

This is a repository copy of *An educational review about using cost data for the purpose of cost-effectiveness analysis*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/141558/>

Version: Published Version

Article:

Franklin, Matthew, Lomas, James Richard Scott orcid.org/0000-0002-2478-7018, Walker, Simon Mark orcid.org/0000-0002-5750-3691 et al. (1 more author) (Accepted: 2019) An educational review about using cost data for the purpose of cost-effectiveness analysis. *PharmacoEconomics*. ISSN: 1179-2027 (In Press)

<https://doi.org/10.1007/s4027>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis

Matthew Franklin¹ · James Lomas² · Simon Walker² · Tracey Young¹

© The Author(s) 2019

Abstract

This paper provides an educational review covering the consideration of costs for cost-effectiveness analysis (CEA), summarising relevant methods and research from the published literature. Cost data are typically generated by applying appropriate unit costs to healthcare resource-use data for patients. Trial-based evaluations and decision analytic modelling represent the two main vehicles for CEA. The costs to consider will depend on the perspective taken, with conflicting recommendations ranging from focusing solely on healthcare to the broader ‘societal’ perspective. Alternative sources of resource-use are available, including medical records and forms completed by researchers or patients. Different methods are available for the statistical analysis of cost data, although consideration needs to be given to the appropriate methods, given cost data are typically non-normal with a mass point at zero and a long right-hand tail. The choice of covariates for inclusion in econometric models also needs careful consideration, focusing on those that are influential and that will improve balance and precision. Where data are missing, it is important to consider the type of missingness and then apply appropriate analytical methods, such as imputation. Uncertainty around costs should also be reflected to allow for consideration on the impacts of the CEA results on decision uncertainty. Costs should be discounted to account for differential timing, and are typically inflated to a common cost year. The choice of methods and sources of information used when accounting for cost information within CEA will have an effect on the subsequent cost-effectiveness results and how information is presented to decision makers. It is important that the most appropriate methods are used as overlooking the complicated nature of cost data could lead to inaccurate information being given to decision makers.

Key Points for Decision Makers

Not appropriately controlling for the nature of cost data and the uncertainty around subsequent cost estimates used in cost-effectiveness analysis (CEA) could lead to inaccurate information being given to decision makers.

Although checklists exist for use alongside CEA to aid transparency in the methods used, there is still poor reporting and rationalisation of statistical methods and covariate adjustments when using cost data.

It is difficult to suggest a ‘one size fits all’ methodology when estimating and analysing cost data for CEA; therefore, it is down to the researcher to assess the nature of the cost data to determine which methods to use.

1 Introduction

Economic evaluation is widely used for the appraisal of healthcare programmes, taking into account both the costs and the effects or outcomes. There are multiple forms of economic evaluation, such as cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and, as a subform of CEA, cost-utility analysis (CUA). However, all forms of economic evaluation are related to ‘value for money’, with ‘costs’ representing an integral aspect of the evaluation process owing to the resultant opportunity costs from resources not being available for other purposes [1]. In the UK, CEA, particularly CUA as the National Institute for Health and Care Excellence (NICE) economic evaluation reference case [2], has become the most common form of economic evaluation. This is also reflected within reimbursement agency guidance internationally [3]. Therefore, this paper focuses specifically on CEA, although many of the aspects described and discussed are relevant to other forms of economic evaluation, such as CBA.

A large amount of research has already been dedicated to how we measure and statistically estimate outcomes for CEA, with particular focus often on utility data and

✉ 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 YO10 5DD, UK

the quality-adjusted life-year (QALY) associated with CUA [4]; in 2017, *Pharmacoeconomics* published a special issue titled ‘Estimating Utility Values for Economic Evaluation’ on this exact subject [5]. Guidance on how we do the same for costs has become dated, do not reflect advances in CEA more generally, new methods are not presented in a user friendly manner in a single source (i.e. methods are often spread across multiple and various papers and publications) and there is not always consensus on the ‘best’ method to use. As a result, some useful and new methods have not been widely adopted.

This educational review summarises relevant methods and research from the published literature about how to identify, estimate and analyse relevant cost data based on two specific vehicles for economic evaluation (within-trial and modelling-based analyses), within which we focussed on the following nine topics: (1) the difference between resource use and costs; (2) vehicles for CEA (e.g. within-trial and modelling-based analysis); (3) what costs to include depending on the costing perspective; (4) sources of resource-use data and unit costs; (5) statistical methods for assessing cost data and its distribution; (6) adjusting for baseline covariates; (7) dealing with different types of missing cost data; (8) uncertainty around cost estimates; and (9) a note on discounting, inflation and using relevant currency. We also identify gaps in the literature based on what the authors perceive are important and overlooked considerations when analysing cost data, and therefore suggest future areas for research. This educational review provides a range of references for further reading, and should therefore be used as a reference guide to get an understanding about using cost data for CEA, rather than a technical document describing in detail specific methodologies associated with using cost data.

2 Resource Use and Costs: What is the Difference?

The focus of this paper is cost data, which can be obtained by applying unit costs to resource-use data; therefore, we need to clarify what are (1) resource-use data; (2) unit costs; and (3) cost data.

Resource-use data represent those resources consumed by the person or population with the health condition of interest for the CEA. For example, common types of healthcare resource-use information include number and type of visits to hospital (e.g. inpatient, outpatient, critical care) and any medications prescribed (e.g. type of drug, dose and duration of treatment).

Resource-use information can be analysed in its own right (e.g. as the response variable in regression analysis or as a descriptive statistic), although due to its close relationship

with costs, and costs being one of the key aspects for CEA, it is cost data that is normally the focus of CEA. To generate cost data from resource use, we need to apply appropriate unit costs to each type of resource use. Unit costs are the cost per unit of item of resource use (e.g. there is a different unit cost for seeing a doctor compared with seeing a nurse) and represent the monetary value of that item (e.g. if a consultation with a doctor has a unit cost of £31, this represents the monetary cost of a consultation). Sources of resource-use data and unit costs are described in more detail in Sect. 5.

3 Vehicles for Cost-Effectiveness Analysis (CEA): Within-Trial and Modelling-Based Analysis

Evidence-based healthcare generally relies on the use of clinical studies to provide the necessary clinical and economic information. Randomised controlled trials (RCTs) are regarded as the ‘gold standard’ type of clinical study [6]. An economic evaluation that uses mainly information from the trial and estimates costs and effects over the trial period is referred to as a ‘within-trial economic evaluation’ [7]; this is in contrast to synthesising evidence from multiple sources (e.g. RCTs, empirical literature, and expert opinion) to estimate costs and effects in ‘modelling-based analyses’ [8].

The focus on RCTs as a ‘gold standard’ has given a perceived level of ‘reliability’ to the evidence they produce, including the associated within-trial CEA; however, they have their limitations. Economic evaluation to inform effective decision making requires that costs and effects can be estimated for all relevant comparators over an appropriate time horizon [9]. Furthermore, all available evidence and potential sources of heterogeneity in cost effectiveness should be considered [10]. RCTs typically consider only a few of the relevant comparators over a short time horizon, often in a particular group of patients with limited external validity of trial results. As a result of these limitations, decision analytic modelling methods are now widely used, although the two vehicles can be complementary such that a decision model can be used to expand on those aspects not addressed as part of the within-trial analysis (e.g. evaluating over a longer time horizon). Models provide a framework for synthesising evidence, comparing all relevant comparators and estimating costs and effects over an appropriate time horizon [11]. A limitation of modelling is that they can be considered a ‘black box’ where the inputs and outputs (and everything in between) may not be transparent to decision makers interested in the CEA. Furthermore, models, by necessity, represent simplified versions of reality and may not be considered to accurately reflect real life. However, these limitations can be partially avoided by being transparent with all methods and sources of information used when

conducting a CEA and following prespecified checklists during the evaluation and write-up process, e.g. the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement [12], which provides guidance on how to report economic evaluations more consistently and with transparency, and the Philips et al. [13] checklist based on a review of guidelines for good practice in decision-analytic modelling for health technology assessment (HTA). Other authors have also suggested pragmatic steps to improve the validation of models [14].

Many different types of decision analytic model have been used for economic evaluation [8, 11, 13, 15], including simple decision trees, state-transition models (e.g. Markov models) [16], individual sampling models (e.g. discrete event simulations [17]), and more complex structures (e.g. dynamic transmission models [18]). Despite differences, the approaches all aim to reflect the possible prognoses of individuals being modelled through the initial interventions they receive, the time they spend in particular health states, and the clinical events they experience, all of which are associated with specified costs. For example, if the initial intervention was a drug treatment, the cost would typically involve defining the dose (which may be conditional on patient characteristics such as weight or body surface area), treatment duration (which may be conditional on survival from the model), and the associated unit cost.

Costs can be applied to specific health states, whereby a health state can represent, for example, a declined state of health (e.g. New York Heart Association Functional Classification states [19]), or be defined by the healthcare received (e.g. outpatient, inpatient, or moving to a care home). Costs associated with being in a health state would usually be calculated by estimating the average costs over a defined time period in that state (e.g. in a Markov model, a cycle length), and then the total health state costs are estimated conditional on the time spent in that health state over the defined time period. Most commonly, the resources associated with being in that health state (e.g. the average number of primary and secondary care appointments associated with a year with stable angina) are estimated and then unit costs are applied to these to estimate the total cost over the defined time period [20]. Methods for the estimation of resource use (described in more detail in Sect. 5) vary from the analysis of resource-use data from patients to the use of clinical opinion [20]. The costs of an event are typically applied as a one-off cost in models and reflect the resource use and costs associated with treating an event (e.g. the cost of treating a myocardial infarction). The resources associated with the treatment of an event are estimated and the unit costs applied to estimate the total cost of an event.

With the costing of health states and events, there is also the possibility of using longitudinal data and panel data analysis to estimate the background costs of being in a health state (using

dummy variables for patients in particular health states in a given period) and the costs of events (again using a dummy in the cycle in which the event occurred) [21]. There is also a growing interest in how observational and real-world data, such as that from ‘Big Data’ (a term often used nowadays to refer to extremely large data sets that include a much higher number of data points, parameters, and/or people than what was historically available for analysis), can be used for the purpose of economic analysis rather than relying on trials. Big Data can also be used to conduct either a (observational) within-study analysis or inform an economic model, the statistical considerations for which are different to when conducting an analysis alongside a trial or a model. As this is a growing area, observational studies as a vehicle for CEA are not extensively described within this manuscript, but form part of a discussion point in Sect. 11. A framework for conducting economic evaluations alongside natural experiments (i.e. “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” [22]), the resource-use data requirements for which will typically rely on the existence of longitudinal ‘Big Data’, has been described by Deidda et al. [23].

4 What Costs to Include: A Brief Introduction to Costing Perspectives

Cost-effectiveness analysis seeks to provide analysis to inform decision making, and crucial to which costs should be included or excluded is the decision makers themselves, their priorities, and their budgets [11]. Thinking about the choice of relevant costs also presents issues of pragmatism; for example, costs common to all mutually exclusive interventions may not be relevant to the decision being informed, since these costs will not impact on incremental costs. From a healthcare or payer perspective, it is important to include costs beyond the initial direct cost (e.g. the acquisition cost of a new pharmaceutical product) and to include other healthcare costs within the time horizon adopted for the analysis. There is a debate about whether to include all future healthcare costs or just those in related disease areas [24, 25], with conflicting recommendations from the US Panel stating all healthcare costs being included [26], and NICE stating that only related future costs should be included [2]. The US Panel also advocates the use of a ‘societal’ perspective, which entails incorporating additional costs into CEA beyond those in the healthcare sector; for example, costs of informal care and lost productivity [26]. The way in which costs from different sectors can be aggregated for use in this perspective is also debated in emerging literature [27, 28]. Finally, irrespective of perspective,

there are questions relating to which types of costs should be incorporated within economic evaluations; for example, if sunk and fixed costs should be included. This may be particularly relevant to evaluations involving large investments, in which case the question becomes what approach should be taken: (1) a ‘short-run’ approach whereby factors of production are fixed, such as capital (e.g. large medical scanners such as magnetic resonance imaging (MRI) machines); or (2) a long-run approach whereby factors of production are not fixed. A useful discussion of these issues can be found in the articles by Drummond et al. [11] and Culyer [29].

5 Sources of Resource-Use Data and Unit Costs: Examples from England

In order to assess costs for CEA, there is a need for accessible and good-quality data that reflect the care people receive. Alongside trials, Ridyard and Hughes [30] identified the following five data collection methods.

1. Medical records (e.g. patient notes, large databases).
2. Prospective forms completed by trial researchers or healthcare professionals:
 - 2.1. not based on patient recall or abstracted from routine sources, or
 - 2.2. based on patient recall.
3. Patient-completed or carer-/non-healthcare professional-completed:
 - 3.1. diaries, or
 - 3.2. forms.

The majority of studies (61/85) identified by Ridyard and Hughes [30] used at least two methods, typically involving patient- or carer-completed forms and medical records (the latter referred to here as ‘routinely collected care data’); for the purpose of discussion, it is these two methods in England that are the focus. Regarding data sources in non-UK settings, a brief discussion of conducting economic evaluations alongside multinational trials (which includes obtaining resource-use and unit cost information) is provided in Sect. 10 for consideration by the interested reader.

Instruments for person-reported cost measurement have been developed, with the Client Service Receipt Inventory (CSRI) the most commonly used example [31], however it has been adapted/modified many times (estimated to be over 200) and is therefore neither fully standardised nor universally applicable [32]. Standardising instruments may be considered useful for cross-study comparisons, allowing for easier comparison across CEAs, which could be perceived

as similar to the debate around using generic outcome measures for CEA [33, 34] and the EQ-5D as the NICE reference case [2, 35]; the counter to which could also be perceived as similar to those rationalisations for using condition-specific measures for outcome measurement [36]—that there are specific aspects of resource use and costs that are important in certain, but not all, decision problems. As such, standardised resource-use measurement instruments may not be specific enough to capture all important resources for particular decision problems. There are other examples of established resource-use instruments [37–39]—those in development [40], and the Database of Instruments for Resource-Use Measurement (DIRUM), which was created as a repository for instruments based on patient recall (<http://www.dirum.org>) [41]. Researchers tend to tailor instruments to focus on specific information needed for analysis, which is particularly important when taking a broader perspective, such as information related to informal care and other wider aspects that are not routinely collected within medical records or in large databases.

The method of using routinely collected data can include anything from raw data extracted straight from a service, to the level of linked datasets across multiple services. Examples of using raw data extracted from multiple care services (i.e. primary, secondary, and social care services) for the purpose of a costing study include the study by Franklin et al. [42] and two subsequent RCT-based CEAs [43, 44]. Within England, National Health Service (NHS) Digital is one major provider of routinely collected data. The most widely used NHS Digital dataset is Hospital Episode Statistics (HES) [45]. For primary care, McDonnell et al. [46] have amalgamated a list of UK datasets that may be of interest (not all contain resource-use information). An issue with electronic datasets is that they may only adequately record data for that particular service; thus, previous studies have described the need for linked data [47–49]. Asaria et al. [50] have described key challenges and opportunities of using linked records to estimate healthcare costs. There are also concerns with data quality and missingness within such large datasets that require assessment and validation [51, 52], and the exchangeability of using person-reported or routine care data is also questionable as they may produce quite different estimates [53, 54]. With the advent of routinely collected data, there is a growing interest in how to utilise these data in trials, observational studies, and modelling, the methodology for which will develop as these data become more accessible and reliable [48, 55]. It is important to note that no data collection method can currently be considered the ‘gold standard’. The trade-off between self-reported and routinely collected electronic data to inform the choice of method for trial-based evaluations has been described by Franklin and

Thorn [55]. The same type of discussion relevant to developing economic models is currently lacking.

There are some common unit cost sources in the UK, including NHS Reference Costs [56], Personal Social Services Research Unit (PSSRU) Unit Costs of Health and Social Care [57], British National Formulary (BNF; for drug costs) [58], and the drugs and pharmaceutical electronic market information tool (eMIT, for generic drugs) [59], among others. NHS reference costs are strongly associated with HES data, particularly in relation to inpatient, outpatient and accident and emergency (A&E) visits. Geue et al. [60] have discussed the implications of using (five) alternative methods for costing hospital episode statistics, and recommend using a Health Resource Grouper (HRG) costing method that can be directly linked with NHS Reference Costs; we also recommend using this method.

6 Statistical Methods for Assessing Cost Data and Its Distribution

As Briggs and Gray [20] point out, within CEA, the parameter of interest is usually the mean cost (or difference between group means) or, less frequently, mean resource use. This is because such research is typically aimed at estimating the expected costs (and associated cost effectiveness) of alternative comparators, reflecting that costs will differ across individuals but it is the mean cost that is important. Cost data are typically non-normal in their distribution, with a mass point at zero, non-negative, and with a long right-hand tail, with the implicit underlying data-generating process resulting in heteroskedastic errors, and non-linear responses to covariates [61]. According to central limit theorem [62], in large samples, estimated mean cost parameters are approximately normally distributed despite the non-normality of the underlying data, which would mean that standard methods based on ordinary least squares (OLS) and the t distribution could be appropriate for characterising the uncertainty in the estimate of the mean. However, with finite samples it is not clear when the sample size is 'large enough' to use these standard methods, especially when data are skewed [20].

Regression methods are used to analyse observable heterogeneity and control for chance imbalance between trial arms to provide an unbiased estimate of the mean cost or incremental cost [63]. Important issues around how to choose covariates are discussed in Sect. 7. In this section, we focus on issues concerning the choice of regression method. A number of alternative approaches have been developed for addressing questions regarding cost data using statistical and econometric methods. A useful review of these methods, particularly for application to datasets from trials, has been conducted by Mihaylova et al. [64]. A comprehensive evaluation of many of these methods using UK observational

data, conducted by Jones et al. [65], found that no single method is dominant and that trade-offs between bias (performance of predictions on average) and accuracy (performance of individual predictions) exist. For researchers performing regression analysis on resource use, different regression methods may be more appropriate, e.g. methods for count data such as negative binomial regression [66]. Depending on the nature of the response variable, other approaches may be warranted, but often the same methods used for costs are transferable [61].

Cost data often possess a large number of zero costs, although there are examples where cost data are not likely to have zero costs (e.g. advanced-stage cancer patients [67, 68]). Literature is available on how best to deal with the presence of a large number of zero costs, which includes the use of a two-part model where separate regressions analyse the chance of non-zero observations and the value of a non-zero observation [66]. It is not favoured to use selection models in this context, although they can and have been used, because zero costs are not the result of censoring but are instead generally considered to be genuine zeros [69].

In the context of economic evaluation, transforming data either via logarithmic transformation or some other function is not favoured [70]. The principle reason for this is that the arithmetic mean cost is the parameter of interest and that retransforming costs to the natural scale requires difficult calculations, particularly when there is heteroskedasticity [70]. In addition, when applying a logarithmic transformation, an adjustment to zero costs or estimation of a two-part model are required. Instead, researchers typically employ generalised linear models (GLMs), which can be applied to data with a mass point at zero and which allow researchers to model costs directly on the scale of interest using a linear or non-linear specification for how covariates affect the mean and variance, with the latter allowing for forms of heteroskedasticity [71]. GLMs comprise a link function, which determines the relationship between a linear index of covariates and the mean, and a distribution function, which determines the relationship between the mean and the variance [70]. While researchers tend to adopt a log-link function with a Gamma distribution function, there is no evidence that this is the dominant form of GLM in terms of model fit for cost data applications, and numerous tests and plots are available to inform model selection [72]. In addition, a more flexible specification of a GLM, the extended estimating equations (EEE) model, can estimate the best-fitting function choices from the data [73]. Despite the flexibility of GLMs available through the choice of link and distribution functions, interested readers may wish to read more around some of the implicit distributional assumptions in GLMs [74, 75] and issues with applying GLMs to heavy-tailed data [76]. Owing to the difficulties in choosing the most appropriate distribution for cost data, researchers may decide to

employ a non-parametric approach to analysis as opposed to a parametric or semi-parametric method. Bootstrapping is one such approach that can be applied to regression, which, instead of making distributional assumptions, makes use of the original data to produce an empirical estimate of the sampling distribution of the statistic in question [20]. Briggs and Gray [20] discuss the validity of this approach, applying the method and contrasting with a transformation approach in five examples. Other methods include model averaging and those based on Bayesian approaches [64].

Cost data in economic evaluation may have additional characteristics that affect the choice of regression method. In particular, in the context of cluster trials, clustering in the data will exist. The appropriate methods for analysing such data include multilevel models, generalised estimating equations (GEE) and two-stage, non-parametric bootstrap [77]. It is worth noting that these approaches can be considered as extensions of OLS, GLMs and bootstrapping. Cost data may also be censored, which is discussed as a special form of missing data in Sect. 8. Finally, cost data are often analysed to inform model parameters, such as the cost of a health state or event, which are likely to be observed at different points in time across patients. Examples of this kind of analysis can be found in the study by Walker et al. [21], and typically require methods for analysing longitudinal data, such as panel fixed effects or GEE (of which panel random effects is one form). Note that analysis of this kind, where analysis is based on an event or health state, assumes causality can be inferred, which may not be the case. If this is not the case, methods for handling unobserved confounding may be required if there is insufficient observable information available to determine causality; an overview of approaches in economic evaluation is presented by Faria et al. [78]. Interested readers may also wish to consider methods that can handle unobserved confounding, along with distributional considerations for cost data (e.g. see the review by Terza et al. [79]).

7 Adjusting for Baseline Covariates: Conceptual and Statistical Considerations

Randomisation is a common method used within trials to control for systematic bias between groups often associated with selection and confounding biases; however, this method will not always result in perfect balance between groups and therefore statistical methods can be used, such as adjusting for covariates. Although almost two decades old, there is guidance on statistical principles for clinical trials that briefly addresses the problem of adjustment for covariates to control for systematic bias and which is still relevant [80]. Key aspects to note from this guidance when adjusting for covariates are:

1. Identify covariates expected to have an important influence on the outcome of interest.
2. Account for these covariates in the analysis in order to improve precision and to compensate for any lack of balance between groups.
3. Apply caution when adjusting for covariates measured after randomisation because they may be affected by the treatments.

It should be noted that adjusting for baseline covariates is not always necessary or recommended [81], and the starting point for any within-trial CEA should be unadjusted analyses; however, in the case of a strong or moderate association between baseline covariate(s) and the outcome of interest, adjustment for such covariate(s) generally improves the efficiency of the analysis and the results should be reported alongside the unadjusted analyses. A pragmatic approach may also be preferred whereby the logistical, time, and resource constraints to collect baseline data may not be possible within a study, particularly if obtaining this data from study participants can overburden the participant (e.g. in terms of time or cognitive burden). As such, researchers should consider the need for baseline adjustment and make appropriate hypotheses rationalising the need for baseline data at the start of the study. There are a number of methods for adjusting for baseline covariates, but regression modelling using baseline covariates as explanatory variables is recommended [81]. There are many types of baseline covariates and their nature depends on the context of the study and the analysis. When analysing trials, the main statistical analysis (i.e. focussed on the trial's primary outcome) will predefine a set of covariates for which to control, and these are a useful starting point for the choice of covariates for the analysis of costs.

In relation to CUA, Manca et al. [82] have discussed the importance of controlling for baseline utilities when the outcome of interest is the (utility-based) QALY. While the focus here is on costs, this does not negate the need to control for other baseline covariates associated with the 'effectiveness' (i.e. QALY or otherwise) aspect of CEA. van Asselt et al. [83] has discussed how to control for baseline cost differences for CEA. They recommend the use of a regression-based adjustment using baseline patient characteristics or, when these are not sufficiently present, baseline costs as a substitute. Franklin et al. [48] have made the case that adjusting for baseline costs is important in their own right, not just as a substitute to patient characteristics. They also suggest that there may be bias related to the data recording mechanisms (routinely collected primary care data in this case study) and 'frequent attenders' (those who utilise services more than the 'norm'), which can be controlled for using a regression-based baseline adjustment. However, before controlling for costs at baseline, there is also the need

to (1) assess the relationship between your baseline covariates, including costs, to avoid collinearity; and (2) consider which covariates to include as hypothesised to be relevant a priori (i.e. per-protocol analysis). In relation to the second point, this does not necessarily restrict the ability to also conduct exploratory, post hoc adjustments with different covariates to assess if there is a relationship (i.e. strength and statistical significance) between baseline variables and the outcomes of interest that could inform future analyses plans. Accounting for various forms of heterogeneity in CEA has been discussed by Sculpher [10] in relation to performing subgroup analysis, although many of the points raised are also relevant when deciding on what covariates to include for the purpose of baseline adjustments and general statistical analysis between groups and subgroups (see also Sect. 6).

8 Dealing with Different Types of Missing and Censored Cost Data

Missing data can be a source of bias and uncertainty, and is a common occurrence with patient-level data. Within trials, this can occur due to factors such as poor reporting rates, incomplete reporting of data, and patients being lost to follow-up (e.g. study dropout; inability to complete a study). Even when using routine datasets, missing data can be problematic for reasons such as ‘censoring’ (i.e. the data are not collected or are not used for analysis). Issues with missing data may be a study design flaw that requires better/further/repeat data collection (see also Sect. 5); however, for the purpose of this section, the focus will be on how to deal with missing data using statistical methods. First, it is up to the analyst to determine the ‘types of missingness’. For example, are the data:

1. *Missing completely at random* (MCAR): no relationship between the data that are missing and for any values in the dataset, missing or observed.
2. *Missing at random* (MAR): no relationship between the missing data and mechanism for missingness, but this mechanism may be related to the observed data.
3. *Missing not at random* (MNAR): there is a relationship between the missing data and the mechanism for missingness.

Dealing with missing data has been explored and described in more detail by Little and Rubin [84]. Note that if the data are MCAR, then little needs to be done statistically at this point as MCAR does not imply systematic bias in your data; however, MCAR may be important if it reduces the statistical power of the analysis. Complete case analysis (CCA; i.e. using only those patients with no

missing data) is valid if complete cases are representative of the whole sample; if CCA is used, then an appropriate rationalisation/assumption should be provided as to why the complete cases are representative of the whole sample in the context of the intended analysis [85]. Like CCA, there are a number of other commonly used methods that have been applied to allow for missing data, a summary of which have been described by Briggs et al. [86] and include available case analysis, mean imputation, regression imputation, last value carried forward, and hot decking, each of which have their issues, e.g. mean imputation and regression imputation tend to not capture variability and can underestimate standard errors. However, despite some limitations, when data are determined to be MAR or MNAR, imputation is still commonly used to predict missing data. There are multiple univariate or multivariate methods for imputation, which can include using linear and logistic regressions of various types [87], predictive mean matching [88], and multiple imputation (MI) using chained equations (MICE) [89]. MICE has become particularly popular over the last decade and guidance on using this method has been described by White et al. [89]. Although MI is generally regarded as being more efficient, inverse probability weighting (IPW) may sometimes be preferred if incomplete cases provide little information [90]. The IPW approach corrects for the selection bias in the observed cases by weighting the complete cases with the inverse of their selection probabilities (i.e. IPW models the selection probability, such that it is a model for the probability that an individual is a complete case), whereas MI models the data distribution (i.e. a model for the distribution of the missing values given the observed data), which can be biased by the data distribution of the observed data, and which is problematic if the observed data suggest little about the unobserved data. As such, there has been a recommendation to combine the two approaches to improve the robustness of estimation in the literature, the practicalities around which have been discussed elsewhere [90, 91]. It is also recommended that sensitivity analyses to assess the robustness of the study’s results to potential MNAR mechanisms should be conducted [85]. Heckman models have also been suggested to extend the validity of MI to some MNAR mechanisms, although as the current authors are not aware of the extent to which Heckman models have been applied specifically for CEA, references are provided here mainly for the interested reader [92, 93]. It has been noted that for HTAs, the CCA or MI methods appear to be the most commonly used for CEA, relative to IPW and including sensitivity analyses [85], the suggestion being a need for improvement when dealing with missing data for CEA as, in many cases, CCA is not appropriately rationalised but is still used without considering and rationalising the most appropriate methods for dealing with, or assessing the effect of, missing data.

Censored data can be considered a special case of missing data. A key assumption when applying conventional survival analysis methods to censored cost data is that the uncensored data are the same as the censored data (i.e. the censored data are ‘uninformative’). However, future resource use and associated costs can be dependent on past resource use and associated costs, and therefore the censored data may be informative. Thus, the assumption that censored data may be uninformative is not valid and therefore censored data can lead to bias within CEA. Methods exist allowing for censoring that overcomes this assumption, using a weighting of costs with or without covariate adjustments. Adjusting for baseline cost differences may result in being able to deal with censoring if it is constant across the baseline and follow-up time periods of interest, but, again, this is an assumption (see also Sect. 7). Dealing with censored data has been discussed by Young [94], Drummond and McGuire [95], Willan et al. [96], and Wijesundera et al. [97]. The interested reader may also wish to explore the related literature regarding the handling of censoring in survival analysis [98].

9 Uncertainty Around Cost Estimates

All economic evaluations are subject to uncertainty, whereby we consider uncertainty to refer to the fact that the expected costs and effects of a treatment are not known with certainty, as opposed to variability (i.e. random differences in costs and effects between identical individuals) and heterogeneity (i.e. variability in costs and effects between individuals that can be attributed to their characteristics) [99, 100]. The importance for economic evaluations of properly characterising uncertainty, estimating its consequences, and appropriate policy actions to address this uncertainty (i.e. research and adoption decisions) have been widely discussed in the literature [101–105]. In this section, we briefly discuss the methods for characterising this uncertainty with regard to costs.

Uncertainty in economic evaluations is normally dealt with in two ways: deterministic sensitivity analysis, whereby single or groups of parameters are varied to assess the impact on results, or probabilistic sensitivity analysis (PSA), which assesses the joint uncertainty (based on the uncertainty across all inputs) [102]. Only the latter of these allows for the full characterisation of uncertainty.

When conducting deterministic sensitivity analysis around costs, unit costs and resource-use estimates may be varied to examine the impact on results, with the ranges chosen often being arbitrary. For PSA, distributions must be assigned to indicate the uncertainty. If the parameter relates to a unit cost that is considered fixed, it may not be appropriate to consider any uncertainty. In contrast, costs are

likely to be subject to uncertainty and distributions should be assigned.

As discussed in Sect. 3, there are two main vehicles for economic evaluation—trial-based and model-based. The approaches to characterising uncertainty in costs using PSA differs between these two vehicles, although there are some common themes.

For trial-based evaluations, typically the total or incremental costs between arms are estimated using econometric analysis. The uncertainty in this estimate can be estimated non-parametrically using techniques such as bootstrapping [106], or parametrically by fitting a distribution to the model (e.g. multivariate normality) [8] to see how the joint uncertainty in explanatory variables in the econometric model impact the outcome. Both types of approach allow the characterisation of uncertainty around the cost estimate (e.g. via confidence intervals) and its subsequent impact on decision uncertainty, i.e. the probability that a given intervention is cost effective, often diagrammatically presented as part of a cost-effectiveness acceptability curve (CEAC) [20].

For model-based evaluations, probabilistic distributions are placed on uncertain parameters and then the uncertainty through the whole model is propagated using Monte Carlo simulation to estimate the impact on costs and effects. While the uncertainty in all parameters may drive the uncertainty in total costs, we focus on appropriate distributions for cost variables (which could include resource use and associated unit costs). As has been previously discussed, costs (and resource use) are bound at zero and are typically right skewed. As such, probabilistic distributions such as the Gamma distribution are appropriate for characterising the uncertainty in mean cost parameters [8].

A further underconsidered issue in uncertainty around costs is how costs may change over time. Claxton et al. [103] have shown how cost changes over time can alter the value of interventions and change the appropriate policy decision.

10 A Note on Discounting, Inflation, and Using Relevant Currency

For CEA, both costs and effects need to be adjusted for the different times that they occur, although discounting is generally only applied when the time horizon of interest is > 12 months. Discounting is a mathematical procedure to adjust future costs to their present value, effectively reducing them by a specified proportion based on the discount rate and how far they are in the future.

Discounting has been rationalised to account for factors such as time preferences in the receipt of costs or effects, such that we value future events lower than current events, as well as the rate of return on capital and change in productivity over time [107]. NICE uses a rate of 3.5% for both

costs and benefits, although 1.5% can be used for sensitivity analyses [2, 108]. Discounting can be performed using group means or individual-level data, although the latter is preferred, particularly due to issues associated with missing data (see also Sect. 8). Recently, Attema et al. [109] have reviewed the issues and debates around discounting, describing and discussing the current discounting recommendations of countries publishing their national guidelines. Other useful references include the work of Claxton et al. [107] and Tinchög [110].

In some cases, there may be the need to inflate different unit costs so that all unit costs reflect a common cost year, which is typically the most recent year at the time of analysis and/or the most recent year for available unit costs (where inflation refers to mathematically scaling up a cost to reflect a general increase in prices). There are various inflation indices that can be used depending on the type of resource associated with the unit cost. For example, the PSSRU Unit Costs of Health and Social Care [57] has a list of inflation indices for the UK that includes the Hospital and Community Health Services (HCHS) index, Personal Social Services (PSS) pay and prices index, Gross Domestic Product (GDP) deflator and the tender price index for public sector buildings, and the Building Cost Information Service (BCIS) house rebuilding cost index and the retail price index.

It is common practice to represent cost in the currency of interest to the decision maker. For example, if the remit of the study is England and the evidence is being generated for NICE, then Great British pounds sterling (GBP; £) would be the common currency. However, in the case of multinational trials and/or if there are multiple decision makers from different countries for whom the evidence is being generated, there may be the need to apply an exchange rate (i.e. a rate representing the value of one currency relative to another) to costs (be it unit costs or overall costs). For multinational trials, how costs are dealt with is dependent on if a pooled (i.e. combine all cost and effects across countries), split (i.e. produce cost and effect results only in their country-specific setting) or partial/combined (e.g. pool only data from countries with comparable healthcare systems or using a common currency such as the Euro (€), with all other country results analysed for their own setting only) approach is taken. There is no clear consensus on how best to conduct economic evaluations in the context of multinational trials. The methods used for conducting an economic evaluation in the context of a multinational trial have been considered by three different literature reviews by Reinhold et al. [111] and Opong et al. [112], with Vemer and Rutten-van Mölken [113] focussing specifically on statistical methods used to enable transferability between country settings. Manca et al. [114] have critically appraised the methodologies used for CEA as part of multinational trials, and a general discussion about the transportability of comparative cost effectiveness between

countries has been discussed by Briggs [115], with ‘good practice’ guidance suggested by Drummond et al. [116].

11 Recommendations for Future Analyses and Research

Although checklists exist to use alongside CEA (e.g. CHEERS), there is still poor reporting and rationalisation of statistical methods and covariate adjustments when using cost data. This aspect does not necessarily need to be part of a new checklist as it can be amalgamated into existing checklists to avoid overburdening researchers with more checklists. What is more important is that researchers give careful consideration as to *why* they have chosen a specific regression model and associated covariates rather than using whatever is observed in a similar study in the empirical literature, e.g. if another study has used OLS adjusting for baseline costs, this does not necessarily mean doing exactly the same is appropriate in a similar study. Researchers should critically assess what covariates to include (e.g. age, gender), and it is up to researchers to suggest and appropriately rationalise what covariates are or are not important, as hypothesised a priori. Such hypotheses should be based on clinical and empirical knowledge, reflecting on (but not simply repeating) existing studies and analyses in the empirical literature. It is also important to consider what adjustments can be conducted post hoc, which could form exploratory analysis (as opposed to per-protocol) to inform developing hypotheses within future studies.

From the current authors’ perspectives, there is a lack of papers where baseline adjustments have been made or appropriately rationalised for the purpose of CEA. Although there is no up-to-date systematic review to support this claim related to costs, a review by Richardson and Manca [117] published in 2004 noted that most published trial-based CUA studies failed to recognise the need to adjust for imbalance in baseline utility. The current authors feel a lack of more generalised (i.e. not just adjusting for baseline utility) and rationalised baseline adjustments for CEA is still apparent; however, we recognise the need to conduct a systematic review to fully support this claim.

The external validity of RCTs has been raised in the literature with regard to, for example, patient selection and use of exclusion criteria. One particular issue with trials for informing costs is the presence of protocol-specific resource use that would not be expected in usual practice (e.g. additional physician appointments for data collection). When such costs are present, it may be sensible to include additional analyses in instances where they are excluded. This is an addition to other methods that can be

used to adjust trial analyses to make them more representative of the target population in real-world use [118].

With the increasing use of observational data in CEA, future research should consider the use of statistical and econometric methods that can control for unobserved confounding. This is the case not only for the treatment effect on costs but also the estimation of event or health state costs for the purposes of modelling. When analysing trials for the purposes of informing decision model parameters such as these, methods to control for unobserved confounding may still be useful since randomisation can be broken at the point of a clinical event or reaching a specific health state.

In this paper, we have taken it as given that unit costs represent the true monetary value of a resource. However, unit costs may not represent the true value of a resource, with resources potentially being more constrained than others, and unit costs not adjusting immediately such that marginal productivity is equalised across healthcare resources. This raises issues for the aggregation of costs across different types of resource use to estimate a total healthcare cost [119, 120]. Consideration may be given to the estimation of cost categories for more disaggregated resource use in such situations. In addition, while the focus of this paper has been cost analysis, for CEA it is the joint distribution of costs and effects which determine cost effectiveness. More research needs to consider the appropriate bivariate modelling of such data, given the underlying statistical properties of the two types of data.

12 Conclusion

There has been a tendency to overlook the complicated nature of cost data and how to include it in CEA, the consequence of which is that inconsistent estimates could be produced across studies that do not appropriately control for the nature of cost data and the uncertainty around subsequent cost estimates. This could lead to inaccurate information being given to decision makers. This paper has outlined important considerations when using cost data for CEA, while noting that further research needs to be conducted in this area. Researchers need to carefully consider how to account for costs within their specific economic evaluation.

Acknowledgements The authors would like to thank all members of the Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber (CLAHRC YH) Health Economics and Outcome Measurement (HEOM) teams at the University of York and University of Sheffield as the writing team and idea for this manuscript was part-conceived during a CLAHRC YH HEOM quarterly team meeting. We also thank Beth Woods (University of York) for her expert input into developing the purview of this paper during the development stage.

Author Contributions All authors contributed to the idea about the content of the manuscript and have provided written contributions to the paper, including edits to draft versions. MF and JL conceived the original idea for the paper before expanding the writing team. MF led the writing of the overall manuscript, including final editing and formatting, and led the writing of the following manuscript sections: Sect. 2 (the difference between resource use and costs); Sect. 5 (sources of resource-use data and unit costs); Sect. 7 (adjusting for baseline covariates); Sect. 8 (dealing with different types of missing cost data); and Sect. 10 (a note on discounting, inflation, and using relevant currency). JL led the writing of Sect. 6 (statistical methods for assessing cost data and its distribution). SW led the writing of the following manuscript sections: Sect. 3 (vehicles for CEA); Sect. 4 (what costs to include depending on costing perspective); and Sect. 9 (uncertainty around cost estimates). SW, JL, and MF developed and wrote the introduction, discussion and conclusion sections. TY provided expert oversight and contributed throughout the manuscript. All authors act as guarantors for the content of the manuscript.

Funding The writing of this manuscript was part-funded by the National Institute for Health Research (NIHR) CLAHRC YH (<http://www.clahrc-yh.nir.ac.uk>). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social care. The funding agreement ensured the authors' independence in developing the purview of the manuscript, and writing and publishing the manuscript.

Compliance with ethical standards

Conflict of interest Matthew Franklin, James Lomas, Simon Walker and Tracey Young have no conflicts of interest to report.

OpenAccess This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. 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.
2. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: NICE; 2013.
3. 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.
4. 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.
5. Lloyd A. Special edition on utility measurement. *PharmacoEconomics.* 2017;35(Suppl 1):5–6.
6. Akobeng A. Understanding randomised controlled trials. *Arch Dis Child.* 2005;90(8):840–4.

7. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials. Oxford: Oxford University Press; 2014.
8. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. New York: Oxford University Press; 2006.
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. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;26(9):799–806.
11. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.
12. 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.
13. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess*. 2004;8(36):1–158.
14. Ghabri S, Stevenson M, Möller J, Caro JJ. Trusting the results of model-based economic analyses: is there a pragmatic validation solution? *Pharmacoeconomics*. 2019;37(1):1–6.
15. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Med Decis Mak*. 2012;32(5):667–77.
16. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. *Med Decis Mak*. 2012;32(5):690–700.
17. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Med Decis Mak*. 2012;32(5):701–11.
18. Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5. *Value Health*. 2012;15(6):828–34.
19. Dolgin M, Fox AC. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. Brown Boston: Little; 1994.
20. Briggs A, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess*. 1999;3(2):1–134.
21. Walker S, Asaria M, Manca A, Palmer S, Gale CP, Shah AD, et al. Long-term healthcare use and costs in patients with stable coronary artery disease: a population-based cohort using linked health records (CALIBER). *Eur Heart J Qual Care Clin Outcomes*. 2016;2(2):125–40.
22. Last JM, Abramson JH, Freidman GD. A dictionary of epidemiology. New York: Oxford University Press; 2001.
23. 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.
24. Morton A, Adler AI, Bell D, Briggs A, Brouwer W, Claxton K, et al. Unrelated future costs and unrelated future benefits: reflections on NICE guide to the methods of technology appraisal. *Health Econ*. 2016;25(8):933–8.
25. de Vries LM, van Baal PH, Brouwer WB. Future costs in cost-effectiveness analyses: past, present, future. *Pharmacoeconomics*. 2018. <https://doi.org/10.1007/s40273-018-0749-8>.
26. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA*. 2016;316(10):1093–103.
27. Claxton K, Walker S, Palmer S, Sculpher M. CHE Research Paper 54: appropriate perspectives for health care decisions. York: University of York; 2010.
28. Remme M, Martinez-Alvarez M, Vassall A. Cost-effectiveness thresholds in global health: taking a multisectoral perspective. *Value Health*. 2017;20(4):699–704.
29. Culyer AJ. CHE Research Paper 154: cost, context and decisions in health economics and cost-effectiveness analysis. York: University of York; 2018.
30. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment program. *Value Health*. 2010;13(8):867–72.
31. Beecham J, Knapp M. Costing psychiatric interventions. In: Thornicroft G, editor. *Measuring mental health needs*. London: Gaskell; 2001. p. 200–24.
32. Thorn JC, Coast J, Cohen D, Hollingworth W, Knapp M, Noble SM, et al. Resource-use measurement based on patient recall: issues and challenges for economic evaluation. *Appl Health Econ Health Policy*. 2013;11(3):155–61.
33. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics*. 2017;35(1):21–31.
34. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Medical Care*. 1989;27(Suppl 3):S217–32.
35. Longworth L, Yang Y, Young T, Mulhern B, Hernandez Alava M, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014;18(9):1–224.
36. Rowen D, Brazier J, Ara R, Zouraq IA. The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics*. 2017;35(1):33–41.
37. Marti J, Hall PS, Hamilton P, Hulme CT, Jones H, Velikova G, et al. The economic burden of cancer in the UK: a study of survivors treated with curative intent. *Psycho-Oncology*. 2016;25(1):77–83.
38. Thompson S, Wordsworth S. An annotated cost questionnaire for completion by patients. Aberdeen: Health Economics Research Unit, University of Aberdeen; 2001.
39. Wimo A, Gustavsson A, Jönsson L, Winblad B, Hsu M-A, Gannon B. Application of Resource Utilization in Dementia (RUD) instrument in a global setting. *Alzheimer's Dementia*. 2013;9(4):429–435.e17.
40. Thorn JC, Brookes ST, Ridyard C, Riley R, Hughes DA, Wordsworth S, et al. Core Items for a Standardized Resource Use Measure (ISRUM): Expert Delphi Consensus Survey. *Value Health*. 2017;21(6):640–9.
41. Ridyard CH, Hughes DA, Team D. Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. *Value Health*. 2012;15(5):650–5.
42. 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.
43. Tanajewski L, Franklin M, Gkoutouras G, Berdunov V, Edmans J, Conroy S, et al. Cost-effectiveness of a specialist geriatric medical intervention for frail older people discharged from acute medical units: economic evaluation in a

- two-centre randomised controlled trial (AMIGOS). *PLoS One*. 2015;10(5):e0121340.
44. Tanajewski L, Franklin M, Gkountouras G, Berdunov V, Harwood RH, Goldberg SE, et al. Economic evaluation of a general hospital unit for older people with delirium and dementia (TEAM Randomised Controlled Trial). *PLoS One*. 2015;10(12):e0140662.
 45. NHS Digital. Hospital Episode Statistics (HES); 2018. <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics>. Accessed 17 Oct 2018.
 46. McDonnell L, Delaney B, Sullivan F. Datasets that may be of interest to Primary Care Researchers in the UK; 2017. <http://www.farrinstitute.org/wp-content/uploads/2017/10/Datasets-that-may-be-of-interest-to-Primary-Care-Researchers-in-the-UK-May-2016.pdf>. Accessed 17 Oct 2018.
 47. Baker R, Tata LJ, Kendrick D, Orton E. Identification of incident poisoning, fracture and burn events using linked primary care, secondary care and mortality data from England: implications for research and surveillance. *Inj Prev*. 2015;22(1):59–67.
 48. 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.
 49. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ*. 2013;346:f2350.
 50. Asaria M, Grasic K, Walker S. Using linked electronic health records to estimate healthcare costs: key challenges and opportunities. *PharmacoEconomics*. 2016;34(2):155–60.
 51. Spencer SA, Davies MP. Hospital episode statistics: improving the quality and value of hospital data: a national internet e-survey of hospital consultants. *BMJ Open*. 2012;2(6):e001651.
 52. Thorn JC, Turner E, Hounsborne L, Walsh E, Donovan JL, Verne J, et al. Validation of the hospital episode statistics outpatient dataset in England. *PharmacoEconomics*. 2016;34(2):161–8.
 53. Byford S, Leese M, Knapp M, Seivewright H, Cameron S, Jones V, et al. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. *Health Econ*. 2007;16(5):531–6.
 54. Noben CY, de Rijk A, Nijhuis F, Kottner J, Evers S. The exchangeability of self-reports and administrative health care resource use measurements: assessment of the methodological reporting quality. *J Clin Epidemiol*. 2016;74(93–106):e2.
 55. 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(8):1–13.
 56. NHS Improvement. Reference costs; 2017. <https://improvement.nhs.uk/resources/reference-costs/>. Accessed 17 Oct 2018.
 57. PSSRU. Unit Costs of Health and Social Care; 2018. <https://www.pssru.ac.uk/project-pages/unit-costs/>. Accessed 17 Oct 2018.
 58. BNF. BNF Publications; 2018. <https://www.bnf.org/>. Accessed 17 Oct 2018.
 59. Department of Health and Social Care. Drugs and pharmaceutical electronic market information tool (eMIT); 2018. <https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit>. Accessed 17 Oct 2018.
 60. Geue C, Lewsey J, Lorgelly P, Govan L, Hart C, Briggs A. Spoilt for choice: implications of using alternative methods of costing hospital episode statistics. *Health Econ*. 2012;21(10):1201–16.
 61. Jones AM. Models for health care. In: Clements M, Hendry D, editors. *Handbook of economic forecasting*. Oxford: Oxford University Press; 2011.
 62. Rouaud M. Probability, statistics and estimation: propagation of uncertainties in experimental measurement; 2017. <http://www.incertitudes.fr/book.pdf>. Accessed 15 Jan 2018.
 63. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ*. 2004;13(5):461–75.
 64. Mihaylova B, Briggs A, Ohagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 2011;20(8):897–916.
 65. Jones AM, Lomas J, Moore PT, Rice N. A quasi-Monte-Carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: an application to healthcare costs. *J R Stat Soc Ser A (Stat Soc)*. 2016;179(4):951–74.
 66. Jones AM. *Health econometrics*. Handbook of health economics. Amsterdam: Elsevier; 2000. p. 265–344.
 67. Bradbury PA, Tu D, Seymour L, Isogai PK, Zhu L, Ng R, et al. Economic analysis: randomized placebo-controlled clinical trial of erlotinib in advanced non-small cell lung cancer. *J Natl Cancer Inst*. 2010;102(5):298–306.
 68. Cromwell I, van der Hoek K, Taylor SCM, Melosky B, Peacock S. Erlotinib or best supportive care for third-line treatment of advanced non-small-cell lung cancer: a real-world cost-effectiveness analysis. *Lung Cancer*. 2012;76(3):472–7.
 69. Deb P, Manning WG, Norton EC. Modeling health care costs and counts. *Annu Rev Public Health*. 2014;39:489–505.
 70. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. *J Health Serv Res Policy*. 2004;9(4):197–204.
 71. Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. *J Health Econ*. 1999;18(2):153–71.
 72. Deb P, Norton EC, Manning WG. *Health econometrics using Stata*. College Station: Stata Press; 2017.
 73. Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*. 2005;6(1):93–109.
 74. Holly A, Monfort A, Rockinger M. Fourth order pseudo maximum likelihood methods. *J Econom*. 2011;162(2):278–93.
 75. Jones AM, Lomas J, Rice N. Healthcare cost regressions: going beyond the mean to estimate the full distribution. *Health Econ*. 2015;24(9):1192–212.
 76. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ*. 2001;20(4):461–94.
 77. Gomes M, Grieve R, Nixon R, Edmunds W. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. *Med Decis Mak*. 2012;32(1):209–20.
 78. Faria R, Hernandez Alava M, Manca A, Wailoo A. NICE DSU Technical Support Document 17: the use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. Sheffield: University of Sheffield; 2016.
 79. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ*. 2008;27(3):531–43.
 80. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med*. 1999;18(15):1903–42.
 81. Committee for Medicinal Products for Human Use. Guideline on adjustment for baseline covariates. London: European Medicines Agency; 2015.
 82. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487–96.

83. van Asselt AD, van Mastrigt GA, Dirksen CD, Arntz A, Severens JL, Kessels AG. How to deal with cost differences at baseline. *Pharmacoeconomics*. 2009;27(6):519–28.
84. Little RJ, Rubin DB. *Statistical analysis with missing data*. New York: Wiley; 2014.
85. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: an incomplete journey. *Health Econ*. 2018;27(6):1024–40.
86. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing.... presumed at random: cost-analysis of incomplete data. *Health Econ*. 2003;12(5):377–92.
87. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol*. 2001;27(1):85–96.
88. Landerman LR, Land KC, Pieper CF. An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociol Methods Res*. 1997;26(1):3–33.
89. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377–99.
90. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278–95.
91. Han P. Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scand J Stat*. 2016;43(1):246–60.
92. Galimard JE, Chevret S, Protopopescu C, Resche-Rigon M. A multiple imputation approach for MNAR mechanisms compatible with Heckman's model. *Stat Med*. 2016;35(17):2907–20.
93. Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann Econ Soc Meas*. 1976;5(4):475–92.
94. Young TA. Estimating mean total costs in the presence of censoring. *Pharmacoeconomics*. 2005;23(12):1229–42.
95. Drummond MF, McGuire A. *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press; 2001.
96. Willan AR, Lin D, Manca A. Regression methods for cost-effectiveness analysis with censored data. *Stat Med*. 2005;24(1):131–45.
97. Wijesundera 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.
98. Latimer N. NICE DSU Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. Sheffield: University of Sheffield; 2011.
99. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Mak*. 2012;32(5):722–32.
100. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Chapter 11: Characterizing, reporting, and interpreting uncertainty. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford: Oxford University Press; 2015. p. 389–426.
101. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ*. 1994;3(2):95–104.
102. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999;18(3):341–64.
103. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, et al. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. *Health Technol Assess*. 2012;16(46):1–323.
104. Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: the consequences of ignoring decision uncertainty. *Health Econ*. 2011;20(2):212–24.
105. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value Health*. 2012;15(3):570–9.
106. Efron B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*. 1981;68(3):589–99.
107. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health-care technologies. *Health Econ*. 2011;20(1):2–15.
108. NICE. *Developing NICE guidelines: the manual*; 2014. <https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview>. Accessed 17 Oct 2018.
109. Attema AE, Brouwer WB, Claxton K. Discounting in economic evaluations. *Pharmacoeconomics*. 2018;36(7):745–58.
110. Tinghög G. Discounting, preferences, and paternalism in cost-effectiveness analysis. *Health Care Anal*. 2012;20(3):297–318.
111. Reinhold T, Brüggemann B, Schlander M, Rosenfeld S, Hessel F, Willich SN. Economic analysis based on multinational studies: methods for adapting findings to national contexts. *J Public Health*. 2010;18(4):327–35.
112. Oppong R, Jowett S, Roberts TE. Economic evaluation alongside multinational studies: a systematic review of empirical studies. *PLoS One*. 2015;10(6):e0131949.
113. Vemer P, Rutten-van Mölken MP. The road not taken: transferability issues in multinational trials. *Pharmacoeconomics*. 2013;31(10):863–76.
114. Manca A, Sculpher MJ, Goeree R. The analysis of multinational cost-effectiveness data for reimbursement decisions. *Pharmacoeconomics*. 2010;28(12):1079–96.
115. Briggs A. Transportability of comparative effectiveness and cost-effectiveness between countries. *Value Health*. 2010;13(Suppl 1):S22–5.
116. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value Health*. 2009;12(4):409–18.
117. Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Econ*. 2004;13(12):1203–10.
118. Hartman E, Grieve R, Ramsahai R, Sekhon JS. From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. *J R Stat Soc Ser A (Stat Soc)*. 2015;178(3):757–78.
119. Revill P, Walker S, Cambiano V, Phillips A, Sculpher MJ. Reflecting the real value of health care resources in modelling and cost-effectiveness studies: the example of viral load informed differentiated care. *PLoS One*. 2018;13(1):e0190283.
120. Van Baal P, Morton A, Severens JL. Health care input constraints and cost effectiveness analysis decision rules. *Soc Sci Med*. 2018;200:59–64.