Empirical Estimates of the Marginal Cost of Health Produced by a Healthcare System: Methodological Considerations from Country Level Estimates

Laura C. Edney, James Lomas, Jonathan Karnon, Laura Vallejo-Torres, Niek Stadhouders, Jonathan Siverskog, Mike Paulden, Ijeoma P. Edoka & Jessica Ochalek

Abstract

Many health technology assessment committees AQ1 have an explicit or implicit reference value (often referred to as a 'threshold') below which new health technologies or interventions are considered value for money. The basis for these reference values is unclear but one argument is that it should be based on the health opportunity costs of funding decisions. Empirical estimates of the marginal cost per unit of health produced by a healthcare system have been proposed to capture the health opportunity costs of new funding decisions. Based on a systematic search, we identified eight studies that have sought to estimate a reference value through empirical AQ2 estimation of the marginal cost per unit of health produced by a healthcare system for the UK, Spain, Australia, The Netherlands, Sweden, South Africa and China. We review these eight studies to provide an overview of the key methodological approaches taken to estimate the marginal cost per unit of health produced by the healthcare system with the aim to help inform future estimates for additional countries. The lead author for each of these papers was invited to contribute to the current paper to ensure all the key methodological issues encountered were appropriately captured. These included consideration of the key variables required and their measurement, accounting for endogeneity of spending to health outcomes, the inclusion of lagged spending, discounting and future costs, the use of analytical weights, level of disease aggregation, expected duration of health gains, and modelling approaches to estimating mortality and morbidity effects of health spending. Subsequent research estimates for additional countries should (1) carefully consider the specific context and data available, (2) clearly and transparently report the assumptions made and include stakeholder perspectives on their appropriateness and acceptability, and (3) assess the sensitivity of the preferred central estimate to these assumptions.

1 Introduction

The estimated costs and effects of investments in healthcare are used to guide funding decisions, but this approach is limited if the health opportunity cost of an investment is unknown. Under a constrained budget, the health opportunity cost of a new investment is the health lost elsewhere from reducing funding to an existing service. An estimate of health opportunity cost can therefore allow decision makers to invest in new health technologies or interventions that are expected to generate net health gains, allowing for the expected health

gains forgone elsewhere in the healthcare system, thus ensuring efficient reimbursement decisions when the goal is to improve population health [2].

Precisely which healthcare intervention(s) are forgone when a new intervention is funded is rarely known. Empirically estimating the marginal cost per unit of health produced by the healthcare system offers a practical alternative to determine an expectation on health opportunity costs. Seminal work from Claxton et al. [1], building on prior work by Martin et al. [2, 3], empirically estimated the health opportunity costs from funding decisions in the English National Health Service (NHS) in this way. This has been followed by estimates in Spain [4], Australia [5], The Netherlands [6, 7], Sweden [8], South Africa [9] and China [10], which all employ different methodological approaches based on available data. While such estimates may be constrained by uncertainty in the data and the methodological approaches taken, they can be explicit about their uncertainty, the assumptions made and the directional impact these may have on the estimated marginal cost per health unit.

This paper provides an overview of previously published methods used to estimate the marginal cost per unit of health produced by the healthcare system. We include discussion of the different empirical approaches taken and discuss these, considering the data available, to help inform future empirical estimates for a country's healthcare system. We include in our discussion the eight papers published to date that we have identified that have estimated the marginal cost per unit of health produced by a healthcare system for seven different countries. These papers were identified through a systematic review of the literature (see Electronic Supplementary Material [ESM] 1). Studies were included that assessed the impact of healthcare spending on health outcomes within a country and translated the results into a cost per quality-adjusted life-year (QALY) or disability-adjusted life-year (DALY) estimate. Studies that sought to estimate the relationship based on cross-country data, sought to estimate the cost per QALY or DALY based on the estimated relationship between spending and outcomes estimated for another country, and those that were not peer reviewed, published in English, or published prior to 2008 were excluded. This led to the final eight studies included in the review. The lead author for each of these papers was invited to contribute to the current paper to ensure all the key methodological issues encountered could be appropriately captured.

Across all eight studies, approaches to estimating the cost per unit of health can be split into two parts: modelling population-level health outcomes against health spending and other control variables to estimate the health spending elasticity; and modelling to extrapolate the estimated effect to impact on a lifetime generic measure of health. We compare and contrast the approaches taken to address key methodological issues critical to both parts. Key issues are those that were identified by the study authors and therefore are also likely to be relevant to researchers wishing to undertake similar research.

Throughout this paper, we refer to the marginal cost per unit of health within the healthcare system, rather than using more conventional terms such as 'threshold', as decision making thresholds are reference values that may or may not reflect health opportunity costs. Thresholds used to inform funding decisions and to draw recommendations in the published literature may reflect a range of considerations other than opportunity costs [11]. In particular, there is an extended view, grounded in welfare economics, that cost-effectiveness thresholds ought to reflect the society's monetary valuation of health gains. Some authors have recently emphasised that decision rules are context-dependent and differ by the perspective taken by decisions

makers and by the budget constraints, whether fixed or variable, faced by them [12, 13]. According to the two-perspective approach framework presented by Brouwer et al., information on the consumption value of health is relevant when decision makers take a broader societal perspective, in which the goal is to maximise social welfare from a flexible budget. However, in the most commonly operating context where fixed budgets are allocated to healthcare and coverage decisions are taken from a healthcare system perspective, information on the health opportunity cost of healthcare funding decisions becomes the relevant information to inform cost-effectiveness thresholds [13].

2 General Challenges to Empirically Estimate the Marginal Cost of Health Produced by a Healthcare System

Estimating the relationship between healthcare spending and health outcomes presents several challenges. The key component to addressing this research question is to identify variation in health spending that is unrelated to variation in health status, and to then estimate the health effects of such exogenous variation. In an ideal world, researchers would link the exogenous variations in healthcare spending to all the affected individuals and calculate the resulting changes in health over their lifetimes. Information on individual-level health spending across all areas of healthcare would be ideal. If the aim is to assess the impact of public spending on health, as would be the case for health opportunity cost estimates attempting to guide public reimbursement decisions, then additional health spending information such as private insurance and patient out-of-pocket spending are relevant as model covariates. The perfect data on health outcomes would include individual-level cause of death and health-related quality of life (HRQoL) data to estimate disease-, age- and sex-specific trajectories in the estimation of mortality- and morbidity-related QALYs. Obtaining exogenous variations in healthcare spending will typically rely on controlling for a large number of observed healthcare need variables and a method to control for any unobserved need and reverse causality. The preferred method to control for this exogenous variation due to both unobserved covariates and reverse causality due to prior health outcomes influencing current health spending is the panel approach. The advantage of this approach is that time-invariant confounding can be removed through the use of multiple cross-sections; however, this approach relies on significant data availability. In the absence of multiple cross-sections of data, instrumental variable (IV) estimation is an alternative approach that has been employed. The use of IVs, discussed further in Sect. 4.2.3 ‘Accounting for Endogeneity’ and in ESM 2, offers an approach to making unbiased causal inferences from observational data by controlling for reverse causality and any unobserved confounding variables. However, it is limited by the quality of the IVs and places a large burden on the researchers to present evidence for the validity of their IVs that cannot be conclusively supported with empirical evidence. Good data are therefore required not only on healthcare spending and health outcomes but also on variables to control for healthcare need, and potentially on candidate IVs.

3 Overview of Methodological Approaches

Table 1 summarises the key methodological approaches taken, split into parts one (modelling population-level health outcomes against health spending and other control variables to estimate the health spending elasticity) and two (modelling to extrapolate the estimated effect to impact on a lifetime generic measure of health).

Table 1 Overview of methodological approaches

Most studies estimate the effect of a change in health spending on mortality in part one, and then approximate and incorporate the morbidity effect of health spending to arrive at the

estimated marginal cost per QALY or DALY in part two, using either area [1, 4, 5, 8,9,10] or patient groups [6, 7] as the unit of analysis. Less frequently, a QALY or DALY outcome measure is used in part one. This is discussed in more detail in Sect. 4.1 on the health outcome measures used. In part one, different econometric specifications are used to model the effect of a change in health spending on population-level health outcomes. As outlined in Table 1, the reviewed studies either employed panel data with unit fixed effects or cross-section data with an IV approach. Control variables used (see Table 1) are similar across the studies and are used to minimise confounding due to area- or group-level determinants of health that can also influence change in healthcare spending in longitudinal approaches, or due to unit- or area-level determinants of health that can also influence change in healthcare spending in cross-sectional approaches. Control variables must also account for potential confounding between the IVs and determinants of health in IV analyses (see Sects. 4.2.2 'Controlling for Healthcare Need', 4.2.3 'Accounting for Endogeneity', and ESM 2 for further details). In part two, some studies have assumed the effect on morbidity to be proportional to the effect on mortality, or have aimed at estimating the morbidity effects of health spending directly.

4 Methodological Issues

Estimation of the health effects of healthcare spending requires methodological decisions to be made by researchers that may have substantial impacts on the final results. In this section, we briefly describe the key decisions of this type that were made in the reviewed papers, splitting these into issues encountered when (1) modelling population-level health outcomes, normally a measure of mortality, against health spending and other control variables to estimate the health spending elasticity, and (2) approximating and incorporating the morbidity effect of health spending to arrive at the estimated marginal cost per unit of lifetime health produced. The majority of the methodological issues identified pertain to part one, therefore this received the bulk of the focus; however, some of the key methodological challenges pertain to both parts. With the exception of the health outcome measures used, these are listed under the part one challenges. Health outcome measures used is given its own section and is discussed first. Where suggestions are made, these are from the view of informing decisions about the funding of new health technologies, but it should be noted that an estimate of the effect of healthcare spending on health outcomes could also be of interest from other perspectives.

4.1 Health Outcome Measures

The population-level health outcomes must be relevant to the specific decision-making context. As decisions are made regarding healthcare interventions spanning a range of disease areas, a generic measure of health that accounts for changes in survival and morbidity (i.e., quality of life) is required, which, in the reviewed studies, was the QALY and DALY. QALYs and DALYs have been shown to be largely interchangeable [14].

Few countries routinely capture national HRQoL information that could be used to inform estimates of QALYs or DALYs by small geographical areas. Therefore, due to limited data, most approaches reviewed herein have modelled the impact of health spending (part one) on some mortality-related measure and subsequently incorporated morbidity effects (part two) [1, 5, 6, 8]. Other approaches included a combined measure of mortality and morbidity as the key outcome variable in their econometric model (part one) through either the use of quality-adjusted life expectancy (QALE), calculated as life expectancy (LE) adjusted by modelled EQ-5D weights [4], or through QALYs lost due to mortality corrected for disease burden and morbidity [7], potentially negating the need for some or all of part two. A single estimate directly modelled

the effect of health spending on DALYs through publicly available data on the Global Burden of Disease [10]; see ESM 3 for further information.

The most commonly estimated mortality-based health outcome measures were mortality rates [6, 9, 10], years of life lost (YLL) [1, 5, 7] or LE [4, 8] (see Table 1.) YLL reflect survival effects and are constructed from data on mortality and conditional LE or a fixed reference age. Careful consideration should be given to how the mortality-based effects contribute to the estimation of YLL. Where mortality rates are used as an outcome measure instead of YLL, assumptions are required to ultimately obtain survival effects. For example, both the South African [9] and the Chinese mortality-based [10] cost per DALY estimates are derived from the change in mortality rates from increases in health spending using the method outlined by Ochalek et al. [15]. This approach uses YLL averted from public disease burden data in an all-cause model to estimate mortality-based survival effects, and incorporates morbidity-based effects by assuming a proportional impact on direct morbidity effects [9, 10], using the same assumption as the English estimate [1]. Using LE, mortality rates are extrapolated to the future to predict YLL. Crucially, and regardless of the mortality measure used, the survival effects must be adjusted for the HRQoL in which they are expected to be lived. Where outcomes are considered by disease area, it is important that the survival effects reflect disease-specific profiles of survival and HRQoL [1].

To account for the morbidity effects of a change in health spending, studies typically employed an assumption about the effect size relative to the change in mortality or YLL [1, 9, 10]. The Australian study estimated the effect directly using available health index data, making the assumption that temporal change in HRQoL, controlling for demographic, societal and other economic variables, was due to change in health spending over the same time period [5]. The English study constructed a measure of the QALY burden of disease at the national level by making the assumption that the effect of change in spending on mortality provided a surrogate for the effect of change in spending on the QALY burden of disease estimated using variation across local health authorities (see ESM 3 for further details). Other studies used a health outcome measure in their econometric models that already reflected changes in mortality and on HRQoL, and thus estimated the effect of spending on survival and morbidity simultaneously [4, 7]. Both approaches used EQ-5D weights to estimate morbidity effects, making additional assumptions about the extent to which an annual survey sample was representative of the Dutch population [7] or modelling temporal change in HRQoL due to limited national-level HRQoL information [4].

4.2 Part One: Modelling Population-Level Health Outcomes Against Health Spending

4.2.1 Health Spending

The measurement of health spending involves the estimation of health spending within defined geographical areas, for which accompanying variables describing health needs and health outcomes are required. If the goal is to inform public reimbursement decisions through the estimation of the opportunity costs of publicly funded health spending, then the main spending independent variable should reflect public spending on health. However, private health spending should be included as an important model covariate as its omission may result in biased estimates on health outcomes where the coefficient on health outcomes may be underestimated (overestimated) if the relationship between public and private spending on health is negative (positive). The preferred approach is to include private health spending as a covariate in part one. Approaches to accounting for the impact of private healthcare spending

on health outcomes in the absence of private health spending data have included using socioeconomic variables as a proxy for private health insurance coverage and spending within the regression model (e.g. Edney et al. [5] and Edoka and Stacey [9]). If total public spending on health is unavailable, then the impact of missing cost categories, either by health service or patient type, on the estimated health spending elasticity must be considered. If the impact is deemed negligible then this should be explicitly justified [5]. Alternatively, conclusions will need to be restricted to the types of healthcare spending included; for example, by estimating the marginal cost per unit of health of hospital-based care [6, 7].

How health spending is measured, as well as characteristics of the specific local context, will also impact on the appropriate econometric models; for example, the English analysis is the only approach to have employed separate disease-specific spending models, reflecting that expenditure data were available for each disease area, with mortality data available by cause of death [1].

4.2.2 Controlling for Healthcare Need

The need for healthcare, defined in relation to the capacity to benefit from healthcare, not only clearly affects health outcomes but can also predict health spending. Variables used to represent healthcare need have included measures of health status, past healthcare use, health supply, and demographic variables such as population size and proportion of elderly people. This is not only most obvious in countries with funding mechanisms that allocate resources on the basis of needs (e.g., England) but is also likely in countries in which resources are not allocated solely on the basis of need. Therefore, healthcare need must be included in the model to estimate the unbiased causal relationship between healthcare spending and outcomes.

While census data, as used in England and Australia, can provide population measures of need via health status or socioeconomic questions, it is typically obtained infrequently, meaning that analyses are restricted to census years [5] or outdated information is used outside of census years [1]. In the context of a model looking at variations over time [4, 7, 9, 16], longitudinal survey or administrative data have been used to account for health need with the addition of fixed effects, including time and age- and sex-specific time trends, to account for any remaining unobserved health need [4, 9].

4.2.3 Accounting for Endogeneity

Regressing outcomes against spending, controlling for differences in health needs, is unlikely to produce estimates of causal effects. There are two principal reasons for this: (1) there are likely unobserved area-specific confounders, that is, omitted variables that are associated with both spending and outcomes, resulting in omitted variable bias; and (2) there may be reverse causality whereby historic health outcomes impact current health spending and current health outcomes (see ESM 4 for a figure depicting the potential temporal causal relationships between health spending and health outcomes). In the context of these factors, spending is endogenous [17] and the resultant coefficient on spending is biased.

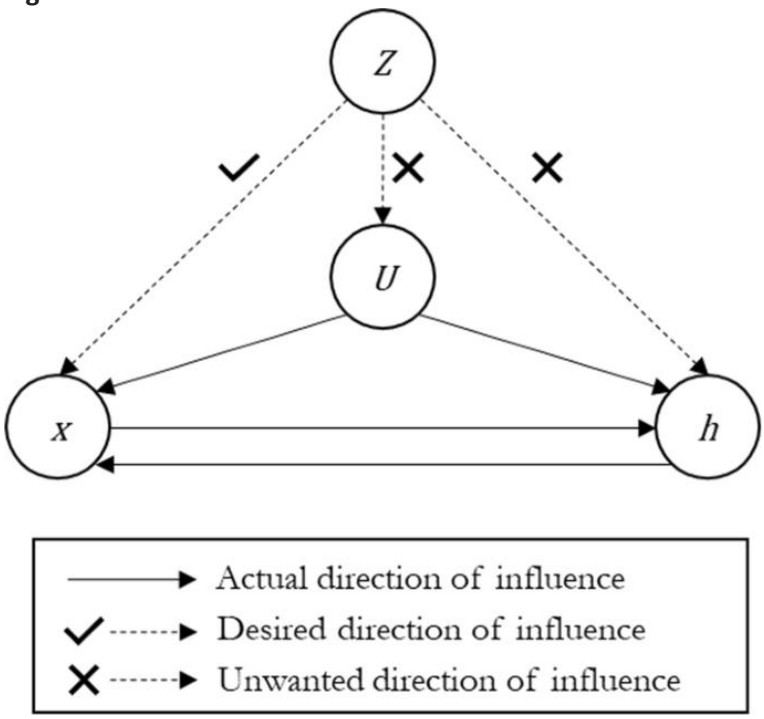
Two popular approaches to accounting for endogeneity that have been employed in the country-level estimates reviewed here include the use of panel data and IVs (see Table 2). Panel data can address endogeneity through eliminating additional unobserved confounding by controlling for time-invariant region effects and time-varying year effects and mitigating some

reverse causality through inclusion of lagged spending or health outcomes in a dynamic panel data analysis [18]. IV regression methods may be particularly useful when panel data are unavailable, or, as in the Spanish approach [4], to assess whether the panel approach has appropriately accounted for endogeneity through evaluation of IV tests of exogeneity of health spending under the assumption that at least one of the set of included instruments is valid (see ESM 2 for further details on IV assumptions). IV regression is undertaken in five [1, 4, 5, 8, 10] of the eight studies reviewed and provides the base-case results for England [1], Australia [5] and Sweden [8].

Table 2 Key predictors and methods to account for endogeneity

An appropriate IV is one that not only has a direct effect on health spending but has no direct or indirect impact on health outcomes, other than through its effect on health spending and other variables included in the outcome equation, i.e., a relevant IV that meets the exclusion restriction. Assuming that observable factors have been adequately controlled for, this can be represented by the Directed Acyclic Graph (DAG) in Fig. 1, where the checked (✔) dotted arrow represents the relevance of the IV required for successful identification and the unchecked (×) dotted arrows represent relationships that would violate the exclusion restriction, thus invalidating the IV. In practice, validating IVs requires extensive theoretical and empirical justification. We outline approaches that could be included in this process in Table 1 of ESM 2.

Fig. 1



Directed acyclic graph (DAG) depicting the role of an instrumental variable (Z) in estimating the causal relationship between health spending (X) and health outcomes (h) with unobserved covariates (U)

Variables that have been employed across the five relevant studies as instruments for public spending on healthcare have included demand-shift IVs such as the proportion of households where unpaid care was provided [1, 5] and the percentage of total public spending assigned to health regions [4], and supply-side shifts, including the number of newly graduated nurses per capita and the proportion of nurses nearing retirement [8] [see Table 2]. Other studies included in this review accounted for endogeneity by employing fixed effects [4, 6, 9] to control for time-

invariant region differences and time-varying year differences not captured by their models. However, these approaches may still suffer contamination due to time-varying unobserved confounding in health status unrelated to measured confounders. The Spanish team [4] employed an IV, the percentage of health spending allocated to healthcare, to assess this. This IV was considered potentially exogenous due to large budget cuts during the study period. However, their analyses indicated that health spending was not endogenous in the health outcome equation, and therefore the results from the fixed effects regression without an IV were used for their central estimate.

Another non-IV approach [7] employed fixed effects and accounted for remaining endogeneity by removing spending incurred in the last year of life from the spending variable. This approach may account for one source of reverse causality under the assumption that surplus costs in the last year of life do not affect mortality, which, as noted by the authors, is more applicable to some disease areas than others. This approach only addresses one source of endogeneity, while other potential time-varying sources of endogeneity are left unaccounted for. For this reason, this approach may be considered when regional variation and valid IVs are unavailable.

Promising for future research in this area are IVs arising from consideration of context-specific national funding arrangements. Using this approach, Andrews et al. identified three new IVs for English estimates: the age index of the local population; the input price index; and the difference between the actual allocation compared with the calculated budget requirement to meet regional healthcare need (allowed to avoid large shocks in allocations following new estimates of need) [19]. These variables are used to determine health funding allocations for regions in England, and therefore, in conjunction with controls for healthcare need and with an age-standardised dependent variable, "… can only plausibly work through the funding rule …" (see Martin et al. [20] for further justification of these specific IVs) [19]. Analyses with these IVs returned similar values to the central estimate reported previously [1] and in reanalysis [20, 21], supporting the estimate of the marginal cost per unit of health produced by the NHS in England. This highlights the importance of understanding the local context when considering appropriate IVs where components of funding rules can be separated from considerations based on demand.

4.2.4 Analytical Weights

To the extent that data observations such as regions [1, 4, 5, 8, 10] have different population sizes or demographic and disease groupings [6, 7] have different sample sizes, then analytical weights may be employed to correct for heteroskedastic error terms and to identify the population-weighted mean effects of health spending on health outcomes [22]. Not weighting gives equal influence to all observations irrespective of the size of the population contained; this does not obviously produce biased estimates, and the degree to which this modelling decision impacts results depends on the amount of heterogeneity between observations. Application of analytical weights should be based on consideration of this between-observation heterogeneity and on the intended use of the estimate of the marginal cost per unit of health produced by the healthcare system. If the aim is to obtain a national average effect, then the application of analytical weights may be preferred; however, the central estimate of the marginal cost per unit of health produced by the healthcare system may then be driven by the level of efficiency within the largest data observations and may not represent the opportunity cost of funding decisions for other regions. Since a difference between weighted and unweighted estimates complicates their interpretation, analytic weights can also be used as a robustness check in which we would hope for results that are fairly insensitive to weighting [23]. We recommend authors present

both the weighted and unweighted estimates, discuss any differences, report robust standard errors and provide clear rationale on the selected approach [22].

4.2.5 All-Cause or Disease-Specific Models

The analysis for England differs from the other studies in that different disease areas were analysed separately by regions [1]. The remaining studies either estimated the marginal cost per unit of health within a single disease area [6], multiple disease areas [7], or across regions [4, 5, 10]. There are advantages and disadvantages to each approach. Disease-specific estimation may better identify the underlying causal effects that may be subject to aggregation bias in all-cause estimation. Therefore, to the extent that the relationship between health spending and outcomes is heterogeneous between disease areas or demographic groups, then the all-cause spending elasticity may differ from the aggregated disease-specific spending elasticity. However, this approach may also introduce measurement error based on methods for classifying spending and outcomes to disease areas and does not reflect the effects of spending and health outcome spillover effects that may occur between disease areas.

4.2.6 Lagged Effects of Healthcare Spending

Estimates of the marginal cost per unit of health produced by the healthcare system for England, Australia and Sweden largely rely on cross-sectional data that comprises contemporaneous spending and outcomes. As such, there is no explicit consideration regarding the future effect on outcomes of spending within the econometric analysis. However, the estimated effect on outcomes of contemporaneous spending will, in principle, equal the long-run effect of spending on health if an equilibrium position has been reached where the effect on current outcomes from past spending exactly offset the effect of current spending on future outcomes. Claxton et al. [24] argue that, in practice, such an equilibrium position would mean that the estimated coefficient on spending represents an underestimate of the long-run effect, because the effect on future outcomes is not captured in the model, whereas the effect of past spending on current outcomes could affect the coefficient where past spending and current spending are correlated. This is supported by the Spanish analysis, which reported a reduced impact of spending on health outcomes when excluding lagged spending, resulting in an increase in the estimated marginal cost per unit of health produced by the healthcare system [4]. Similarly, van Baal et al. [6] reported that the impact of cardiovascular disease (CVD)-related health spending was significant in the subsequent year [6], highlighting that different care types and disease areas are likely to have different lag structures from spending on health. Taken together, these results suggest that estimates based solely on contemporaneous effects of spending on health outcomes may overestimate the marginal cost per unit of health produced by the healthcare system.

4.3 Part Two: Approximating and Incorporating the Morbidity Effect of Health Spending to Arrive at the Estimated Marginal Cost Per Unit of Lifetime Health Produced

4.3.1 Future Healthcare Costs

Healthcare spending in one time period is likely to generate health benefits over multiple future time periods, and health benefits in one time period are likely to impact health spending over multiple future time periods. Therefore, a decision must be made about how future costs and benefits should be discounted to reflect their present value. In general, higher discount rates will result in a higher estimate of the marginal cost per unit of health produced by the healthcare system. Discount rates employed across the eight studies have ranged from 0 [1] to 5% [5]; we

suggest that discount rates for both future costs and benefits should be in accordance with the relevant national guidelines for economic evaluations [5, 7], but that sensitivity estimates also be presented for a range of discount rates including the application of none.

Extending survival means that patients will receive healthcare in these years and thus incur additional spending; conversely, reducing morbidity may lessen demands for healthcare in future years. If future costs are not incorporated, then the implicit assumption is that gaining a QALY is cost neutral. Only one study considered future costs through incorporating CVD hospital spending and hospital spending for other diseases by age and sex in their calculation of per capita hospital spending [6]. Debating the appropriateness of considering future unrelated costs is outside the scope of this paper; when the estimate of the marginal cost per unit of health produced by the healthcare system is sought to help guide funding decisions, then we recommend applying the same approach regarding future costs as that required by the relevant national health technology assessment committees.

4.3.2 Expected Duration of Health Gains

The expected duration of the estimated health effects should also be considered and included in the final estimate where feasible. If a death is avoided in the present year, then a stream of health effects are generated up until the eventual delayed death. Only the Australian analysis included estimates of the duration of morbidity-related QALY effects based on conservative clinician input into duration effects within disease areas and accounting for age- and sex-specific LEs of the Australian population. From this, an aggregated average length of improvement from interventions delivered by health spending in the year of analysis was calculated [5]. Insofar as the impact of current health spending on morbidity-related QALYs extends beyond a single year, the impact of spending in prior years on current and subsequent morbidity-related QALYs to the value of the estimated duration effects should also be removed from the final morbidity-related QALY improvement. The Australian study provides one method to estimate the duration of morbidity-related QALY gains by adjusting the annual change in HRQoL by excluding ongoing effects of health spending in the years prior to the year of interest and including ongoing effects of spending on HRQoL in years subsequent to the year of interest using clinical judgement on the duration of HRQoL improvement from the annual increase in health spending [5] (see ESM 3 for further details).

4.4 Presenting Uncertainty in the Estimated Marginal Cost Per Unit of Health Produced by the Healthcare System

Estimates of the marginal cost per unit of health produced by the healthcare system are all subject to parameter and structural uncertainty, which must be clearly reported. For parameter uncertainty, this includes reporting confidence intervals around the estimated marginal cost per unit of health produced by the healthcare system and the probability that the estimate is below different cut-off values. Different estimation will require different methods to estimate confidence intervals. For example, the English analysis included a spending elasticity estimated for each of the 23 disease areas, and outcome elasticities estimated for 11 disease areas; therefore they employed simulation analyses to propagate the uncertainty in the parameter estimates through to the final estimate [1]. Other analyses estimated a single econometric model for change in mortality-related QALYs [4,5,6,7] and presented the uncertainty in the parameter estimates via bootstrapped confidence intervals [5].

Structural uncertainty refers to "… simplifications and scientific judgements that have to be made when constructing and interpreting a model of any sort…" [25]. In the context of this review, a key source of structural uncertainty is the set of choices made in estimating plausible econometric models and determining the preferred central estimate based on these choices, which includes the strategy chosen to identify causal effects. The nature of these choices depends on the approach taken; in the case of the IV approaches undertaken in four of the eight reviewed studies [1, 4, 5, 8], this includes reporting tests of IV relevance and of the exclusion restriction, including sensitivity analyses. For potential violations of the exclusion restriction when the IV is only plausibly exogenous, methods such as the Union of Confidence Intervals approach [26] should also be reported [1] (see ESM 2 for more information on testing IV assumptions).

Stadhouders et al. distinguished parameter uncertainty, data transformation uncertainty and structural uncertainty with respect to modelling choices [7]. Parameter uncertainty was presented via bootstrapped confidence intervals; uncertainty with respect to the values used in data transformations (e.g., YLL values) was incorporated using Monte Carlo simulations, and structural uncertainty regarding modelling choices was incorporated by presenting outcomes of alternative specifications. Several other approaches also used simulation methods to reflect the combined uncertainty in estimates, reported as the probability of the marginal cost per unit of health produced by the healthcare system being below a certain cut-off [1, 5].

Finally, the full set of assumptions made to link the econometrics to the estimate of the the marginal cost per unit of health produced by the healthcare system should be outlined in detail, including assumptions related to the (1) measurement of the key healthcare spending, mortality and morbidity outcomes, and healthcare need variables; (2) econometric modelling; and (3) incorporation of morbidity effects. This should include an outline of the issue, what assumptions were made and the associated impact on the estimate of the marginal cost per unit of health produced by the healthcare system, based either on empirical estimates or the theorised impact as reported by Claxton et al. [1] and Edney et al. [5]. Some assumptions will have limited formal empirical data to inform them; where feasible, we recommend authors attempt to validate these types of assumptions with, for example, broader expert elicitation, as conducted for the English estimate [27].

5 Discussion

Estimates of the marginal cost per unit of health produced by the healthcare system present an important policy tool to guide funding decisions for new health technologies and interventions by facilitating the interpretation of cost-effectiveness analyses that represent a key input to many national reimbursement decisions.[Footnote1] Building on earlier work by Martin et al. [2, 3], the seminal work by Claxton et al. [1] presented the first empirical estimate of the marginal cost per unit of health of the NHS in England, followed by similar estimates for Spain [4], Australia [5], Sweden [8], The Netherlands [6, 7], South Africa [9] and China [10]. The broad rationale for estimating the marginal cost per unit of health produced by the healthcare system in these reviewed studies was to guide health funding decisions. Where the true opportunity cost is unknown, that is, when decisions to fund a new health technology or intervention do not specify what is to be forgone, then an empirical estimate can provide the next best option to quantify the net health impact.

We have purposefully referred to the estimates in the reviewed papers as estimates of the marginal cost per unit of health produced by the healthcare system rather than the more conventionally used 'threshold', as a threshold implies that only new health technologies or interventions with an incremental cost-effectiveness ratio below this should be publicly funded. In reality, rational decision makers may fund healthcare with a higher incremental cost-effectiveness ratio than the estimated marginal cost per unit of health because the 'unit of health', for example the QALY, may not reflect all factors that influence funding decisions, such as their social value, innovation, rarity or impact on outcome equity [30]. An estimate of the marginal cost per unit of health produced by the healthcare system simply allows decision makers to articulate the expected net health benefit from publicly funding new health technologies or interventions. Presentation of net health benefit is intuitive, can be expressed at the population level to communicate health magnitudes and budget impact, and can provide a transparent method to justify funding decisions, particularly for those that result in a net health loss that are accepted based on other criteria [31]. Analyses by disease areas may be further useful in this scenario as they can provide decision makers with an indication of where the opportunity costs may fall [1], thus highlighting when net health losses also fall on patient groups with similar equity considerations. Such transparency in decision making becomes more important as national reimbursement authorities come under increasing scrutiny from industry, patients, and the broader community. Empirical estimates of the marginal cost per unit of health produced by the healthcare system can also be used to guide decisions about other financial investments within the healthcare sector including the value of health research through the expected value of perfect information and of implementation programmes to increase uptake of health interventions with well-established health benefits.

The studies reviewed here have all taken different approaches and made different assumptions in order to generate estimates of the marginal cost per unit of health produced by the healthcare system, driven largely by the local decision making context and data availability[Footnote2]. Given heterogeneity across existing and future studies, it is important that all assumptions are explicitly stated and the robustness of the estimated marginal cost per unit of health to specific assumptions is examined through extensive sensitivity analysis. The acceptability of different assumptions will likely vary across different decision-making contexts, therefore key stakeholders should be involved in the research where feasible, and researchers should provide estimates based on a range of different assumptions relevant to the specific decision-making context. Our combined experience suggests that involving stakeholders is important throughout and their input is particularly valuable on issues such as knowledge of available data, acceptability of the assumptions made and knowledge of the local context. Factors relevant to the local context include the health budget, the demand for and technical efficiency of existing technologies, development of new technologies [33], the specific decision-making context, how healthcare is financed, the different factors that may influence healthcare spending and health outcomes, and potential policy changes or health system shocks, such as the coronavirus disease 2019 (COVID-19) pandemic, that may impact conclusions drawn. While these factors will not influence the analysis of the marginal cost per unit of health produced by the healthcare system, they will influence the interpretation and long-term applicability of the estimate. Researchers should carefully discuss the likely directional impact of any policy changes or health system shocks on their central estimate of the marginal cost per unit of health produced by the healthcare system.

Transparency in reporting results and in determining the preferred central estimate, if one is presented, helps decision makers interpret research findings, particularly where the range of plausible estimates is wide. Probabilistic analyses can be used to represent parameter and

structural uncertainty by estimating probabilities that the marginal cost per unit of health is below different values. The expected or hypothesised direction and magnitude of the impact of different assumptions on the central estimate should also be explicitly stated, with reference to the rationale for the expectations, and supporting empirical or qualitative evidence where appropriate. Uptake of the estimated marginal cost per unit of health by decision makers may be supported by comparisons with existing 'cost-effectiveness thresholds', and the extent to which the new and existing estimates are evidence-based.

6 Conclusion

There is an increasing need for resource allocation decisions to be transparent to all key stakeholders, as tight fiscal budgets are increasingly challenged by high-cost technologies and interventions, increasing LE and expectations regarding the public funding of new healthcare. An empirical estimate of the marginal cost per unit of health produced by the healthcare system can provide a key method to inform, justify and legitimise funding decisions. Such estimates make trade-offs between population health maximisation and health equity explicit by quantifying the expected population health effects of decisions to fund healthcare for particular patient groups [31]. Estimates of the marginal cost per unit of health also inform the monitoring of health system efficiency and the funding of research [34], implementation processes and disinvestment activities. Finally, such estimates can also simplify the process of making resource allocation decisions for decision makers and simplify the process of communicating how these resource allocation decisions were made to key stakeholders through use of the net health impact from new investments [30, 35]. Empirical estimates of the marginal cost per unit of health produced by the healthcare system derived from within-country analyses now exist for seven countries internationally based on eight studies. It is anticipated that many other countries may wish to empirically estimate the health opportunity costs of reimbursement decisions; this paper provides the first discussion of the range of approaches taken within these published papers to empirically estimate health opportunity costs of funding decisions to guide these future research attempts. Specific recommendations for future country estimates include careful consideration of the specific context and data available, clarity and transparency regarding the assumptions made and stakeholder perspectives on their appropriateness and acceptability, and assessment of the sensitivity of the preferred central estimate to these assumptions**.**

Notes

1. For local decisions, the marginal cost per unit of health produced by the local healthcare system could be estimated by exploring the heterogeneity between geographically-defined observations. This is an area for further research that may draw inspiration from some existing work that employs quantile regression methods (i.e. Edney et al. [28] and Hernandez-Villafuerte et al. [29]). Such work could also inform national decision making with respect to the geographical distribution of effects of a new health technology.

2. Another study aims to estimate health opportunity costs in the US and finds that 1 QALY is lost for every $104,000 spent on a new technology [32]. This study provides a policy-relevant estimate for the US that merits consideration but differs from the studies reviewed here in two ways. First, it is based on a different approach where private insurance premiums are increased to fund a new technology, with a resultant impact on the health of people dropping insurance coverage (i.e., it does not estimate the marginal cost per unit of health produced by the healthcare system). Second, rather than performing any econometric data analysis, it is instead a simulation exercise that draws on previously published econometric estimates.

References

1. Claxton K, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess. 2015;19(14):1–503.

   Article Google Scholar

2. Martin S, Rice N, Smith PC. Does health care spending improve health outcomes? Evidence from English programme budgeting data. J Health Econ. 2008;27(4):826–42.

   Article Google Scholar

3. Martin S, Rice N, Smith PC. Comparing costs and outcomes across programmes of health care. Health Econ. 2012;21(3):316–37.

   Article Google Scholar

4. Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. Estimating a cost-effectiveness threshold for the Spanish NHS. Health Econ. 2018;27(4):746–61.

   Article Google Scholar

5. Edney LC, et al. Estimating the reference incremental cost-effectiveness ratio for the Australian health system. Pharmacoeconomics. 2018;36(2):239–52.

   Article Google Scholar

6. van Baal P, et al. A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. Health Econ. 2019;28(1):87–100.

   Article Google Scholar

7. Stadhouders N, et al. The marginal benefits of healthcare spending in the Netherlands: estimating cost-effectiveness thresholds using a translog production function. Health Econ. 2019;28(11):1331–44.

   Article Google Scholar

8. Siverskog J, Henriksson M. Estimating the marginal cost of a life year in Sweden's public healthcare sector. Eur J Health Econ. 2019;20(5):751–62.

   Article Google Scholar

9. Edoka IP, Stacey NK. Estimating a cost-effectiveness threshold for health care decision-making in South Africa. Health Policy Plan. 2020;35(5):546–55.

   Article Google Scholar

10. Ochalek J, et al. Informing a cost-effectiveness threshold for health technology assessment in China: a marginal productivity approach. Pharmacoeconomics. 2020;38(12):1319–31.

    Article Google Scholar

11. Sculpher M, Claxton K, Pearson SD. Developing a value framework: the need to reflect the opportunity costs of funding decisions. Value in Health. 2017;20(2):234–9.

Article **Google Scholar**

12. Culyer AJ. Cost, context, and decisions in health economics and health technology assessment. Int J Technol Assess Health Care. 2018;34(5):434–41.

Article **Google Scholar**

13. Brouwer W, et al. When is it too expensive? Cost-effectiveness thresholds and health care decision-making. Springer; 2019.

Book **Google Scholar**

14. Feng X, et al. Using QALYs versus DALYs to measure cost-effectiveness: How much does it matter? Int J Technol Assess Health Care. 2020;36(2):96–103.

Article **Google Scholar**

15. 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(6):e000964.

Article **Google Scholar**

16. van Baal PH, et al. Standardizing the inclusion of indirect medical costs in economic evaluations. Pharmacoeconomics. 2011;29(3):175–87.

Article **Google Scholar**

17. Gravelle H, Backhouse M. International cross-section analysis of the determination of mortality. Soc Sci Med. 1987;25(5):427–41.

Article **CAS** **Google Scholar**

18. Bond S, Dynamic panel data models: a guide to microdata methods and practice. CeMMAP working papers CWP09/02. Centre for Microdata Methods and Practice, Institute for Fiscal Studies; 2002.

19. Andrews M, et al. Inference in the presence of redundant moment conditions and the impact of government health expenditure on health outcomes in England. Economet Rev. 2017;36(1–3):23–41.

Article **Google Scholar**

20. Martin S, et al. How effective is marginal healthcare expenditure? New evidence from England for 2003/04 to 2012/13. Appl Health Econ Health Policy. 2021. https://doi.org/10.1007/s40258-021-00663-3.

Article **PubMed** **Google Scholar**

21. 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(6):1017–23.

Article Google Scholar

22. Solon G, Haider SJ, Wooldridge JM. What are we weighting for? J Human Resour. 2015;50(2):301–16.

Article Google Scholar

23. Angrist JD, Pischke JS. Mastering'metrics: The path from cause to effect. Princeton, NJ: Princeton University Press; 2014.

Google Scholar

24. Claxton K, Sculpher M. Response to the OHE critique of CHE Research paper 81. 2017.

25. Bojke L, et al. Characterizing structural uncertainty in decision analytic models: a review and application of methods. Value Health. 2009;12(5):739–49.

Article Google Scholar

26. Conley TG, Hansen CB, Rossi PE. Plausibly exogenous. Rev Econ Stat. 2012;94(1):260–72.

Article Google Scholar

27. Soares MO, Sculpher M, Claxton KP. Health opportunity costs: assessing the implications of uncertainty using elicitation methods with experts. Med Decis Making. 2020;40(4):448–59.

Article Google Scholar

28. Edney L, et al. Mortality reductions from marginal increases in public spending on health. Health Policy. 2018;122(8):892–9.

Article CAS Google Scholar

29. Hernandez-Villafuerte K, et al. Exploring variations in the opportunity cost cost-effectiveness threshold by clinical area: results from a feasibility study in England. OHE Research Paper. London. Office of Health Economics; 2019.

30. Paulden M. Why it's time to abandon the ICER. Pharmacoeconomics. 2020;38(8):781–4.

Article Google Scholar

31. Siverskog J, Henriksson M. On the role of cost-effectiveness thresholds in healthcare priority setting. Int J Technol Assess Health Care. 2021;37:e23.

Article Google Scholar

32. Vanness DJ, Lomas J, Ahn H. A health opportunity cost threshold for cost-effectiveness analysis in the United States. Ann Intern Med. 2021;174(1):25–32.

Article Google Scholar

33. Paulden M, O'Mahony J, McCabe C. Determinants of change in the cost-effectiveness threshold. Med Decis Mak. 2017;37(2):264–76.

Article Google Scholar

34. Claxton K, Fenwick E, Sculpher MJ. Decision-making with uncertainty: the value of information. In: Jones AM, editor. The Elgar companion to health. Second edition (Chapter 51). Berlin: Edward Elgar Publishing; 2012.

Google Scholar

35. O'Mahony JF. Does cost-effectiveness analysis really need to abandon the incremental cost-effectiveness ratio to embrace net benefit? Pharmacoeconomics. 2020;38(8):777–9.

Article Google Scholar

Download references

Acknowledgements

The authors are grateful for feedback received from the Institute of Health Economics (Canada), Dr. Etienne Gaudette and Nicholas Ouellet. They also thank Professor Christopher McCabe for useful direction and discussions.

Author information

Authors and Affiliations

1. **Flinders Health and Medical Research Institute, Flinders University, Bedford Park, SA, Australia**
   Laura C. Edney & Jonathan Karnon
2. **Centre for Health Economics, University of York, Heslington, York, UK**
   James Lomas & Jessica Ochalek
3. **Department of Quantitative Methods in Economics and Management, University of Las Palmas de Gran Canaria, Canary Islands, Las Palmas de Gran Canaria, Spain**
   Laura Vallejo-Torres
4. **IQ Healthcare, Radboud University and Medical Center, Nijmegen, The Netherlands**
   Niek Stadhouders
5. **Centre for Medical Technology Assessment, Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden**
   Jonathan Siverskog
6. **School of Public Health, University of Alberta, Edmonton, AB, Canada**
   Mike Paulden
7. **Health Economics and Epidemiology Research Office, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa**
   Ijeoma P. Edoka

8. **School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa**
Ijeoma P. Edoka

Corresponding authors

Correspondence to Laura C. Edney or Jessica Ochalek.

Ethics declarations

Funding

This work draws on research supported by the Patented Medicines Price Review Board (PMPRB). The views expressed herein do not necessarily represent the official policy of the PMPRB. The agreement with the PMPRB ensured the authors' independence in designing, writing and publishing the report. Laura Vallejo-Torres would additionally like to acknowledge the support of the project RTI2018-096365-J-I00.

Conflicts of interest

Laura C. Edney, James Lomas, Jonathan Karnon, Laura Vallejo-Torres, Niek Stadhouders, Jonathan Siverskog, Mike Paulden, Ijeoma P. Edoka, and Jessica Ochalek have no conflicts of interest to declare.

Availability of data and material

Not applicable.

Code availability

Not applicable.

Author contributions

LCE, JO, JL and JK conceived the study design. LCE, JO and JL drafted the initial manuscript. All authors commented on previous versions of the manuscript and approved the final manuscript.

Supplementary Information

Below is the link to the electronic supplementary material.

Supplementary file1 (DOCX 45 kb)

Supplementary file2 (DOCX 29 kb)

Supplementary file3 (DOCX 63 kb)

Supplementary file4 (DOCX 28 kb)