REVIEW ARTICLE

Conducting Value for Money Analyses for Non-randomised Interventional Studies Including Service Evaluations: An Educational Review with Recommendations

Matthew Franklin[1] · **James Lomas[2]** · **Gerry Richardson[2]**

Abstract

This article provides an educational review covering the consideration of conducting 'value for money' analyses as part of non-randomised study designs including service evaluations. These evaluations represent a vehicle for producing evidence such as value for money of a care intervention or service delivery model. Decision makers including charities and local and national governing bodies often rely on evidence from non-randomised data and service evaluations to inform their resource allocation decision-making. However, as randomised data obtained from randomised controlled trials are considered the 'gold standard' for assessing causation, the use of this alternative vehicle for producing an evidence base requires careful consideration. We refer to value for money analyses, but reflect on methods associated with economic evaluations as a form of analysis used to inform resource allocation decision-making alongside a finite budget. Not all forms of value for money analysis are considered a full economic evaluation with implications for the information provided to decision makers. The type of value for money analysis to be conducted requires considerations such as the outcome(s) of interest, study design, statistical methods to control for confounding and bias, and how to quantify and describe uncertainty and opportunity costs to decision makers in any resulting value for money estimates. Service evaluations as vehicles for producing evidence present different challenges to analysts than what is commonly associated with research, randomised controlled trials and health technology appraisals, requiring specific study design and analytic considerations. This educational review describes and discusses these considerations, as overlooking them could affect the information provided to decision makers who may make an 'ill-informed' decision based on 'poor' or 'inaccurate' information with long-term implications. We make direct comparisons between randomised controlled trials relative to non-randomised data as vehicles for assessing causation; given 'gold standard' randomised controlled trials have limitations. Although we use UK-based decision makers as examples, we reflect on the needs of decision makers internationally for evidence-based decision-making specific to resource allocation. We make recommendations based on the experiences of the authors in the UK, reflecting on the wide variety of methods available, used as documented in the empirical literature. These methods may not have been fully considered relevant to non-randomised study designs and/or service evaluations, but could improve and aid the analysis conducted to inform the relevant value for money decision problem.

✉ Matthew Franklin
matt.franklin@sheffield.ac.uk

[1] Health Economics and Decision Science (HEDS), School of Health and Related Research (ScHARR), University of Sheffield, West Court, 1 Mappin Street, Sheffield S1 4DT, UK

[2] Centre for Health Economics, University of York, Heslington, York, UK

△ Adis

Key Points

Service evaluations and associated evidence do not tend to receive the same peer-reviewed scrutiny, governing oversight, time or budgetary allowances relative to research as defined from a study ethics perspective. However, as a vehicle for producing evidence, the same rigorous methods that are associated with conducting research should be considered if permitted

Guidance and checklists exist for conducting economic evaluations alongside randomised controlled trials and as part of health technology assessments. However, as service evaluations usually serve a different purpose, useful and fundamental methods may be overlooked when assessing 'value for money'

It is difficult to suggest a single method to produce the 'value for money' evidence needed. However, evaluators need to consider the needs of the commissioner, but also what is required to produce the 'best' possible evidence (i.e. less uncertain and biased, but potentially costly) to inform the resource allocation decision problem

Evaluators should be transparent about the limitations of analyses that they conduct and should reflect on the impact that choices in the methods used may have on the results, conclusions and recommendations

1 Introduction

A consensus exists that policy and clinical decisions that affect the public and individual patients should be evidence-based [1, 2]. However, there is less clarity as to what constitutes 'appropriate' economic or 'value for money' (VfM) evidence, by which resource-allocation decision-making can be informed. In many cases within healthcare, particularly health technology assessments (HTAs) of new and existing medicines and treatments [3], the primary source of evidence may derive from research in the form of randomised controlled trials (RCTs). In the UK, these HTA processes are undertaken on behalf of the National Institute for Health and Care Excellence (NICE) and referred to as technology appraisals (TAs) [3]. NICE has influenced international processes, guidelines [4], and subsequently the HTA evidence base with Sculpher and Palmer [5] suggesting NICE could be considered a method innovator. Even before NICE's influence, internationally RCTs are often considered the 'gold standard' vehicle for producing evidence related to causation

such as 'treatment effects' [6]. In this article, we argue that RCTs and research more broadly, here defined as an exercise designed to inform decisions with clearly defined questions, aims and objectives to provide generalisable results [7], represent a subset of the possible relevant evidence base for a given decision problem. Indeed, in many cases conducting RCTs to gather evidence may not be appropriate, practical, affordable, ethical or even possible [8, 9].

In this article, we focus on interventional studies as non-randomised study designs including those conducted as service evaluations. From an ethics perspective, research is a vehicle for obtaining evidence with an ethical and underpinning legislative framework [7, 10]. As an alternative to research, service evaluations do not require research ethics (often referred to as 'NHS ethics' in the UK) [7]. Obtaining research ethics approval can be time consuming and may delay the start of a study [11]. Therefore, avoiding research ethic processes can be perceived as desirable in some circumstances (e.g. in the case of tight time and budget restrictions). However, an associated restriction for service evaluations is random allocation to treatment options is not permitted [7]. Additionally, the use of routine data is encouraged for service evaluations relative to primary data collection [7]. Service evaluations are not intended to provide generalisable estimates of intervention efficacy or effectiveness relative to an existing intervention, but they may be the source of the best evidence available. Although internationally alternatives to 'research' might be called or defined differently, evaluations that utilise non-randomised data are often used to inform decision-making such as by charities and local and national governing bodies [12, 13]. Non-randomised data and service evaluations present a different challenge to analysts than what is commonly associated with research, RCTs and HTAs. These evaluations usually serve a different purpose and require specific study design and analytic considerations. However, there are useful and fundamental methods associated with conducting research and HTA processes that can be appropriately applied if governing, time and budget restrictions allow.

This article provides an educational review of considerations when conducting VfM analyses in non-randomised studies including service evaluations. In doing so, we often make direct comparisons to RCTs and research as the current 'gold standard' to highlight relative strengths and weaknesses. We discuss to what extent there is a difference between 'service evaluation' and 'research' in terms of governing ethics and methodologies that could be used. We describe alternative methods for VfM analyses used depending on the resource allocation interests of the decision maker. Aspects to consider in terms of study design, statistical methods to control for confounding and bias, and how to quantify and describe uncertainty and opportunity costs to decision makers in any resulting VfM estimates are

also described. The possible implications of producing VfM evidence when these aforementioned aspects are not taken into account are a key point for discussion. We provide a range of references for further reading, including a glossary of cross-referenced key terms and methods provided in Appendix S1 of the Electronic Supplementary Material (ESM). Therefore, this article should be used as a reference guide when generating VfM evidence from non-randomised data including service evaluations, rather than a technical document describing in detail specific methodologies.

2 Evidence-Based Healthcare Decision-Making: Are Randomised Controlled Trials the Only Choice?

Randomised controlled trials are considered the 'gold standard' for producing an evidence base related to causation, but their limitations have been noted [13, 14]. Validity comparisons of the evidence generated from non-randomised studies and RCTs have dispelled many misperceptions of the former as a viable option for producing an evidence base for healthcare decision-making [12–15]. Although the controlled nature of RCTs is often what enables them to produce explanatory results and have high 'internal' validity [16, 17], moving to real-world decision-relevant settings may be considered to improve the relative 'external' validity of such results [13].

Decision makers such as charities and local and national governing bodies are unlikely to commission RCTs in part because of their associated time and financial costs. In the UK, this includes Clinical Commissioning Groups (CCGs) who are clinician-led bodies responsible for commissioning healthcare services within a local area. Since the UK's Health and Social Care Act of 2012 [18], CCGs have become increasingly responsible for the health needs of their district of responsibility. Furthermore, in the UK, nationally funded initiatives such as projects within the National Health Service (NHS) Test Beds programme [19] have relied on service evaluations to produce an evidence base on intervention effectiveness and VfM in localised areas. Although these are UK examples, the financial sustainability of healthcare systems internationally is reliant on the careful management and commissioning of the plethora of services and interventions that are available for a variety of decision makers to fund. The role of regional and local decision makers is explicitly mentioned in the Helsinki Statement on Health in All Policies (HiAP) [20, 21]. This call upon governments worldwide states that "health authorities at all levels (national, regional, local) are key actors in promoting HiAP" (p. 17), which includes "building knowledge by providing evidence of success and lessons learnt" (p. 18) [21]. The limited ability of many decision makers internationally

to generate high-quality evidence and analyses necessitates the use of appropriate and timely approaches to inform their decision-making based on an accurate and relevant evidence base. Internationally, few decision makers have the available finances to produce evidence such as the National Institute for Health Research (NIHR) as the largest national funder of clinical research in Europe [22]. As such, there has been much international interest in alternatives to RCTs to produce an evidence base including related to VfM [23].

3 Research and Service Evaluations: Why are They Used and is There a Difference?

In the UK, service evaluations are generally used when the intervention of interest has been, or is about to be, implemented within a specific care setting. Unlike research, service evaluations are generally not based on well-formed aims and objectives to answer specific hypotheses to produce generalisable results. Instead, they are used to assess 'what standard does this service achieve' in a more general sense, with aspects of interest often being related to effectiveness and/or VfM.

As an overview of our perspective when assessing VfM, we do not believe there is a clear dichotomy between research and service evaluation other than from an ethics perspective and associated legislative underpinnings. Even from an ethics perspective, this dichotomy is not always clear. What defines 'research' from an ethics perspective internationally is complex, with aspects for consideration described by Gevers [10] including the seminal Declaration of Helsinki [24] for informing international human research ethics. From a UK perspective, Table 1 provides an overview of what the NHS Health Research Authority (HRA) Research Ethics Service (RES) considers 'research' relative to 'service evaluation'. The HRA also has set standard operating procedures for Research Ethics Committees (RECs) to try and make processes more standardised and transparent including when judging what is 'research' [25]. A key point from Table 1 is that randomisation is only permitted within research. Therefore, a key consideration for service evaluations is conducting analyses using non-randomised data.

Suggesting what interventions can be evaluated within a 'research' ethics framework or otherwise (e.g. as a service evaluation), in the UK or internationally, is outside the scope of this article. However an evaluation is defined ethically, when it comes to evaluating VfM, we suggest the same rigorous considerations are required. This includes when developing the study design (without purposeful or random allocation of the intervention for service evaluations), analytical methods and reporting of evidence. We further discuss this perceived dichotomy between 'research'

Table 1 Defining service evaluations relative to research from UK Health Research Authority guidance. Source: NHS HRA Research Ethics Service [7]

No	Consideration	Research	Service evaluation
1	Overarching purpose	"The attempt to derive generalisable or transferable new knowledge to answer questions with scientifically sound methods including studies that aim to generate hypotheses as well as studies that aim to test them, in addition to simply descriptive studies."	"Designed and conducted solely to define or judge current care."
2	Study design	"Quantitative research—can be designed to test a hypothesis as in a randomised controlled trial or can simply be descriptive as in a postal survey."	"Designed to answer: "What standard does this service achieve?""
3	Study question	"Quantitative research—addresses clearly defined questions, aims and objectives."	"Measures current service without reference to a standard."
4	Intervention of interest	"Quantitative research—may involve evaluating or comparing interventions, particularly new ones. However, some quantitative research such as descriptive surveys, do not involve interventions."	"Involves an intervention in use only. The choice of treatment, care or services is that of the care professional and patient/service user according to guidance, professional standards and/or patient/ service user preference."
5	Intervention allocation	"Quantitative research—study design may involve allocating patients/service users/healthy volunteers to an intervention."	"No allocation to intervention: the care professional and patient/ service user have chosen intervention before service evaluation."
6	Randomisation	"May involve randomisation."	"No randomisation."
7	Data collection	"Usually involves collecting data that are additional to those for routine care but may include data collected routinely. May involve treatments, samples or investigations additional to routine care. May involve data collected from interviews, focus groups and/or observation."	"Usually involves analysis of existing data but may also include administration of interview(s) or questionnaire(s)."
8	Research Ethics Committee (REC) review	"Normally requires REC review but not always."	"Does not require REC review."

The text within this table has been directly quoted from the "Defining research" (Oct 2017) table developed by the National Health Service's (NHS's) Health Research Authority's (HRA's) Research Ethics Service [1]. Any mention of qualitative research in the original table has been omitted in the above table due to the fact value for money analysis is typically classified as quantitative analysis, and so qualitative analysis is not pertinent to the objectives of the manuscript

and 'service evaluation' from a UK perspective with some international considerations in Appendix S2 of the ESM.

4 Economic Evaluations and Partial Evaluations: Methods and Distinctions

Economic evaluation is widely used for the appraisal of healthcare programmes, taking into account both the costs and the consequences (i.e. effects or outcomes) of two (or more) alternatives [26]. There are multiple forms of economic evaluation. However, all are VfM analyses, with 'costs' representing an integral aspect of the evaluation process owing to the resultant opportunity costs (Appendix S1 of the ESM) from resources not being available for other purposes [27–29]. Traditionally, at the local level, there has been very little use of economic evaluation evidence, although there is suggestion this has increased overtime particularly in the UK [30, 31]. There is a need to explain and rationalise the purpose of economic evaluation to decision makers with a particular focus on commissioners of the evaluation and decision makers they represent. A commissioner is an individual (or group) who has a legitimate authority to make decisions, such as a representative of a local governing body being put in charge of identifying relevant experts to conduct the VfM analyses (e.g. health economists). From a commissioning perspective, requesting a VfM analysis may be more pertinent than requesting an economic evaluation. The term 'value for money' tends to mean something to commissioners more than 'economic evaluation', partly because it is a politically motivated and widely recognised term. The NHS Constitution for England states: "The NHS is committed to providing best value for taxpayers' money" [32]. Thus, making 'value for money' a political 'buzz word' when discussing care funding and provision.

In the UK, cost-effectiveness analysis (CEA) in the form of cost-utility-analysis (CUA) based on quality-adjusted life years (QALYs) has become the most popular form of economic evaluation. This is partly because it is NICE's 'gold standard' reference case, with NICE's guidelines [3] informing HTA processes internationally [4]. There are several papers debating the extent that economic evaluations of different forms of interventions fit a typical HTA [33,

34], such that current guidelines may not be fully applicable, including: public health [35–37], antimicrobials [38], diagnostics [39], medical devices [40], genetics [41], digital [42], environmental [43], and service and delivery interventions [33, 44]. Whatever the intervention of interest, Drummond et al. [26] suggest that an economic evaluation would "explicitly consider the relative consequences of the alternatives and compare them with the relative costs" (p. 5). Anything else than the aforementioned "economic evaluation" is a "partial evaluation" (see Table 2).

Common economic evaluation methods include cost–benefit analysis (CBA), which uses individuals' values for their outcomes to convert into a monetary unit, and CEA/CUA, which values outcome in natural units (usually health outcomes that for CUA are preference-based) [26]. Each of the aforementioned tend to represent VfM as a single outcome, normally as a ratio of outcomes relative to costs, e.g. incremental cost-effectiveness ratios (ICERs) and benefit–cost ratios (the equations for these ratios are presented in Appendix S3 of the ESM). Alternative economic evaluation methods include cost-minimisation analysis (CMA) and cost-consequence analysis (CCA). Cost-minimisation analysis is considered flawed based on its underlying assumption that outcomes can be equivalent between alternatives [26, 45] (a review on using CMA to inform NICE is underway [46]). Cost-consequence analysis offers flexibility for representing VfM in a disaggregated or aggregated manner [47–50], but using a single outcome for cross-comparison decision-making has comparative advantages and disadvantages [51, 52]. Budget impact analysis (BIA) [48, 53] is a 'partial evaluation' method that only accounts for costs to addresses the expected changes in the expenditure of a healthcare system after the adoption of a new intervention [48]. Budget impact analysis is recommended to be included alongside economic evaluation methods such as CEA [48] and CCA [54, 55].

Other cost-related methods used within service evaluations, commonly associated with evaluating public health programmes [56], include return on investment (ROI) analysis often based on cost savings as the 'return' (thus a partial evaluation) [57, 58] and social ROI analysis based on natural outcomes given monetary weights [59–61]. Whether social ROI analyses can be considered a full economic relative to partial evaluation depends on what costs are taken

Table 2 Overview of types of analysis focussed on costs and consequences. Source: Adapted from Drummond, Sculpher [26]

Options for analysis	Examine only consequences*	Examine only costs**	Examine consequences and costs
No comparison of alternatives	Outcome description	Cost description	Cost-outcome evaluation
Comparison of alternatives	Efficacy or effectiveness evaluation	Cost-comparison analysis	Economic evaluation

*Consequences can be referred to as 'benefits' or 'outcomes', noting that 'effect' implies causality. These consequences are the impact on people or a population in terms of health or 'welfare'; although, well-being is of growing concern for economic evaluations

** Costs often include cost beyond intervention costs (e.g. treatment costs) dependent on the costing perspective (see Sect. 5)

into account and if there is a comparison with an alternative (which are also factors for consideration with ROI analyses, with Sect. 5 including a relevant discussion related to costs). Often there is confusion between evaluation approaches because of how outcomes are presented to decision makers. However, certain approaches may contain the same information, but with a different presented outcome, e.g. as an ICER rather than a benefit–cost ratio (Appendix S3 of the ESM further discusses this aspect).

5 Costing Perspective: Intervention Costs, Future Costs and Other Considerations

The costs to include in a VfM analysis are dependent on what question needs to be answered for the decision context being informed [26, 28, 62]. Fundamentally, when a care service introduces an intervention or delivery model, the evaluation should include the direct costs of this aspect referred to as 'intervention costs' (Appendix S1 of the ESM). For example, if a new member of staff was introduced within a care system, the cost of this staff member over the time horizon of interest should be included. When comparing between two alternatives, incremental costs associated with the intervention relative to the alternative(s) assessed (e.g. usual or previous care model) are of interest. These direct intervention costs should always be included in the VfM analysis. However, what *other* costs should be included is the focus of considerable debate and research. de Vries et al. [63] have attempted to classify potential other costs into three categories, each described as 'future costs' (Appendix S1 of the ESM): (1) future *related* medical costs; (2) future *unrelated* medical costs; and (3) future non-medical costs. de Vries et al. [63] state that the literature suggests that inclusion of 'related' and 'unrelated' medical costs is required to obtain optimal outcomes from available resources irrespective of the costing perspective adopted. The inclusion of medical costs is referred to as the care-payer perspective (Appendix S1 of the ESM). However, 'unrelated' costs are typically difficult to define and thus often excluded/ignored [64, 65]. A case for also collecting non-medical costs as part of a societal perspective (Appendix S1 of the ESM) has been made by Jönsson [66]. A framework for including the societal perspective and associated complications is described by Walker et al. [67].

Obtaining future cost data can be costly, timely and resource intensive (depending on if data are readily available or not) [68]. As such for service evaluations, what future costs are included alongside the intervention costs may be limited to where benefits could be observed rather than accounting for opportunity costs across the wider care system. This may be more pertinent with ROI analyses, which often focusses just on intervention costs relative to beneficial returns without accounting for wider cost implications. For example, introducing set 'inpatient bed days' could reduce short-term hospital costs, which in a ROI analysis could seem beneficial, but this ignores wider morbidity, mortality, readmission and care costs associated with discharging patients too early. It seems reasonable to suggest that a full economic evaluation should attempt to account for future costs over an appropriate time horizon to capture resource use implications of the intervention of interest. The exact future costs to include have yet to be firmly established and thus may be informed by the decision maker's perspective with potential implications for the evaluation (see Sect. 9). In some cases, there is also a suggestion to include 'implementation' costs, such that timely implementation of recommended interventions can provide health benefits to patients and cost savings to health service providers [69]. There are debates and complications with the inclusion of such costs [69–71], with a discussion on the economic evaluation of implementation strategies in healthcare by Hoomans and Severens [72].

6 Routine Data for Estimating Resource Use, Costs and Non-monetary Consequences

Primary data collection can be time consuming and costly, but may also have implications that consider ethical consideration (e.g. if talking to vulnerable groups). As such, for service evaluations using routinely collected data is recommended [7]. The data available for analysis may restrict the VfM method (Sect. 4) and costing perspective (Sect. 5), but could also inform the potential study design (Sect. 7).

Cost data for VfM analyses are estimated based on care resource-use data to which unit costs are attached (Appendix S1 of the ESM) [28]. There are self-reported and routinely collected resource-use data methods as described by Franklin and Thorn [68], noting their (and our) examples are based on routine data sources in England. If the consequences of interest are also related to resource use (e.g. cost per inpatient bed days avoided), then the source of routine data for costs and consequences may be the same. The range of resource-use information required will depend on the costing perspective (Sect. 5). However, different care services tend to collect their resource-use information on different electronic systems, which often do not tend to be linked at the patient or service level (e.g. primary care and hospital care) [68]. Although large linked databases may exist, their use has complications [68, 73]. In England, commissioners can supplement their local data flows with data from the Secondary Uses Service (SUS) [74]. Such data could be used for service evaluations, albeit with a variety of time, monetary, technical and information governance restrictions

[68]. As such, routine data are recommended but often difficult to utilise [28, 68, 75].

For CEA, consequences are measured in natural units that could be based on routinely collected data (e.g. 'cost per death avoided' based on mortality data). For CUA, preference-based values for health-related outcomes are required to elicit the QALY (Appendix S1 of the ESM) [3]. Obtaining such preference-based values can be problematic if not routinely collected (e.g. routinely collecting the EQ-5D, as the NICE-preferred preference-based measure [3]). As an example of using indirectly collected preference-based data, Franklin et al. [76] suggest a method for attaching preference-based values to routinely collected, health-related events of interest (i.e. asthma exacerbations) to conduct a CUA. Preference-based values, often referred to as 'utility' values, could be sourced via the ScHARR Health Utilities Database (ScHARRHUD [77]: www.scharrhud.org). If clinical or condition-specific measures relevant to the intervention of interest are routinely collected which could be used for CEA, then a 'mapping' or 'cross-walk' algorithm may exist to allow the statistical prediction of utility values from that measure to conduct CUA. The purpose and procedures of statistical mapping are described by Longworth and Rowen [78], with a systematic review of mapping studies by Mukuria et al. [79], and an online database of mapping studies also currently available (HERC database of mapping studies [80]: www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies).

7 Statistical Considerations Based on Study Design, Underlying Data Attributes, and for Reporting Uncertainty

The desired form of VfM analysis should be accounted for at the study design stage, given that not all study designs are good vehicles for VfM analyses [26]. When time and finance are restricted, the study design might be dependent on data availability (Sect. 6). Value for money analyses are 'analytical' in nature because there is an intention to infer VfM as a causal factor related to an intervention compared to an alternative (e.g. usual or previous care model). As such, there will be the need to choose an analytical study design and associated appropriate statistical method(s). Hinde et al. [81] have explored the possible scenarios that could occur when seeking to conduct a quantitative evaluation of an intervention at the local level, specifically with regard to availability of evidence, the subsequent statistical method chosen and the resulting impact on 'effectiveness' evidence.

Analytic study designs as required for VfM analyses can be broadly classified as observational (e.g. case series, cohort, cross-sectional and case–control study designs [82, 83]), or experimental (e.g. before-and-after studies [84, 85], comparative/controlled trials and RCTs [82]). These are different to 'descriptive' studies, which could include describing costs, qualitative studies or cross-sectional surveys [86]. In analytic studies, participants are identified and observed, and characteristics including outcomes and costs are recorded. Additionally, for experimental studies, the setting should be equivalent across all participants, an intervention is used and is part of the assessment and there is an observation/evaluation of the effects of the intervention with causality being of particular interest (relative to association as a common interest in observational studies). When causality is of particular interest, there is a need to reduce *chance*, eliminate *bias* and account for *confounding* (Appendix S1 of the ESM). Although these aspects can be accounted for using statistical methods, good study designs reduce reliance on statistical methods with experimental studies generally regarded as being less susceptible to bias than observational studies.

Experimental designs such as comparative trials are generally preferred when inferring causality, with a preference for randomised trials [82]. Randomisation to treatment groups is preferred as the process reduces chance and bias in resulting study estimates, but RCTs themselves have limitations [13, 14]. In any case, randomisation to treatment groups is not ethically permitted outside of research (see Table 1 for a UK perspective). Therefore, non-randomised and historical control designs may be options for service evaluations. Historical controls alone have been shown to overestimate new treatment benefits [85, 87]. Authors such as Goodacre [84] have made a case why before-and-after studies without a comparison group and/or appropriate statistical methodology (e.g. interrupted time-series analysis, described later in this section) should be discouraged for evaluations. For non-randomised comparative trials (supplemented with or without historical data for both groups) particularly as part of a service evaluation, there are often difficulties when trying to recruit and perform primary data collection for a control group (or obtain relevant and necessary historical data retrospectively). Primary data collection and recruitment is expensive, time consuming, may have ethical considerations, and is thus often deemed undesirable by the service evaluation funder. Service evaluations could be conducted as 'natural experiments'. Natural experiments are defined as: "naturally occurring circumstances in which subsets of a population have different level of exposure to a supposed causal factor, in a situation resembling an actual experiment where human subjects would be randomly allocated to groups" [88, 89]. Deidda et al. [88] have developed an economic evaluation framework when using natural experiments with a specific focus on public health interventions. As the framework was developed mainly to evaluate public health interventions, not all aspects of the framework may be relevant nor necessary for all service evaluations. For

example, under 'costs', the framework suggests a societal perspective would be of interest; however, such a perspective may not be required/possible for all interventions to be evaluated (Sect. 5).

When economic evaluations are conducted as part of non-randomised study designs, the need to account for the non-randomised nature of the data is not always recognised [23]. There are suggested statistical methods/guidance to mitigate confounding and bias using observational [90] and 'real-world' data [91]. As an example, guidance by Faria et al. [90] provides an overview of a method described as 'Matching', which aims to replicate randomisation by identifying/matching control individuals who are similar to those receiving the intervention in one or more characteristic. Matching could be conducted within routinely collected datasets, assuming enough patient characteristics exist for matching, and subsequently used for the VfM analysis as part of a service evaluation (Sect. 6). There has been much interest in how such methods can be applied to improve VfM analyses (particularly CEA) using non-randomised data. As examples, using propensity score matching methods for CEAs has been explored by Manca and Austin [92]. Using regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluations has been explored by Kreif et al. [93]. The use of propensity score matching against other methods used in observational data such as difference-in-difference and regression models for (health) economic analysis has been explored by Crown [94]. Guidance on choosing an appropriate weighting mechanism for propensity score matching is described by Desai and Franklin [95]. These 'matching' methods are useful when there is interest in better defining a group for comparison to reduce bias. In comparison, interrupted time-series analysis, a statistical method using longitudinal data, has been preferred for single-arm before-and-after studies without a comparator [84] and has been used to inform modelling-based (Appendix S1 of the ESM) economic evaluations [96]. A short tutorial for using interrupted time-series to evaluate public health interventions is described by Bernal et al. [97], which outlines the data needed for interrupted time-series analyses. How to combine statistical methods that account for the non-randomised aspects of the data among other considerations pertinent to VfM analyses (e.g. comparison between alternatives, and accounting for costs and outcomes) is still an area for further research and guidance.

There are specific statistical considerations pertinent to VfM analyses that need to be accounted for alongside the non-randomised nature of the data. Two educational reviews already describe the use of utility data for CUA [98] and costs for CEA [28], both of which describe statistical considerations such as: assessing cost and consequence (utility) data and its distribution; baseline covariate adjustments;

and dealing with missing data. We do not wish to repeat these educational reviews. Instead, we shall summarise a few key points and suggested statistical methods, focussing particularly on costs as a common factor in all VfM analyses. Controlling for baseline covariates (i.e. aspects that influence costs, e.g. age, frailty, health status) is a simple method for making adjustments to improve precision and correct for between-group imbalances [99], particularly for non-randomised groups. Regression-based methods are typically used to account for baseline covariates when making estimations. In the case of costs, the case has been made to use [28]: (1) parametric methods including ordinary least squares (OLS), generalised linear models (GLMs), extended estimating equations (EEE), multi-level models and generalised estimating equations (GEE) models; and (2) non-parametric methods including bootstrapping and the two-stage bootstrap. All forms of VfM should include unadjusted (i.e. observed) and adjusted analyses [28]. Although such methods have long been used as part of statistical analyses related to clinical outcomes [99], their use for VfM analyses has not always been fully recognised [28, 98].

There are also statistical methods that allow a better reflection of the uncertainty around estimates, which should be applied to costs and consequences. Common methods applied to economic evaluations include bootstrapping for within-trial evaluations and probabilistic sensitivity analysis (PSA) using Monte Carlo simulation for modelling-based analyses (Appendix S1 of the ESM). These methods allow the random resampling of the observed data over a specified number of iterations either non-parametrically or parametrically. The estimates from which can be presented to decision makers in cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs) to indicate the probability of achieving a specified outcome over a range of monetary valuations of consequence outcomes [100], e.g. 'cost-effectiveness thresholds' [101, 102]. An alternative that allows for a quantification of a change in parameter(s) value if we are particularly unsure about point estimate value(s) is one- or multi-way sensitivity analyses, whereby point-estimate input values are changed (e.g. average intervention cost) and the resulting change in the outcome is reported (e.g. change in ICER value). An overview of the application of these methods to costs is described by Franklin et al. [28]. The use of such methods should be applied irrespective of the type of VfM analysis conducted, as they represent statistical methods to quantify and account for the uncertainty around the parameters associated with the VfM analyses that should be presented to decision makers.

Another aspect for consideration is using a relevant and appropriate time horizon. For service evaluations, particularly if informing policy decisions that require timely evidence, the ability to collect primary data over a relevant time horizon (whereby 'relevant' is dependent on the decision

context) is potentially limited. There are various methods for extrapolating results beyond an observed time horizon, such as survival analysis [103] and other methods to account for censoring [104–106] dependent on the parameter of interest (e.g. longer-term mortality or costs). Economic modelling is often rationalised based on the inability to collect sufficient parameter information over a relevant time horizon within a single study to inform the decision problem and thus could be an alternative option [107]. However, modelling analyses and subsequent estimates will be driven by the data used to inform the model, e.g. a key model driver will be the input parameter estimate of intervention treatment effectiveness. If the intention is to use a service evaluation to produce the estimates on treatment effectiveness that will drive the model, then the aforementioned statistical methods described in this section will still be needed when estimating treatment effects from non-randomised data. Examples of CUA modelling studies born of a service evaluation include Franklin and Hunter [108] (fall-screening and fall-prevention intervention) and Hunter et al. [109] (major system change in acute stroke services). It should be noted, however, some decision makers may be interested in short-term costs and consequences (monetary or otherwise). For example, over 1 year because of yearly budget allocations (often associated with the 'financial year'), rather than long-term planning dependent on the decision problem (see Sect. 9).

8 Quantifying the Value of Information

As stated by Sculpher, Claxton [9]: "It is argued that any framework for economic analysis can only be judged insofar as it can inform two key decisions and be consistent with the objectives of a health care system subject to its resource constraints. The two decisions are, firstly, whether to adopt a health technology given existing evidence and, secondly, an assessment of whether more evidence is required to support this decision in the future". The methods described in this article so far relate to the aforementioned first point, but value of information (VOI) [Appendix S1 of the ESM] is associated with the second point. Value of information represents the monetary value of collecting more information that could inform an investment decision. There are three types of VOI worth considering: expected value of perfect information (EVPI), expected value of partial perfect information (EVPPI) and expected value of sample information (EVSI). An overview of these methods is described by Jackson et al. [110], with simplified descriptions provided in Appendix S1 of the ESM.

There are two key issues with VOI in general and specific to service evaluations. First, traditionally, VOI analyses are computationally complicated and time consuming. However, there are suggested methods [111–113] and free-licence software (e.g. Sheffield Accelerated Value of Information tool [112]: https://savi.shef.ac.uk/SAVI/) that can speed up and simplify the process, with a practical VOI guide by Wilson [114] and a description of emerging good practice VOI analytical methods by Rothery et al. [115]. Furthermore, although you need the output from a PSA or other Bayesian framework to be able to calculate VOI [111], parameter values from within-study VfM analyses can be placed within a simple model (e.g. decision tree) to run a PSA and subsequent VOI analysis (Appendix S1 of the ESM) [116]. The second issue is related to understanding the outputs from VOI, explaining the implications to decision makers and why they should pay attention to VOI. For trials, VOI can be particularly useful for pilot and feasibility studies, as they will place a monetary value on the worth of conducting the next stage trial design (e.g. RCT). When informing local or national decision makers, the purpose is to highlight the potential monetary consequence of beginning or continuing to invest in an intervention based on the current information available. As stated in Sect. 3: "service evaluations are generally used when the intervention of interest has been, or is about to be, implemented within a specific care setting". As the decision maker may have already made the investment in an intervention, point estimates from any VfM analysis should confirm the already made decision to invest or not. However, such point estimates do not suggest if the service evaluation has provided enough information to inform the decision to invest in the future for as long as the investment decision is relevant (e.g. over the next 1–5 years). In addition, as many decision makers are responsible for a plethora of care interventions, another consideration is which interventions should be the focus of further evaluation in the future to check on their investment. Value of information can help prioritise and monetarise the investment in the service evaluation as well as the investment in the intervention. Decision makers may not be able to fully comprehend the impact of investing in an intervention or service evaluation based on the information provided to them, particularly related to the uncertainty around estimates. Value of information can quantify this aspect into a monetary value to be considered alongside other evidence provided.

9 Informing Decisions in Healthcare: A Discussion Related to Value for Money

Within this article, we have described a variety of matters to consider when conducting VfM analyses alongside non-randomised study designs including service evaluations. Reflecting on the UK, NICE has issued guidance on how to conduct economic evaluations for HTAs that were developed with allocative efficiency in mind across the whole NHS [3]. NICE's processes and guidelines have influenced

reimbursement agencies internationally specific to their HTA processes [4, 5]. However, such guidelines have been described to not always be practical nor relevant in every decision-making context [33, 35–44]. Such guidelines may also align more with research practices than service evaluations, whereby the former could include conducting expensive RCTs whereas the latter may be a more 'budget and time' conscious approach. There may be a need to move away from guidelines such as NICE's HTA processes as the 'gold standard' for evaluating care interventions, but careful consideration and rationale need to be given when moving away from 'gold standards'. This includes moving away from RCTs that dominant the HTA evidence base and CUA preferred by many reimbursement agencies internationally [4].

There are key differences between producing evidence for a reimbursement agency like NICE compared to decision makers that are part of government. For example, in the UK, NICE currently acts as an independent reimbursement agency for the NHS (although it was once a special health authority for the NHS), is not part of any government body, and NICE's evidence review groups (as an external academic organisation independent of NICE) review the evidence that informs the HTA process [117, 118]. NICE also has principles [119] that align with the NHS Constitution [32], which is to provide "the best value for taxpayers' money and the most effective, fair and sustainable use of finite resources" (NICE principle 7, point 22) [119]. NICE's principle when rationalising its stance on cost per QALY states as part of its allocative efficiency objective: "[Cost per QALY] takes into account the *'opportunity cost'* of recommending one intervention instead of another, highlighting that there would have been other potential uses of the resource. It includes the needs of other people using services *now or in the future* who are not known and not represented" (NICE principle 7, point 23) [119]. NICE has incorporated a level of independence between the evidence reviewers and final decision makers, while also producing guidance that aligns with its allocative efficiency objectives. In contrast, when producing evidence for local and national governing bodies, the decision maker may also be the commissioner of the evaluation, may have a narrower perspective when assessing 'opportunity costs' and a shorter time horizon of interest for the evaluation. Each of the aforementioned may not be good factors. The role of local relative to national government when providing healthcare has long been a point of debate internationally, with the World Health Organization reflecting on localised decision-making in their 1997 report "The role of local government in health: comparative experiences and major issues" [120]. More recently and focussed on the NHS, a question has been raised of "should local government run the NHS?" [121], which aligns with the powers given to local agencies within the Health and Social Care Act of 2012 [18]. The advantages for local government made

by Furber [121] mainly focus on local government's ability to deal with localised public health concerns and inequality issues, relative to national concerns including opportunity costs across the whole NHS budget. From a sceptic's perspective, the extent to which obtaining good-quality unbiased estimates for a relevant time horizon is desirable relative to confirming an investment was 'correct' and wanting evidence to confirm this aspect can represent the different desires of local government agencies [122]. Additionally, as an example, localised decision makers may only wish to focus only on the opportunity costs in their jurisdiction of interest. For example, in England, an NHS foundation trust may only want to focus on care provided in hospitals as their care jurisdiction of interest. As such, their costing perspective may ignore other care services within which opportunity costs might be observed (e.g. primary care). Such decision makers may therefore opt to ignore other relevant opportunity costs that are recommended to be included by independent reimbursement agencies such as NICE. There is a case to be made that focussing just on the decision makers' perspective may not always be the appropriate perspective to take if, on the whole, it may lead to inefficient resource allocation across the whole care budget.

Based on NICE and other international reimbursement agencies' guidance, CUA is preferred for HTA [3, 4]. A key rationale for using CUA is it allows for comparable outputs in terms of economic evidence (i.e. cost per QALY estimates) for cross-care decision-making. Although the QALY framework is not perfect with a key debate questioning the concept of 'a QALY is a QALY', which enables cross-comparability [123, 124], using a single outcome metric such as the QALY still has its advantages albeit with the need for some suggested improvements [26, 125]. Using CUA is not restricted within service evaluations. Given the advantages of using a single metric and the stance of many reimbursement agencies internationally, perhaps cost per QALY analysis should be given priority across all care resource-use decision-making (noting we make this case with a UK perspective specifically in mind). However, it should be noted that this perspective might differ dependent on the care funding system incorporated including the use of social and private health insurance systems, rather than a tax-based system mainly used for the NHS. There has also been attempts and/or suggested frameworks to make CEA and QALYs applicable to more decision-making contexts. For example, equity concerns that are a key factor for consideration by many decision makers are suggested to be accounted for within distributional CEA for which there is a published tutorial [126], with a case study related to the UK Bowel Cancer Screening programme [127] and rotavirus vaccination programme in Ethiopia [128]. Furthermore, methods associated with estimating a new evidence-based cost-effectiveness threshold for NICE [27] (relative to NICE's

current non-evidence-based £20,000–£30,000 per QALY threshold [3]) has sprouted a range of other studies, debate [129–131] and associated methods for making CEA applicable to more healthcare decision-making contexts. For those unaware of cost-effectiveness thresholds, McCabe et al. [102] (specific to NICE) and Culyer [101] (more general concept) outline what they are and their uses. These methods for estimating cost-effectiveness thresholds are based on the concept of health opportunity costs, i.e. the health benefits that could have been achieved had the resources been used elsewhere in the healthcare system [132]. These methods incorporate wider opportunity cost concerns within CEA, QALYs and even disability-adjusted life years for use in low- and middle-income countries [133]. New approaches born from these methods include estimating social variation in the health effects of changes in healthcare expenditure [134]. Relatedly, Lomas [135] suggests a framework for incorporating affordability concerns alongside cost-effectiveness estimates highlighting an example that using a BIA alongside a CEA does not deal with such concerns. The "cost-effective but unaffordable paradox" [132] has been discussed in priority setting for global health programmes [136] but also in the context of high-income countries (e.g. UK and USA) [137, 138], which has relevance for local and national decision makers with finite budgets.

Despite the case for using a single metric and advances made in its potential use (with limitations), cost per QALY is rarely used within service evaluations. We suggest two key reasons: (1) the study design and data collection methodology means it is too difficult or not possible to collect the data to inform the CUA or (2) the CUA is not of interest to the decision maker for various reasons, including not understanding how to interpret QALY gains and associated ICERs. Detsky and Laupacis [139], in their paper 'Relevance of Cost-effectiveness Analysis to Clinicians and Policy Makers', suggest that: "In addition to the problems of looking at cost-effectiveness ratios individually, interpreting those ratios can be difficult for clinicians and decision-makers. *It is not easy to understand what a QALY is*" (p. 223). How best to explain the QALY including how it is (or could be) used to inform decision-making is certainly an area of interest. An HTA report on 'The use of economic evaluations in NHS decision-making' by Williams et al. [51] suggests: "Committee [including local] members raised concerns about lack of understanding of the economic analysis but felt that a single measure of benefit, e.g. the quality-adjusted life-year, was useful in allowing comparison of disparate health interventions and in providing a benchmark for later decisions" (p. iii). This report suggests the QALY could have a place in local decision-making if stakeholders better understood its construct and purpose. Within a service evaluation context, there is no guidance to suggest you have to, or how to,

conduct a CUA, meaning it can be overlooked or avoided for the right or wrong reasons.

For service evaluations, if CUA is not desirable and/or possible, then other forms of VfM analyses may offer potential alternatives. Arguably, producing any form of economic evidence to inform decision-making is better than no evidence, provided the analysis and outcomes have been conducted and reported 'appropriately' (e.g. appropriate study design, statistical analysis, and uncertainty and opportunity cost reporting options). An issue seems to stem from the lack of guidance and monitoring of the use of VfM analyses when informing localised and even national decision-making as part of service evaluations, which is more common for research. For example, NICE's evidence review groups provide independent reviews of evidence used to inform decision-making [117, 118], and the NIHR has an independent peer-review process pre-funding and at the reporting stage associated with its funding programme journals [140]. Evidence used by some decision makers and associated methods may not be properly peer reviewed, thus allowing for commissioners and decision makers to base their decisions on a multitude of evidence with varying quality. It should be noted that such peer-review processes themselves can be time consuming and costly, and thus may not be a practical nor cost-efficient option. However, the use of checklists such as Drummond et al. [26] and the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [141] checklists still have a place alongside VfM for service evaluations, as these standards should still enable cross-comparable evaluations with some suggested/recommended methods included.

There are practical examples as part of the debate for [142] (here focussed on the Cancer Drugs Fund) and against [143] using RCTs, the 'against' here being a comedic look at using RCTs to assess parachute use to prevent death and major trauma when jumping from aircraft. The use of RCTs is debated relative to other options including non-randomised, real-world, and observational data either with or without a comparison group, and using historical controls [13]. Further work should be conducted on how VfM methodology can be adapted to deal with these different study designs and data sources. For example, an economic model by Franklin et al. [96] combined treatment effect evidence from an interrupted time-series analysis using routine data with modelling methodology to conduct a CUA. Although this example represents one potential solution and relied on a natural experiment design to conduct the interrupted time-series analysis, further work is required to examine how to combine existing statistical methods for determining treatment effects in observational [90] and real-world [91] data with VfM methodology.

Multi-criteria decision-making (MCDM) and multiple-criteria decision analysis (MCDA) could deal with

some limitations associated with single evaluation-based approaches when informing decision makers who require a range of information (not just related to VfM analyses) [144, 145], with good practice guidance of emerging methodologies [146]. Edwards and McIntosh [35] discuss a method called "programme budgeting and marginal analysis" for economic evaluation and prioritisation between public health interventions. They suggest programme budgeting and marginal analysis is an example of multiple-criteria decision analysis and describe steps for its use, but as we are not familiar with the approach, we suggest the interested reader refer to Edwards and McIntosh [35]. Some authors have called for the application of realist evaluation methodologies [147] to better explain cost-effectiveness mechanisms within more "explanatory economic evaluations". Anderson and Hardwick [147] describe the premise, comparing aspects of both realist evaluations and economic evaluations. Although we approve of the idea of more explanatory economic evaluations, we believe the practical application and understanding of such methodology to inform decision-making is in its early days.

It is important to note that even as part of NICE-based decision-making, there is a range of evidence produced (e.g. via non-randomised studies) that requires appropriate methodologies that are not always taken into account. A review of NICE appraisals of pharmaceuticals (2000–16) by Anderson et al. [148] found variations in establishing comparative clinical effectiveness. Of 489 individual pharmaceutical technologies assessed by NICE, 22 (4%) used non-RCT data to estimate comparative clinical effectiveness, with the methods for establishing external controls including: 13 (59%) used published trials, 6 (27%) used observational data, 2 (9%) used expert opinion and 1 (5%) used a responder vs non-responder analysis. Only five (23%) used a regression model to adjust for covariates, indicating that fundamental statistical methodology is missing even from evidence presented to NICE. Interestingly, the authors did not observe a notable difference in the proportion of pharmaceutical technologies that received a positive recommendation from NICE based on RCT or non-RCT data (83% vs 86%). This suggests that even NICE recognises the need to use evidence from non-randomised study designs to inform decision-making, although the quality of such evidence still requires substantial scrutiny and appropriate statistical methodology. For an example of a HTA that uses considerable evidence from single-arm trials using statistical comparator techniques, see Llewellyn et al. [149].

10 Conclusions

The time and budgetary restrictions placed on decision makers might mean that service evaluations are required as a vehicle for producing VfM evidence. As such evidence tends not to be peer reviewed and without formal guidance related to VfM analyses for service evaluations, there is the opportunity for sub-optimal analyses to be carried out. Although NICE and other guidance specific to economic evaluations might not perfectly fit these analyses to inform all decision makers, there are some fundamental aspects that should be taken into account including study design, data collection methods and sources, statistical methods and reporting standards. The use of study designs and statistical methods to account for confounding factors and potential biases, and methods to control and report on uncertainty around estimates and opportunity costs, are important aspects to consider. In terms of costs, even if considered outside the scope of the decision maker, related future costs should be included in the evaluation alongside intervention costs to account for potential opportunity costs in a care system that obtains its funding from the same overall care budget. In terms of relevant outcomes and associated VfM method, although alternative VfM analyses than CUA might be considered more appropriate or practical to use, CUA should be given priority. Alternative methods could then be rationalised, but still reported to current standards expected from using the CHEERS checklist. Accounting for the time horizon of the decision problem is also important, which for longer time horizons could be accounted for using statistical and/or modelling-based methods. However, it is important to note that for some decision makers, the time horizon of interest may be more immediate short-term gains (i.e. over 1 year) than longer term planning. We suggest that the time horizon over which all costs and effects relevant to the decision problem occur should be considered for the evaluation, with estimates reported over a relevant short (e.g. 1 year) and long term (e.g. 20 years). Value of information methods can then be used to monetarise the decision uncertainty over the relevant time horizons.

References

1. Akobeng AK. Principles of evidence based medicine. Arch Dis Child. 2005;90(8):837–40.
2. Hunink MM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, et al. Decision making in health and medicine: integrating evidence and values. Cambridge: Cambridge University Press; 2014.
3. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: 2013.
4. Rowen D, Zouraq IA, Chevrou-Severac H, van Hout B. International regulations and recommendations for utility data for health technology assessment. Pharmacoeconomics. 2017;35(1):11–9.
5. Sculpher M, Palmer S. After 20 years of using economic evaluation, should NICE be considered a methods innovator? Pharmacoeconomics. 2020;38(3):247–57.
6. Akobeng A. Understanding randomised controlled trials. Arch Dis Child. 2005;90(8):840–4.
7. Research Ethics Service. Defining research. 2017. https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2017-1.pdf. Accessed 22 Feb 2019.
8. Petrou S. Rationale and methodology for trial-based economic evaluation. Clin Invest. 2012;2(12):1191–200.
9. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006;15(7):677–87.
10. Gevers S. Medical research involving human subjects: towards an international legal framework. Eur J Health Law. 2001;8:293–8.
11. Hunter D. Efficiency and the proposed reforms to the NHS research ethics system. J Med Ethics. 2007;33(11):651–4.
12. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25):1887–922.
13. Frieden TR. Evidence for health decision making: beyond randomized, controlled trials. N Engl J Med. 2017;377(5):465–75.
14. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard: lessons from the history of RCTs. N Engl J Med. 2016;374(22):2175–81.
15. Chavez-MacGregor M, Giordano SH. Randomized clinical trials and observational studies: is there a battle? J Clin Oncol. 2016;34(8):772–3.
16. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Clin Epidemiol. 2009;62(5):499–505.
17. Spieth PM, Kubasch AS, Penzlin AI, Illigens BM-W, Barlinn K, Siepmann T. Randomized controlled trials: a matter of design. Neuropsychiatr Dis Treat. 2016;12:1341–9.
18. Legislation.gov.uk. Health and Social Care Act 2012. 2012. https://www.legislation.gov.uk/ukpga/2012/7/contents/enacted. Accessed 5 May 2019.
19. NHS England. NHS England Test Beds Programme: evaluation learning from Wave 1. 2018. https://www.england.nhs.uk/wp-content/uploads/2018/11/test-beds-programme-evaluation-learning-from-wave-1.pdf. Accessed 17 Oct 2019.
20. Van den Broucke S. Implementing health in all policies post Helsinki 2013: why, what, who and how. Health Promot Int. 2013;28(3):281–4.
21. World Health Organization. Health in all policies: Helsinki statement. Framework for country action. Helsinki; 2014: Report No. 9241506903.
22. Davies SC, Walley T, Smye S, Cotterill L, Whitty CJ. The NIHR at 10: transforming clinical research. Clin Med. 2016;16(6):501–2.
23. Rovithis D. Do health economic evaluations using observational data provide reliable assessment of treatment effects? Health Econ Rev. 2013;3(1):21.
24. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191–4.
25. NHS HRA. Research Ethics Committee: standard operating procedures (update: March 2020). 2020. https://www.hra.nhs.uk/about-us/committees-and-services/res-and-recs/research-ethics-committee-standard-operating-procedures/. Accessed 9 Mar 2020.
26. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4th ed. Oxford: Oxford University Press; 2015.
27. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, 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.
28. Franklin M, Lomas J, Walker S, Young T. An educational review about using cost data for the purpose of cost-effectiveness analysis. Pharmacoeconomics. 2019;37(5):631–43.

29. Booth N. On value frameworks and opportunity costs in health technology assessment. Int J Technol Assess Health Care. 2019;35(5):367–72.

30. Eddama O, Coast J. A systematic review of the use of economic evaluation in local decision-making. Health Policy. 2008;86(2–3):129–41.

31. Eddama O, Coast J. Use of economic evaluation in local health care decision-making in England: a qualitative investigation. Health Policy. 2009;89(3):261–70.

32. Department of Health and Social Care. The NHS constitution for England. 2015. https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england. Accessed 9 Mar 2020.

33. Sutton M, Garfield-Birkbeck S, Martin G, Meacock R, Morris S, Sculpher M, et al. Economic analysis of service and delivery interventions in health care. Health Serv Deliv Res. 2018;4:16. https://doi.org/10.3310/hsdr06050.

34. Kristensen FB, Husereau D, Huić M, Drummond M, Berger ML, Bond K, et al. Identifying the need for good practices in health technology assessment: summary of the ISPOR HTA Council Working Group Report on Good Practices in HTA. Value Health. 2019;22(1):13–20.

35. Edwards RT, McIntosh E. Applied health economics for public health practice and research. 2019.

36. Weatherly H, Drummond M, Claxton K, Cookson R, Ferguson B, Godfrey C, et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. Health Policy. 2009;93(2–3):85–92.

37. Edwards RT, Charles JM, Lloyd-Williams H. Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. BMC Public Health. 2013;13(1):1001.

38. Schaffer SK, West P, Towse A, Henshall C, Mestre-Ferrandiz J, Masterton R, et al. Assessing the value of new antibiotics: additional elements of value for health technology assessment decisions. London: The Office of Health Economics; 2017.

39. Schaafsma JD, van der Graaf Y, Rinkel GJ, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. J Clin Epidemiol. 2009;62(12):1248–52.

40. Drummond M, Griffin A, Tarricone R. Economic evaluation for devices and drugs: same or different? Value Health. 2009;12(4):402–4.

41. Buchanan J, Wordsworth S, Schuh A. Issues surrounding the health economic evaluation of genomic technologies. Pharmacogenomics. 2013;14(15):1833–47.

42. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and undertaking a health economics study of digital health interventions. Am J Prev Med. 2016;51(5):852–60.

43. Bojke L, Schmitt L, Lomas J, Richardson G, Weatherly H. Economic evaluation of environmental interventions: reflections on methodological challenges and developments. Int J Environ Res Public Health. 2018;15(11):2459.

44. Meacock R. Methods for the economic evaluation of changes to the organisation and delivery of health services: principal challenges and recommendations. Health Econ Policy Law. 2019;14(1):119–34.

45. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? Health Econ. 2001;10(2):179–84.

46. NICE Decision Support Unit (DSU). Cost minimisation. 2019. https://nicedsu.org.uk/cost-minimisation/. Accessed 16 Dec 2019.

47. Mauskopf JA, Paul JE, Grant DM, Stergachis A. The role of cost-consequence analysis in healthcare decision-making. Pharmacoeconomics. 1998;13(3):277–88.

48. Sullivan SD, Mauskopf JA, Augustovski F, Caro JJ, Lee KM, Minchin M, et al. Budget impact analysis: principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health. 2014;17(1):5–14.

49. National Institute for Health and Care Excellence (NICE). Medical technologies evaluation programme methods guide. London: National Institute for Health and Care Excellence (NICE); 2017.

50. Optimity Advisors. Community engagement: approaches to improve health and reduce health inequalities: cost-consequence analysis. National Institute for Health and Care Excellence (NICE); 2016.

51. Williams I, McIver S, Moore D, Bryan S. The use of economic evaluations in NHS decision-making: a review and empirical investigation. Health Technol Assess. 2008;12(7):1–175.

52. Loomes G, McKenzie L. The use of QALYs in health care decision making. Soc Sci Med. 1989;28(4):299–308.

53. NICE. Budget impact test. 2017; https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/budget-impact-test. Accessed 1 Apr 2020.

54. NICE. Evidence standards for digital health technologies. ; 2019.

55. York Health Economics Consortium (YHEC). National Institute for Health and Care Excellence evidence standards framework for digital health technologies: cost consequences and budget impact analyses and data sources. England: York; 2019.

56. Masters R, Anwar E, Collins B, Cookson R, Capewell S. Return on investment of public health interventions: a systematic review. J Epidemiol Commun Health. 2017;71(8):827–34.

57. NICE. Return on investment tools: beta versions. 2019. https://www.nice.org.uk/about/what-we-do/into-practice/return-on-investment-tools. Accessed 1 Apr 2020.

58. Public Health England (PHE). Health economics: a guide for public health teams. 2018. https://www.gov.uk/guidance/health-economics-a-guide-for-public-health-teams#the-cost-effectiveness-of-specific-topic-areas. Accessed 1 Apr 2020.

59. Nicholls J, Mackenzie S, Somers A. Measuring real value: a DIY guide to social return on investment. London: New Economics Foundation; 2007.

60. Social Value UK. Global value exchange. 2019. https://www.socialvalueuk.org/resources/global-value-exchange/. Accessed 17 Oct 2019.

61. Millar R, Hall K. Social return on investment (SROI) and performance measurement: the opportunities and barriers for social enterprises in health and social care. Public Manag Rev. 2013;15(6):923–41.

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

63. de Vries LM, van Baal PH, Brouwer WB. Future costs in cost-effectiveness analyses: past, present, future. Pharmacoeconomics. 2019;37(2):119–30.

64. Grima DT, Bernard LM, Dunn ES, McFarlane PA, Mendelssohn DC. Cost-effectiveness analysis of therapies for chronic kidney disease patients on dialysis. Pharmacoeconomics. 2012;30(11):981–9.

65. van Baal P, Meltzer D, Brouwer W. Pharmacoeconomic guidelines should prescribe inclusion of indirect medical costs! A response to Grima et al. Pharmacoeconomics. 2013;31(5):369–73.

66. Ten JB. arguments for a societal perspective in the economic evaluation of medical innovations. Eur J Health Econ. 2009;10(4):357–9.

67. Walker S, Griffin S, Asaria M, Tsuchiya A, Sculpher M. Striving for a societal perspective: a framework for economic evaluations when costs and effects fall on multiple sectors and decision makers. Appl Health Econ Health Policy. 2019;17(5):577–90.

68. Franklin M, Thorn J. Self-reported and routinely collected electronic healthcare resource-use data for trial-based economic evaluations: the current state of play in England and considerations for the future. BMC Med Res Methodol. 2019;19(1):8.

69. Whyte S, Dixon S, Faria R, Walker S, Palmer S, Sculpher M, et al. Estimating the cost-effectiveness of implementation: is sufficient evidence available? Value Health. 2016;19(2):138–44.

70. Hoomans T, Evers SM, Ament AJ, Hübben MW, Van Der Weijden T, Grimshaw JM, et al. The methodological quality of economic evaluations of guideline implementation into clinical practice: a systematic review of empiric studies. Value Health. 2007;10(4):305–16.

71. Vale L, Thomas R, MacLennan G, Grimshaw J. Systematic review of economic evaluations and cost analyses of guideline implementation strategies. Eur J Health Econ. 2007;8(2):111–21.

72. Hoomans T, Severens JL. Economic evaluation of implementation strategies in health care. Implement Sci. 2014;9:168.

73. CPRD. Clinical practice research datalink. 2019. https://www.cprd.com/. Accessed 16 May 2019.

74. NHS Digital. Secondary uses service (SUS). 2019. https://digital.nhs.uk/services/secondary-uses-service-sus. Accessed 17 Oct 2019.

75. Franklin M, Berdunov V, Edmans J, Conroy S, Gladman J, Tanajewski L, et al. Identifying patient-level health and social care costs for older adults discharged from acute medical units in England. Age Ageing. 2014;43(5):703–7.

76. Franklin M, Davis S, Horspool M, Kua WS, Julious S. Economic evaluations alongside efficient study designs using large observational datasets: the PLEASANT trial case study. Pharmacoeconomics. 2017;35(5):561–73.

77. Rees A, Paisley S, Brazier J, Cantrell A. Development of the Scharr HUD (Health Utilities Database). Value Health. 2013;16:A580. https://doi.org/10.1016/j.jval.2013.08.1585.

78. Longworth L, Rowen D. NICE DSU technical support document 10: the use of mapping methods to estimate health state utility values. Sheffield: University of Sheffield; 2011.

79. Mukuria C, Rowen D, Harnan S, Rawdin A, Wong R, Ara R, et al. An updated systematic review of studies mapping (or crosswalking) measures of health-related quality of life to generic preference-based measures to generate utility values. Appl Health Econ Health Policy. 2019;17(3):295–313.

80. Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. Health Qual Life Outcomes. 2018;16:31.

81. Hinde S, Bojke L, Richardson G. Understanding and addressing the challenges of conducting quantitative evaluation at a local level: a worked example of the available approaches. BMJ Open. 2019;9(11):e029830.

82. Coggon D, Rose G, Barker D. Epidemiology for the uninitiated, 4th ed. BMJ; 1997. Accessed 2019.

83. Dekkers OM, Egger M, Altman DG, Vandenbroucke JP. Distinguishing case series from cohort studies. Ann Intern Med. 2012;156(1):37–40.

84. Goodacre S. Uncontrolled before-after studies: discouraged by Cochrane and the EMJ. Emerg Med J. 2015;32(7):507–8.

85. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med. 1982;72(2):233–40.

86. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. Lancet. 2002;359(9301):145–9.

87. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol. 2006;35(2):337–44.

88. Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. Soc Sci Med. 2019;220:353–61.

89. Last JM, Spasoff RA, Harris SS, Thuriaux MC. A dictionary of epidemiology. New York: International Epidemiological Association, Inc; 2001.

90. Faria R, Alava MH, Manca A, Wailoo AJ. NICE Decision Support Unit (DSU) Technical Support Document (TSD) 17: the use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. Sheffield: National Institute for Health and Care Excellence (NICE); 2015.

91. Bell H, Wailoo AJ, Hernandez M, Grieve R, Faria R, Gibson L, et al. NICE Decision Support Unit (DSU) Technical Support Document (TSD): the use of real world data for the estimation of treatment effects in NICE decision making. Sheffield: National Institute for Health and Care Excellence (NICE); 2016.

92. Manca A, Austin PC. Using propensity score methods to analyse individual patient level cost effectiveness data from observational studies. York: The University of York; 2008.

93. Kreif N, Grieve R, Radice R, Sekhon JS. Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation. Health Serv Outcomes Res Methodol. 2013;13(2–4):174–202.

94. Crown WH. Propensity-score matching in economic analyses: comparison with regression models, instrumental variables, residual inclusion, differences-in-differences, and decomposition methods. Appl Health Econ Health Policy. 2014;12(1):7–8.

95. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ. 2019;367:l5657.

96. Franklin M, Wailoo A, Dayer MJ, Jones S, Prendergast B, Baddour LM, et al. The cost-effectiveness of antibiotic prophylaxis for patients at risk of infective endocarditis. Circulation. 2016;134(20):1568–78.

97. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348–55.

98. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N. An educational review of the statistical issues in analysing utility data for cost-utility analysis. Pharmacoeconomics. 2015;33(4):355–66.

99. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. Stat Med. 1999;18(15):1903–42.

100. Briggs A, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess. 1999;3(2):1–134.

101. Culyer AJ. Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use. Health Econ Policy Law. 2016;11(4):415–32.

102. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold. Pharmacoeconomics. 2008;26(9):733–44.

103. Latimer NR. Survival analysis for economic evaluations alongside clinical trials: extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Med Decis Mak. 2013;33(6):743–54.

104. Young TA. Estimating mean total costs in the presence of censoring. Pharmacoeconomics. 2005;23(12):1229–422.

105. Willan AR, Lin D, Manca A. Regression methods for cost-effectiveness analysis with censored data. Stat Med. 2005;24(1):131–45.

106. Wijeysundera HC, Wang X, Tomlinson G, Ko DT, Krahn MD. Techniques for estimating health care costs with censored data: an overview for the health services researcher. Clinicoecon Outcomes Res. 2012;4:145–55.

107. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

108. Franklin M, Hunter RM. A modelling-based economic evaluation of primary-care-based fall-risk screening followed by fall-prevention intervention: a cohort-based Markov model stratified by older age groups. Age Ageing. 2019;49(1):57–66.

109. Hunter RM, Fulop NJ, Boaden RJ, McKevitt C, Perry C, Ramsay AI, et al. The potential role of cost-utility analysis in the decision to implement major system change in acute stroke services in metropolitan areas in England. Health Res Policy Syst. 2018;16(1):23.

110. Jackson C, Presanis A, Conti S, De Angelis D. Value of information: Sensitivity analysis and research design in Bayesian evidence synthesis. J Am Stat Assoc. 2019;114(528):1436–49.

111. Heath A, Manolopoulou I, Baio G. A review of methods for analysis of the expected value of information. Med Decis Making. 2017;37(7):747–58.

112. Strong M, Oakley JE, Brennan A. Estimating multi-parameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a non-parametric regression approach. Med Decis Making. 2014;34(3):311–26.

113. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: a fast, nonparametric regression-based method. Med Decis Mak. 2015;35(5):570–83.

114. Wilson EC. A practical guide to value of information analysis. Pharmacoeconomics. 2015;33(2):105–21.

115. Rothery C, Strong M, Koffijberg H, Basu A, Ghabri S, Knies S, Murray JF, Schmidler GDS, Steuten L, Fenwick E. Value of information analytical methods: report 2 of the ISPOR Value of information analysis emerging good practices task force. Value Health. 2020;23(3):277–86.

116. Cox M, O'Connor C, Biggs K, Hind D, Bortolami O, Franklin M, et al. The feasibility of early pulmonary rehabilitation and activity after COPD exacerbations: external pilot randomised controlled trial, qualitative case study and exploratory economic evaluation. Health Technol Assess. 2018;22(11):1–204.

117. NICE. NICE technology appraisal guidance. 2020. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance. Accessed 9 Mar 2020.

118. Kaltenthaler E, Boland A, Carroll C, Dickson R, Fitzgerald P, Papaioannou D. Evidence Review Group approaches to the critical appraisal of manufacturer submissions for the NICE STA process: a mapping study and thematic analysis. Health Technol Assess. 2011;15(22):1–82, iii–iv.

119. NICE. The principles that guide the development of NICE guidance and standards. 2020. https://www.nice.org.uk/about/who-we-are/our-principles. Accessed 9 Mar 2020.

120. World Health Organization. The role of local government in health: comparative experiences and major issues. Geneva: World Health Organization; 1997.

121. Furber A. Should local government run the NHS? BMJ. 2016;355:i5962.

122. Cullis J, Jones P. Public finance and public choice. 2nd ed. New York: Oxford University Press; 1998.

123. Pettitt D, Raza S, Naughton B, Roscoe A, Ramakrishnan A, Ali A, et al. The limitations of QALY: a literature review. J Stem Cell Res Ther. 2016;6:4.

124. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health-related QALY. Appl Health Econ Health Policy. 2015;13(6):557–65.

125. Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. Value Health. 2009;12:S27–30.

126. Asaria M, Griffin S, Cookson R. Distributional cost-effectiveness analysis: a tutorial. Med Decis Making. 2016;36(1):8–19.

127. Asaria M, Griffin S, Cookson R, Whyte S, Tappenden P. Distributional cost-effectiveness analysis of health care programmes: a methodological case study of the UK bowel cancer screening programme. Health Econ. 2015;24(6):742–54.

128. Dawkins BR, Mirelman AJ, Asaria M, Johansson KA, Cookson RA. Distributional cost-effectiveness analysis in low-and middle-income countries: illustrative example of rotavirus vaccination in Ethiopia. Health Policy Plan. 2018;33(3):456–63.

129. Barnsley P, Towse A, Sussex J. Critique of CHE research paper 81: methods for the estimation of the NICE cost effectiveness threshold. London: Office of Health Economics (OHE); 2013.

130. Claxton K, Sculpher M. Response to the OHE critique of CHE Research Paper 81: University of York; 2017.

131. Raftery J. NICE's cost-effectiveness range: should it be lowered? Pharmacoeconomics. 2014;32:613–5.

132. 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 Health. 2018;21(3):266–75.

133. 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 Global Health. 2018;3(6):e000964.

134. Love-Koh J, Cookson R, Claxton K, Griffin S. Estimating social variation in the health effects of changes in health care expenditure. Med Decis Mak. 2020;40(2):170–82.

135. Lomas JR. Incorporating affordability concerns within cost-effectiveness analysis for health technology assessment. Value Health. 2019;22(8):898–905.

136. Bilinski A, Neumann P, Cohen J, Thorat T, McDaniel K, Salomon JA. When cost-effective interventions are unaffordable: Integrating cost-effectiveness and budget impact in priority setting for global health programs. PLoS Med. 2017;14(10):e1002397.

137. Charlton V, Littlejohns P, Kieslich K, Mitchell P, Rumbold B, Weae A, et al. Cost effective but unaffordable: an emerging challenge for health systems. BMJ. 2017;22(356):j1402.

138. Pearson SD. The ICER value framework: integrating cost effectiveness and affordability in the assessment of health care value. Value Health. 2018;21(3):258–65.

139. Detsky AS, Laupacis A. Relevance of cost-effectiveness analysis to clinicians and policy makers. JAMA. 2007;298(2):221–4.

140. NIHR. Journals. https://www.journalslibrary.nihr.ac.uk/journals/. Accessed 9 Mar 2020.

141. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS): explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. Value Health. 2013;16(2):231–50.

142. Grieve R, Abrams K, Claxton K, Goldacre B, James N, Nicholl J, et al. Cancer Drugs Fund requires further reform. BMJ. 2016;354:i5090.

143. Yeh RW, Valsdottir LR, Yeh MW, Shen C, Kramer DB, Strom JB, et al. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. BMJ. 2018;363:k5094.

144. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics. 2014;32(4):345–65.

145. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making: an introduction: report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19(1):1–13.

146. Marsh K, Izerman M, Thokala P, Baltussen R, Boysen M, Kaló Z, et al. Multiple criteria decision analysis for health care decision making: emerging good practices: report 2 of the ISPOR

MCDA Emerging Good Practices Task Force. Value Health. 2016;19(2):125–37.

147. Anderson R, Hardwick R. Realism and resources: towards more explanatory economic evaluation. Evaluation. 2016;22(3):323–41.

148. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of NICE appraisals of pharmaceuticals 2000–2016 found variation in establishing comparative clinical effectiveness. J Clin Epidemiol. 2019;105:50–9.

149. Llewellyn A, Faria R, Woods B, Simmonds M, Lomas J, Woolacott N, et al. Daclatasvir for the treatment of chronic hepatitis C: a critique of the clinical and economic evidence. Pharmacoeconomics. 2016;34(10):981–92.