment decisions
 - Clinical guidelines
 - Clinical decision-making

There are multiple decision-makers within any health service. Each may have a different perspective and thus seek to employ a different threshold or decision-making framework. A 'national' threshold estimate may differ from that relevant to local budget holders, where different prices and economies of scale or scope may apply. It may also be important to consider contexts beyond the health service and outcomes beyond health. There are many contributors to health outside of health care, particularly social care, which might justify a multisectoral approach [83].

If the ambition of policymakers is to use thresholds to optimise efficiency across a range of services, within a total health care budget, then a standard approach is needed. However, it is common practice for thresholds to be applied

to only a subset of funding decisions (e.g. primarily to new medicines in Australia, Canada, England, the Netherlands, Korea, Scotland, and Sweden). Spending on new technologies, infrastructure investments, staffing decisions, or any decisions about allocative efficiency could be considered within the same paradigm.

As a tool of economic evaluation, a threshold is a means to achieving politically and socially acceptable ends. The costs of meeting alternative or changing political goals might differ, and different estimates are required for different goals. For instance, a decision-maker might reasonably prioritise a less cost-effective treatment for a disadvantaged group over a more cost-effective treatment. Empirical estimates of average displacement alone cannot guide such decisions as they provide no information about distributional impacts. It is necessary that policymakers clearly formulate the constraints of the problem—such as concerns for distributional impacts—to support an optimisation exercise, as these constraints determine a shadow price [29].

The empirical estimation of thresholds must consider the context in which they will be applied. Different countries use thresholds in different ways, and this is not by chance. In countries with complex systems of health care financing and provision, with multiple budget constraints in operation, the role (and benefit) of a single threshold approach becomes clouded. This may be especially applicable in LMIC settings. It is also important to consider how the use of a threshold relates to the policy and regulatory context. For example, expenditure on branded medicines in the UK is capped, yet this budgetary threshold is not considered in either the estimation or application of CETs [10].

It is beyond the scope of this article to explore the various ways in which a CET may be used in decision-making. However, thresholds should be seen as an input to HTA, and HTA as an input to a broader framework of priority setting in health [84]. The extent to which a CET may determine recommendations can therefore vary as CETs interact with other mechanisms. Similarly, the nature of recommendations that are supported by thresholds should inform their estimation. For instance, a threshold might be used to determine the availability of a medicine, or it might be used to inform clinical guidelines. A given threshold in each context is likely to have different implications for resource allocation and health outcomes. Furthermore, investment decisions might be considered as distinct from routine decisions about health care provision or disinvestment decisions [85]. A threshold might simply be used as a price negotiation tool, to manage competition, in which case its relationship to the budget constraint might be considered incidental. Where a CET is adopted as an ‘approval norm’ [86], it is incumbent on policymakers to specify how this CET should be informed by evidence, with full consideration of the ethical and economic consequences.

6.2 Develop an Evidence Assessment Process

As more evidence on opportunity costs is generated, and as budgeting arrangements change over time, it is inevitable that evidence will be conflicting, collected using different theoretical bases, methodologies, and data sources. As previously stated, there are few, if any, analogies in policy to the process of setting an evidence-based CET. Thus, a bespoke dedicated assessment process for new evidence to inform the threshold should be established. In Box 2, we propose some characteristics for such a process and the questions that it ought to consider.

The assessment process will need to determine whether, and under what circumstances, there is enough relevant evidence to justify a change in threshold. Where an explicit threshold is already used by policymakers, some formal process is required to consider whether evidence is sufficient to justify a change, given the possibility of harm if the wrong decision is made (or if an inappropriate threshold is maintained). Where an explicit threshold is not used, an assessment of the implications of adopting a threshold should be conducted.

HTA agencies might adopt the task of interpreting evidence and setting a threshold, or there might be a separate body given this task. For example, an analogous organisation to an interest rate setting committee of an independent central bank (such as the Monetary Policy Committee in the UK or the Federal Reserve Board in the USA) might be suitable [57]. Explicit adoption of such a task may support better availability and transparency of relevant data from health systems, which would support further empirical work.

In adopting a CET, it is important to understand what is displaced or foregone when a new technology is adopted (see, for example, [87, 88]). Currently, there is little or no data available on either (1) what is displaced in practice or (2) the cost-effectiveness of those things that are displaced, diluted, or delayed. Without information about the cost-effectiveness of prevailing care provision which might be displaced, the impact of employing a threshold to a decision cannot be assessed.

Crucially, the establishment of such a process depends on agreeing the theoretical basis for CETs and the scope of their application (as described above). There is a need for openness, scrutiny, and triangulation. For any policy to rely on a single study or methodological approach is risky; validation and replication are important tools. Reproducibility in different settings or under different conditions should be part of an assessment of evidentiary sufficiency. The process should involve a systematic assessment of sources of bias as well as of sources of structural and parameter uncertainty.

There are numerous stakeholder perspectives to consider in setting CETs [88]. Some stakeholders may hold opposing views about the evidence, so it is vital that the process

is inclusive. The limited influence of empirical estimates to date—as described above—may have arisen in part from industry stakeholders' opposition to those estimates. Similarly, the complexity of the research challenge calls for a multi-disciplinary approach. Given the international application of such approaches, there is a need to build capacity globally. This is particularly important for LMICs. The application of evidence from the UK to other countries [19], with no assessment of the relevance of such evidence, is problematic [89].

The assessment process should acknowledge the complexity of decision-making in health care. Thresholds are part of a broader process, so changes (up or down) may be countered by other parts of the process; there are numerous levers available to influence decisions. For instance, the definition of a reference case for economic evaluations has a significant bearing on the relative cost-effectiveness of technologies. Furthermore, in view of the evidence, criteria should be established for deviating from a reference threshold in individual cases or contexts. We discuss this in the next section.

We have focussed on supply-side threshold estimates, but demand-side estimates may also be considered important to policy-making [90] where health expenditure impacts social insurance, individual insurance premiums, or rates of taxation. The assessment process may therefore seek to reconcile alternative types of evidence.

In addition to the considerations for new evidence listed in Box 2, it is also important to consider the costs of adopting or changing a threshold. These costs may be considered from a political, normative, or economic perspective. At the very least, tangible costs associated with the communication and realignment of policy could be identified.

More generally, there is a need to understand the consequences of threshold changes. It seems logical that a threshold should change over time as populations, budgets, and prevailing health care evolve [12]. A policy response will be required for the historical inconsistency in decision-making that arises from a threshold change. For example, there may be a desire to revisit decisions based on earlier CETs used in policy.

For a jurisdiction that is looking to change a threshold based on new evidence, a further question arises as to how frequently the threshold should be reassessed. NICE's threshold has remained unchanged for around 15 years. In principle, the threshold could be changed every year as the latest evidence becomes available. A Bayesian approach to integrating evidence could be adopted, though this must recognise the trade-off between constant updates for accuracy, the upheaval and costs of change, and the long-term nature of drug development. Regulatory impact assessments and value of information approaches could be used to assess whether a new estimate represents a basis for revising a CET

or if a higher hurdle was needed, perhaps with additional research to address uncertainty.

We are not aware of any such process being employed by policymakers. However, existing research (e.g. in Spain [91]) has sought to answer some of the questions that we have specified in Box 2, and such research could form the basis of assessment processes.

Box 2: Developing an evidence assessment process

Developing an evidence assessment process

An evidence assessment process should adopt the following characteristics:

- Systematic, *e.g. identification of all relevant evidence*
- Critical, *e.g. assessment of the quality of the evidence*
- Transparent, *e.g. pre-specified terms of assessment*
- Inclusive, *e.g. multi-disciplinary and cross-stakeholder approach*

And address the following questions:

- What is the theoretical basis for the empirical estimate?
- Are the methods employed appropriate?
- Are the data suitable?
- What are the sources and level of uncertainty?
- What are the sources and extent of bias?
 - Have the findings been validated?
 - Internal, external, and ecological validity
- How should the current evidence inform policy?
 - What additional evidence could be collected to address concerns about uncertainty or bias?
 - What are the implications of changing the threshold now?
 - What are the implications of waiting for additional evidence before acting?

6.3 Maintain Flexibility

Policymakers should recognise that CETs are a blunt instrument because they do not capture all that is relevant to resource allocation decisions. Thresholds should not be used in a mechanistic manner and, because the extent to which they are appropriate may vary according to the context of the decision, their application should be flexible. Prevailing approaches to the use of thresholds highlight the benefits of maintaining flexibility. There are at least four respects in which flexibility should be maintained.

6.3.1 In Recognition of Uncertainty

Uncertainty in economic evaluations can be managed and quantified (through sensitivity analysis), but cannot be resolved. Inflexible application of a fixed threshold across all decisions implies confidence about the implications of displacing budgets and hence the cost-effectiveness of technologies being assessed. Decision-makers rarely have confidence in either the expected cost-effectiveness of technologies or the opportunity cost of a particular decision. The application of a (quantitative) threshold should allow for a (qualitative) consideration of the nature of uncertainty in a decision.

There are accepted methods for presenting evidence on uncertainty about the cost-effectiveness of a new technology, but not for characterising uncertainty in a CET. Ways of presenting uncertainty relevant to the econometric approaches used are valuable, and researchers have adopted a variety of approaches [40]. Recommendations for parameter uncertainty to be conducted around relevant structural uncertainties in key modelling assumptions—and not just authors' preferred estimates—are also relevant [92].

6.3.2 In Recognition of the True Health Opportunity Cost of a Decision

There are specific scenarios in which a purposeful flexibility can be used to adjust thresholds depending on the inputs to (and consequences of) the decision process. For instance, the opportunity cost of a decision depends on its budget impact [60, 93]. There may also be grounds for adjusting thresholds in light of additional QALY benefits relating to dynamic efficiency [94] and the value of innovation [95].

6.3.3 In Accounting for Non-QALY Sources of Value

Health services do not proclaim the sole aim of maximising the health of the population. To this end, HTA agencies have made use of equity weights or modifiers to adjust thresholds in particular cases, such as severity or rarity of illness [2]. In this sense, adjustments may be applied to a threshold estimate for some decisions. Nevertheless, it should be recognised that thresholds cannot capture everything that is of value. HTA agencies that make explicit use of thresholds do not employ them in a mechanistic manner, but rather use them as part of a broader deliberative process. Where cost-per-QALY thresholds are used, it is important to consider non-QALY values as part of the process [96].

6.3.4 In Recognition of Moral Principles

The true opportunity cost of a decision will be different at the local level compared with the national level.

It is necessary to make a judgment about the trade-off between local variation and national solidarity, which cannot rely solely on evidence of opportunity cost. For instance, use of a CET may imply that a local health care investment is not cost-effective, but if the rest of the country already has access to this investment, parity of care may be a priority, and the CET may be deemed less relevant to the decision. More generally, it is important to recognise the political nature of CETs and the conflicting political objectives of decision-makers that use them.

6.4 Support (Local) Decision-Makers

Consistency needs to be maintained between the estimation and application of thresholds and their consequences for decision-making. However, there will remain inconsistencies in the priorities of decision-makers at different levels. Those setting thresholds for use in policy will not necessarily operate on the same terms as those making decisions about the use of resources.

Evidence suggests that local decision-makers adopt a 'satisficing' approach rather than maximisation [97], which undermines the use of thresholds at the national level. Managers and commissioners in health care are subject to numerous imperatives and policy initiatives, such as targets for waiting times and treatment uptake, that have little to do with the objectives underlying the use of a CET.

If thresholds are to support an optimal allocation of resources, local decision-makers require guidance and support to act in accordance with national thresholds. In particular, the creation of frameworks for disinvestment that are relevant to local decision-makers may support consistency in decision-making.

In countries without national HTA agencies or the capacity to make legally binding decisions about resource allocation in health care, decision-makers are likely to require even greater support. There may be many different value frameworks employed in a fragmented health care system. Each of these may choose to adopt a different threshold and the relevance of any national threshold needs to be clearly articulated by policymakers.

Policymakers may provide targeted support where decision-makers formally employ alternative decision-making frameworks, such as multi-criteria decision analysis (MCDA) or programme budgeting and marginal analysis (PBMA). Such alternative approaches are, in some respects, a valid recognition of the limitations of a threshold approach. Guidance in specific scenarios should also be considered, such as where a provider is in deficit or has large reserves.

7 Conclusions

It is essential that any use of CETs should be informed by evidence on the opportunity costs of health care expenditure. We have described some limitations in the current evidence base concerning theory, methods, and data, which may explain and justify the limited extent to which estimates have influenced policy. Further research, as described elsewhere [98], would be valuable in supporting policymakers in making instrumental or conceptual use of evidence.

Uncertainties and inconsistencies in the evidence base will persist, while policymakers retain the task of using (or not using) CETs, currently in the absence of a framework for appraising evidence for use in a decision context. We have provided a set of actionable recommendations for policymakers that can support the use of evidence in the application of CETs. In particular, policymakers should establish a process for the assessment of evidence on opportunity costs in health care, which recognises the complexity of decision-making, and clearly communicate their objectives to researchers and decision-makers.

Acknowledgements We are grateful to Pieter van Baal for discussion and comments on earlier versions of the manuscript, and to three anonymous reviewers who provided highly constructive input.

Declarations

Author contributions CS assembled the writing group and led the drafting of the manuscript. All authors contributed to writing and editing all parts of the manuscript.

Funding This work was funded by an unconditional grant from the Association of the British Pharmaceutical Industry (ABPI). The funder had no role in establishing the writing group, the decision to publish, or the preparation of the manuscript. The ABPI funded open access publication of this article. The views expressed in this article are those of the authors and do not represent the views of the ABPI or its members.

Conflict of interest CS, PCM, and AT are employees of the Office of Health Economics, which receives financial support from the ABPI and provides consultancy services to pharmaceutical companies. BZ, SW, JC, KC, ND, BGL, DAH, and AL declare that they have no conflicts of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication (from patients/participants) Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction

in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>

References

1. Santos AS, Guerra-Junior AA, Godman B, Morton A, Ruas CM. Cost-effectiveness thresholds: methods for setting and examples from around the world. *Expert Rev Pharmacoecon Outcomes Res.* 2018;18:277–88.
2. Zhang K, Garau M. International cost-effectiveness thresholds and modifiers for HTA decision making. OHE Consulting Report. 2020. <https://www.ohe.org/publications/international-cost-effectiveness-thresholds-and-modifiers-hta-decision-making>. Accessed 25 Apr 2022.
3. Thokala P, Ochalek J, Leech AA, Tong T. Cost-effectiveness thresholds: the past, the present and the future. *Pharmacoeconomics.* 2018;36:509–22.
4. Lomas J, Ochalek J, Faria R. Avoiding opportunity cost neglect in cost-effectiveness analysis for health technology assessment. *Appl Health Econ Health Policy* [Internet]. 2021. <https://doi.org/10.1007/s40258-021-00679-9> (cited 2021 Sep 16).
5. Ryen L, Svensson M. The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Econ.* 2015;24:1289–301.
6. Danzon P, Towse A, Mestre-Ferrandiz J. Value-based differential pricing: efficient prices for drugs in a global context. *Health Econ.* 2015;24:294–301.
7. Bertram MY, Lauer JA, De Joncheere K, Edejer T, Hutubessy R, Kieny M-P, et al. Cost-effectiveness thresholds: pros and cons. *Bull World Health Organ.* 2016;94:925–30.
8. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the Estimation of the NICE Cost Effectiveness Threshold. *Health Technol Assess* 2015;19(14)
9. Department of Health, Office for Life Sciences. Accelerated Access Collaborative for health technologies [Internet]. 2017. Report No.: 13003. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/663094/Accelerated_Access_Collaborative_-_impact_assessment.pdf. Accessed 25 Apr 2022.
10. Department of Health and Social Care, Association of the British Pharmaceutical Industry. The 2019 Voluntary Scheme for Branded Medicines Pricing and Access—Chapters and Glossary [Internet]. 2018. <https://www.gov.uk/government/publications/voluntary-scheme-for-branded-medicines-pricing-and-access>. Accessed 25 Apr 2022.
11. Department of Health and Social Care. Cost-effectiveness methodology for Immunisation Programmes and Procurements (CEMIPP): the government's decision and summary of consultation responses [Internet]. 2019. <https://www.gov.uk/government/consultations/cost-effectiveness-methodology-for-vaccination-programmes>. Accessed 25 Apr 2022.
12. Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Econ.* 2018;27:746–61.

13. Edney LC, Haji Ali Afzali H, Cheng TC, Karnon J. Estimating the reference incremental cost-effectiveness ratio for the Australian health system. *Pharmacoeconomics*. 2018;36:239–52.
14. Sacristán JA, Oliva J, Del Llano J, Prieto L, Pinto JL. ¿Qué es una tecnología sanitaria eficiente en España? [What is an efficient health technology in Spain?]. *Gac Sanit*. 2002;16:334–43.
15. Vallejo-Torres L, García-Lorenzo B, García-Pérez L, Castilla I, Valcárcel Nazco C, Linertová R, et al. Valor Monetario de un Año de Vida Ajustado por Calidad: Revisión y Valoración Crítica de la Literatura [Internet]. 2014. https://www3.gobiernodecanarias.org/sanidad/scs/content/e6da6619-d18b-11e5-a9c5-a398589805dc/SESCS%202014_AVAC.pdf. Accessed 25 Apr 2022.
16. Vallejo-Torres L, García-Lorenzo B, Castilla I, Valcárcel Nazco C, García-Pérez L, Linertová R, et al. Valor Monetario de un Año de Vida Ajustado por Calidad: Estimación empírica del coste de oportunidad en el Sistema Nacional de Salud. Ministerio de Sanidad, Servicios Sociales e Igualdad. Servicio de Evaluación del Servicio Canario de la Salud [Internet]. 2015. https://www3.gobiernodecanarias.org/sanidad/scs/content/3382aaa2-cb58-11e5-a9c5-a398589805dc/SESCS%202015_Umbral%20C.O.%20AVAC.pdf. Accessed 25 Apr 2022.
17. Vallejo-Torres L, García-Lorenzo B, Rivero-Arias O, Pinto-Prades JL, Serrano-Aguilar P. Disposición a pagar de la sociedad española por un Año de Vida Ajustado por Calidad. Ministerio de Sanidad, Servicios Sociales e Igualdad. Servicio de Evaluación del Servicio Canario de la Salud [Internet]. 2016. https://www3.gobiernodecanarias.org/sanidad/scs/content/c6b59111-420f-11e7-952b-a987475f03d3/SESCS_2016_DAP_AVAC.pdf. Accessed 25 Apr 2022.
18. Wang S, Gum D, Merlin T. Comparing the ICERs in medicine reimbursement submissions to NICE and PBAC—does the presence of an explicit threshold affect the ICER proposed? *Value Health*. 2018;21:938–43.
19. Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. *Value Health*. 2016;19:929–35.
20. Ochalek J, Lomas J, Claxton K. Assessing health opportunity costs for the Canadian health care systems. 2018. http://www.pmprb-cepmb.gc.ca/CMFiles/Consultations/new_guidelines/Canada_report_2018-03-14_Final.pdf. Accessed 25 Apr 2022.
21. Patented Medicine Prices Review Board (PMPRB). Working Group to Inform the Patented Medicine Prices Review Board (PMPRB) Steering Committee on Modernization of Price Review Process Guidelines: Final Report [Internet]. 2019. <http://www.pmprb-cepmb.gc.ca/view.asp?ccid=1449>. Accessed 25 Apr 2022.
22. Vallejo-Torres L, García-Lorenzo B, Edney LC, Stadhouders N, Edoxa I, Castilla-Rodríguez I, et al. Are estimates of the health opportunity cost being used to draw conclusions in published cost-effectiveness analyses? A scoping review in four countries. *Appl Health Econ Health Policy* [Internet]. 2021. <https://doi.org/10.1007/s40258-021-00707-8> (cited 2022 Jan 17).
23. Simon HA. *Administrative behavior; a study of decision-making processes in administrative organization*. Oxford: Macmillan; 1947.
24. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness—the curious resilience of the \$50,000-per-QALY threshold. *N Engl J Med*. 2014;371:796–7.
25. Towse A. Should NICE’s threshold range for cost per QALY be raised? Yes. *BMJ* [Internet]. British Medical Journal Publishing Group; 2009 [cited 2020 Mar 25];338. <https://www.bmj.com/content/338/bmj.b181>. Accessed 25 Apr 2022.
26. Raftery J. Should NICE’s threshold range for cost per QALY be raised? No. *BMJ* [Internet]. British Medical Journal Publishing Group; 2009 [cited 2020 Mar 25];338. <https://www.bmj.com/content/338/bmj.b185>. Accessed 25 Apr 2022.
27. Niessen LW, Grijseels EWM, Rutten FFH. The evidence-based approach in health policy and health care delivery. *Soc Sci Med*. 2000;51:859–69.
28. Culyer A, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, et al. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *J Health Serv Res Policy*. 2007;12:56–8.
29. Williams A. Cost-benefit analysis: Bastard science? And/or insidious poison in the body politick? *J Public Econ*. 1972;1:199–225.
30. Lavis JN, Robertson D, Woodside JM, McLEOD CB, Abelson J. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Q*. 2003;81:221–48.
31. Cubi-Molla P, Mott D, Henderson N, Zamora B, Grobler M, Garau M. Resource allocation in public sector programmes: does the value of a life differ between governmental departments? OHE Research Paper. 2021. <https://www.ohe.org/publications/resource-allocation-public-sector-programmes-does-value-life-differ-between>. Accessed 25 Apr 2022.
32. Weinstein M, Zeckhauser R. Critical ratios and efficient allocation. *J Public Econ*. 1973;2:147–57.
33. Pekarsky BAK. *The New Drug Reimbursement Game: A Regulator’s Guide to Playing and Winning* [Internet]. ADIS; 2015 [cited 2021 May 19]. <https://www.springer.com/gp/book/9783319089027>. Accessed 25 Apr 2022.
34. Blumstein JF. The Oregon experiment: the role of cost-benefit analysis in the allocation of Medicaid funds. *Soc Sci Med*. 1997;45:545–54.
35. Lomas J, Martin S, Claxton K. Estimating the Marginal Productivity of the English National Health Service From 2003 to 2012. *Value Health*. 2019;22(9):995–1002.
36. Edoxa IP, Stacey NK. Estimating a cost-effectiveness threshold for health care decision-making in South Africa. *Health Policy Plan*. 2020;35:546–55.
37. Ochalek J, Lomas J. Reflecting the health opportunity costs of funding decisions within value frameworks: initial estimates and the need for further research. *Clin Ther*. 2020;42:44–59.e2.
38. Cameron D, Ubels J, Norstrom F. On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. *Global Health Action*. 2018;11(1447828).
39. Vallejo-Torres L, García-Lorenzo B, Castilla I, Valcárcel-Nazco C, García-Pérez L, Linertová R, et al. On the estimation of the cost-effectiveness threshold: why, what, how? *Value Health*. 2016;19:558–66.
40. Edney LC, Lomas J, Karnon J, Vallejo-Torres L, Stadhouders N, Siverskog J, et al. Empirical estimates of the marginal cost of health produced by a healthcare system: methodological considerations from country-level estimates. *Pharmacoeconomics*. 2022;40:31–43.
41. Birch S, Gafni A. On the margins of health economics: a response to ‘resolving NICE’S nasty dilemma.’ *Health Econ Policy Law*. 2015;10:183–93.
42. Gafni A, Birch S. Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Soc Sci Med*. 2006;62:2091–100.
43. Cleemput I, Neyt M, Thiry N, Laet CD, Leys M. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *Int J Technol Assess Health Care*. 2011;27:71–6.
44. Culyer AJ. Cost, context, and decisions in health economics and health technology assessment. *Int J Technol Assess Health Care*. 2018;34:434–41.
45. Martin S, Rice N, Smith PC. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *J Health Econ*. 2008;27:826–42.
46. Martin S, Rice N, Smith PC. Comparing costs and outcomes across programmes of health care. *Health Econ*. 2012;21:316–37.

47. Siverskog J, Henriksson M. Estimating the marginal cost of a life year in Sweden's public healthcare sector. *Eur J Health Econ* [Internet]. 2019. <https://doi.org/10.1007/s10198-019-01039-0>.
48. van Baal P, Perry-Duxbury M, Bakx P, Versteegh M, van Doorslaer E, Brouwer W. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ*. 2019;28:87–100.
49. Stadhouders N, Koolman X, van Dijk C, Jeurissen P, Adang E. The marginal benefits of healthcare spending in the Netherlands: estimating cost-effectiveness thresholds using a translog production function. *Health Econ*. 2019;28:1331–44.
50. Vanness DJ, Lomas J, Ahn H. A health opportunity cost threshold for cost-effectiveness analysis in the United States. *Ann Intern Med* [Internet]. 2020. <https://doi.org/10.7326/M20-1392> (cited 2020 Nov 3).
51. Ochalek J, Lomas J, Claxton K. Estimating health opportunity costs in low-income and middle-income countries: a novel approach and evidence from cross-country data. *BMJ Glob Health*. 2018;3:e000964.
52. University of York. Increasing value for money in healthcare [Internet]. University of York. [cited 2021 Jan 25]. <https://www.york.ac.uk/research/impact/healthcare-value/>. Accessed 25 Apr 2022.
53. Dillon A. Carrying NICE over the threshold [Internet]. NICE. NICE; 2015 [cited 2021 Jan 25]. <https://www.nice.org.uk/news/blog/carrying-nice-over-the-threshold>. Accessed 25 Apr 2022.
54. Appleby J, Devlin N, Parkin D, Buxton M, Chalkidou K. Searching for cost effectiveness thresholds in the NHS. *Health Policy*. 2009;91:239–45.
55. Karlsberg Schaffer S, Sussex J, Devlin N, Walker A. Searching for Cost-effectiveness Thresholds in NHS Scotland. OHE Research Paper. 2013. <https://www.ohe.org/publications/searching-cost-effectiveness-thresholds-nhs-scotland>. Accessed 25 Apr 2022.
56. Karlsberg Schaffer S, Sussex J, Hughes D, Devlin N. Opportunity costs of implementing NICE decisions in NHS Wales. OHE Research Paper. 2014. <https://www.ohe.org/publications/opportunity-costs-implementing-nice-decisions-nhs-wales> Accessed 25 Apr 2022.
57. Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. *BMJ*. 2007;335:358–9.
58. Siverskog J, Henriksson M. On the role of cost-effectiveness thresholds in healthcare priority setting. *Int J Technol Assess Health Care*. 2021;37:e23.
59. Paulden M, O'Mahony J, McCabe C. Determinants of change in the cost-effectiveness threshold. *Med Decis Mak*. 2017;37:264–76.
60. Lomas J, Claxton K, Martin S, Soares M. Resolving the “cost-effective but unaffordable” paradox: estimating the health opportunity costs of nonmarginal budget impacts. *Value in Health*. 2018;21:266–75.
61. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health-care technologies. *Health Econ*. 2011;20:2–15.
62. Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. Discounting in economic evaluations: stepping forward towards optimal decision rules. *Health Econ*. 2007;16:307–17.
63. Hernandez-Villafuerte K, Zamora B, Feng Y, Parkin D, Devlin N, Towse A. Exploring variations in the opportunity cost cost-effectiveness threshold by clinical area: results from a feasibility study in England. OHE Research Paper. 2019. <https://www.ohe.org/publications/exploring-variations-opportunity-cost-cost-effectiveness-threshold-clinical-area>. Accessed 25 Apr 2022.
64. Charlton V. NICE and fair? Health technology assessment policy under the UK's National Institute for Health and Care Excellence, 1999–2018. *Health Care Anal* [Internet]. 2019. <https://doi.org/10.1007/s10728-019-00381-x>.
65. Reckers-Droog VT, van Exel NJA, Brouwer WBF. Looking back and moving forward: on the application of proportional shortfall in healthcare priority setting in the Netherlands. *Health Policy*. 2018;122:621–9.
66. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables. *BMJ Br Med J Publ Group*. 1993;306:570–2.
67. Epstein DM, Chalabi Z, Claxton K, Sculpher M. Efficiency, equity, and budgetary policies: informing decisions using mathematical programming. *Med Decis Mak*. 2007;27:128–37.
68. Claxton K, Lomas J, Martin S. The impact of NHS expenditure on health outcomes in England: alternative approaches to identification in all-cause and disease specific models of mortality. *Health Econ*. 2018;27:1017–23.
69. Bojke C, Castelli A, Grašič K, Street A. Productivity growth in the English National Health Service from 1998/1999 to 2013/2014. *Health Econ*. 2017;26:547–65.
70. Gallet CA, Doucouliagos H. The impact of healthcare spending on health outcomes: a meta-regression analysis. *Soc Sci Med*. 2017;179:9–17.
71. Guindon GE, Contoyannis P. A second look at pharmaceutical spending as determinants of health outcomes in Canada. *Health Econ*. 2012;21:1477–95.
72. Ivashchenko O. The patterns and determinants of longevity in Russia's regions: evidence from panel data. *J Comp Econ*. 2005;33:788–813.
73. Crémieux P-Y, Meilleur M-C, Ouellette P, Petit P, Zelder M, Potvin K. Public and private pharmaceutical spending as determinants of health outcomes in Canada. *Health Econ*. 2005;14:107–16.
74. Nakamura R, Lomas J, Claxton K, Bokhari F, Moreno-Serra R, Suhreke M. Assessing the Impact of Health Care Expenditures on Mortality Using Cross-Country Data. *World Scientific Series in Global Health Economics and Public Policy* [Internet]. WORLD SCIENTIFIC; 2020 [cited 2020 Aug 6]. p. 3–49. https://doi.org/10.1142/9789813272378_0001.
75. Wickens MR, Breusch TS. Dynamic specification, the long-run and the estimation of transformed regression models. *Econ J*. 1988;98:189–205.
76. Hendry DF, Pagan AR, Sargan JD. Chapter 18 dynamic specification. *Handbook of Econometrics*, vol 2. Elsevier; 1984. p. 1023–1100.
77. Banerjee A, Dolado JJ, Hendry DF, Smith GW. Exploring equilibrium relationships in econometrics through static models: some monte carlo evidence*. *Oxf Bull Econ Stat*. 1986;48:253–77.
78. Lee Y. Bias in dynamic panel models under time series misspecification. *J Econometr*. 2012;169:54–60.
79. Nagendran M, Kiew G, Raine R, Atun R, Maruthappu M. Financial performance of English NHS trusts and variation in clinical outcomes: a longitudinal observational study. *BMJ Open*. 2019;9:e021854.
80. Soares MO, Sculpher MJ, Claxton K. Health opportunity costs: assessing the implications of uncertainty using elicitation methods with experts. *Med Decis Mak*. 2020;40:448–59.
81. Boseley S. Patients suffer when NHS buys expensive new drugs, says report. *The Guardian* [Internet]. 2015 Feb 19 [cited 2019 Oct 10]. <https://www.theguardian.com/society/2015/feb/19/nhs-buys-expensive-new-drugs-nice-york-karl-claxton-nice>. Accessed 25 Apr 2022.
82. Daniels N, Sabin JE. *Accountability for reasonableness. Setting limits fairly: can we learn to share medical resources?* Oxford: Oxford University Press; 2002.
83. Remme M, Martinez-Alvarez M, Vassall A. Cost-effectiveness thresholds in global health: taking a multisectoral perspective. *Value Health*. 2017;20:699–704.
84. Mitton C, Seixas BV, Peacock S, Burgess M, Bryan S. Health technology assessment as part of a broader process for priority

- setting and resource allocation. *Appl Health Econ Health Policy* [Internet]. 2019. <https://doi.org/10.1007/s40258-019-00488-1> (cited 2019 Aug 16).
85. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ*. 2004;13:437–52.
 86. Woods B, Fox A, Sculpher M, Claxton K. Estimating the shares of the value of branded pharmaceuticals accruing to manufacturers and to patients served by health systems. *Health Econ*. 2021;30:2649–66.
 87. Polisen J, Clifford T, Elshaug AG, Mitton C, Russell E, Skidmore B. Case studies that illustrate disinvestment and resource allocation decision-making processes in health care: a systematic review. *Int J Technol Assess Health Care*. 2013;29:174–84.
 88. Chen TC, Wanniarachige D, Murphy S, Lockhart K, O'Mahony J. Surveying the cost-effectiveness of the 20 procedures with the largest public health services waiting lists in Ireland: implications for Ireland's cost-effectiveness threshold. *Value Health*. 2018;21:897–904.
 89. Chi Y-L, Blecher M, Chalkidou K, Culyer A, Claxton K, Edeka I, et al. What next after GDP-based cost-effectiveness thresholds? *Gates Open Res*. 2020;4:176.
 90. Brouwer W, van Baal P, van Exel J, Versteegh M. When is it too expensive? Cost-effectiveness thresholds and health care decision-making. *Eur J Health Econ*. 2019;20:175–80.
 91. García-Lorenzo B, Vallejo-Torres L, del Trujillo-Martín MM, Perestelo-Pérez L, Valcárcel-Nazco C, Serrano Aguilar P. Evaluación económica busca umbral para apoyar la toma de decisiones. *Rev Esp Salud Publica*. 2015;89:537–44.
 92. Sculpher MJ, Basu A, Kuntz KM, Meltzer DO. Reflecting uncertainty in cost-effectiveness analysis. In: Neumann PJ, Ganiats TG, Russell LB, Sanders GD, Siegel JE, editors. *Cost-effectiveness in health and medicine*. Oxford University Press; 2016. p. 289–318.
 93. Pearson SD. The ICER value framework: integrating cost effectiveness and affordability in the assessment of health care value. *Value Health*. 2018;21:258–65.
 94. Refoios Camejo R, Miraldo M, Rutten F. Cost-effectiveness and dynamic efficiency: does the solution lie within? *Value Health*. 2017;20:240–3.
 95. Charlton V, Rid A. Innovation as a value in healthcare priority-setting: the UK experience. *Soc Just Res*. 2019;32:208–38.
 96. Caro JJ, Brazier JE, Karnon J, Kolominsky-Rabas P, McGuire AJ, Nord E, et al. Determining value in health technology assessment: stay the course or tack away? *Pharmacoeconomics* [Internet]. 2018. <https://doi.org/10.1007/s40273-018-0742-2> (cited 2018 Nov 13).
 97. Karlsberg Schaffer S, Sussex J, Hughes D, Devlin N. Opportunity costs and local health service spending decisions: a qualitative study from Wales. *BMC Health Serv Res*. 2016;16:103.
 98. Schaffer SK, Cubi-Molla P, Devlin N, Towse A. Shaping the research agenda to estimate cost-effectiveness thresholds for decision making. OHE Consulting Report. 2016. <https://www.ohe.org/publications/shaping-research-agenda-estimate-cost-effectiveness-thresholds-decision-making>. Accessed 25 Apr 2022.