

Title:

Conducting Economic Evaluations Alongside Randomised Trials: Current Methodological Issues and Novel Approaches

Authors:

Dyfrig Hughes*¹, Joanna Charles¹, Dalia Dawoud², Rhiannon Tudor Edwards¹, Emily Holmes¹, Carys Jones¹, Paul Parham³, Catrin Plumpton¹, Colin Ridyard¹, Huw Lloyd-Williams¹, Eifiona Wood¹, Seow Tien Yeo¹

¹Centre for Health Economics and Medicines Evaluation, Bangor University, Wales, UK, LL57 2PZ.

²National Clinical Guidelines Centre, 11 St Andrews Place, Regent's Park, London, UK, NW1 4LE.

³Department of Public Health and Policy, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK, L69 3GL.

***Corresponding Author:**

Centre for Health Economics and Medicines Evaluation, Arduwy, Bangor University, Holyhead Road, Wales, UK, LL57 2PZ.

Tel: +44 (0)1248 382950 Email: d.a.hughes@bangor.ac.uk

Running title:

Trial-based economic evaluations

Compliance with Ethical Standards:

The preparation of this review was funded by the Medical Research Council North-West Hub in Trials Methodology Research, reference number G0800792.

Dyfrig Hughes, Joanna Charles, Dalia Dawoud, Rhiannon Tudor Edwards, Emily Holmes, Carys Jones, Paul Parham, Catrin Plumpton, Colin Ridyard, Huw Lloyd-Williams, Eifiona Wood and Seow Tien Yeo declare they have no conflict of interest.

Author contributions:

Dyfrig Hughes conceived this review and, with Eifiona Wood, revised the manuscript critically for important intellectual content. Dyfrig Hughes, Joanna Charles, Dalia Dawoud, Rhiannon Tudor Edwards, Emily Holmes, Carys Jones, Paul Parham, Catrin Plumpton, Colin Ridyard, Huw Lloyd-Williams, Eifiona Wood and Seow Tien Yeo made substantial contributions to the design and drafting of the work, approved the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Dyfrig Hughes will act as the guarantor of the work presented in this paper.

Abstract:

Trial-based economic evaluations are an important aspect of health technology assessment. The availability of patient-level data coupled with unbiased estimates of clinical outcomes means that randomised controlled trials are effective vehicles for the generation of economic data. However there are methodological challenges to trial-based evaluations, which include the collection of reliable data on resource use and cost, choice of health outcome measure, calculating minimally important differences, dealing with missing data, extrapolating outcomes and costs over time and the analysis of multinational trials. This review focuses on the state of the art of selective elements concerning the design, conduct, analysis and reporting of trial-based economic evaluations. The limitations of existing approaches are detailed and novel methods introduced. The review is internationally relevant but with a focus towards practice in the UK.

Key Messages:

- Trial-based economic evaluations require careful consideration of design, conduct, analysis and reporting.
- Areas for improvement include the need for more robust methods to estimate resource utilisation, use of multiple imputation for missing data, and application of appropriate modelling for the assessment of the impact of non-adherence.
- This review identifies economic methods that may also enhance clinical trials such as for calculating minimally important differences using discrete choice experiments, and undertaking data extrapolation based on modelling.

1. Introduction

Randomised controlled trials (RCTs) are an important foundation for economic evaluation. RCTs generate patient-level data that provide an unbiased estimate of the efficacy or effectiveness of an intervention being tested, as well as an opportunity for collecting resource use and cost data necessary for calculating an intervention's cost-effectiveness. RCTs, however, are not a panacea and rarely do they provide the full information necessary for an economic evaluation or to inform a decision [1]. Approaches to trial-based economic evaluation have consequently evolved, as detailed in recent guidelines and reviews [2-4]. This article focuses on specific aspects of trial-based evaluations where there is scope for improvement and introduces novel methodologies that may complement or supplant current approaches. The topics discussed can be broadly categorised into the methods of interpreting trial-based data, and techniques for using trial-based data in model-based evaluation. The former category considers methods for resource use estimation, choice of outcome measure, handling missing data, accounting for patient non-adherence and analysis of multi-national trial data. Factors relating to the application of trial-based data in modelling include consideration and discussion of issues on calculating minimally-important differences and methods for extrapolation. Based on a purposive review of the literature, the concepts discussed are internationally relevant but with a focus towards practice in the UK.

2. Trial-based economic evaluations

2.1. When to conduct economic evaluation?

Several factors signpost the need to undertake an economic evaluation, including:

Anticipated difference in therapeutic benefit between interventions. New health technologies are less likely to be cost-effective (and thus more reason to conduct an economic analysis) if any difference in health outcome is anticipated to be small when compared with existing interventions.

Large cost differences. An economic analysis is warranted if the new health technology is potentially very costly and /or the two (or more) interventions being compared are of greatly different cost; or, if there are considerable down-stream costs that occur or continue beyond the trial time horizon, that may differ between interventions such as increased intensity of care or frequency of testing.

High levels of uncertainty around the cost-effectiveness, or economic impact of health technology; the greater the uncertainty, the stronger the case for economic evaluation.

Sizeable budget impact as a result of a potentially large population or population subgroup. Health economic analyses may be less appropriate if the decision rule is based on budget impact and the overall budgetary impact is very small, as the opportunity cost associated with an incorrect decision may be limited.

2.2. Does the economic evaluation need to be based on RCT evidence?

Where an economic evaluation is indicated, it is imperative that due consideration is given as to whether an RCT is the most suitable platform for evidence acquisition. Decisions may have to be made before results are available from clinical trials; or put another way, the value of generating economic evidence from a clinical trial might be less than the costs of postponing a decision. In such circumstances, an economic model may be preferred. Trials which include comparators that differ from routine care, include selective sub-populations of patients, or which are not pragmatic in design may not lead to reliable estimates of the comparative effectiveness of treatments, and may therefore bias the cost-effectiveness result.

Organisations such as the UK National Institute for Health Research Health Technology Assessment (NIHR-HTA) programme, aim to inform clinical decisions by funding pragmatic studies that address issues of clinical and cost effectiveness. A review of 78 randomised studies funded by the NIHR-HTA programme published to June 2009, identified 75 (96%) of studies to include an economic analysis [5]. However not all trials are amenable to economic analysis, and not all interventions require trial-based economic evidence. It may be the case, for instance, that the intervention is low cost but non-inferior to current practice. While this might only be known *ex post*, there are specific instances where RCTs are less relevant for economic evaluations. A trial-based economic evaluation would be unnecessary in the case of a comparison of two similar interventions that are delivered in different settings, whereby incremental benefits might mean less use of expensive resources. The cost of either saline or albumin for resuscitation in patients with severe sepsis, for instance, is dwarfed by the potential savings from a reduction in expensive intensive care should one prove to be more effective. Cost would not be a factor to influence the decision to change clinical practice. Economic evaluation can also, in the appropriate situation, take data from non-RCT sources. For example, a cost-minimisation analysis comparing endometrial ablation performed as a day case under general anaesthesia or as an outpatient using local anaesthetic used hospital data that were routinely collected [6]. Similarly, a cost-minimisation analysis, for instance, would be appropriate without the need to source economic evidence directly from trials of generic medicines that are bioequivalent to their branded counterparts.

2.3. Modelling of RCT data, or a trial-based evaluation?

It is generally accepted that some form of economic evaluation is necessary to support funding decisions on new healthcare interventions. However, it is important to note that only a small proportion of healthcare interventions and services are prioritised for appraisal by HTA organisations. The National Institute for Health and Care Excellence (NICE) in the UK expect an economic model to form part of the assessment, but whether this is a trial-based economic evaluation is not a high priority [7].

A model can flexibly evaluate different patient populations, sub-groups, interventions, outcomes, resources and cost effects. Such an approach is almost always justified in order to synthesise the evidence base and to take into account the uncertainty around the cost-effectiveness point estimate. However the choice of whether to use RCT data, as one of several information sources to populate an economic model, or to undertake a trial-based economic evaluation and subsequent model whereby the majority of the model inputs are obtained from the trial, is dependent upon a number of factors.

Situations that would point towards a trial-based economic evaluation being more appropriate might include those where costs are largely driven by, or estimated from, the primary outcomes of the trial (such as for cardiovascular outcomes or COPD exacerbations). Other factors that point to a preference for trial-based evaluations include situations where considerable heterogeneity of costs is expected, thereby justifying a more individual patient-specific data collection method, or where adherence, only apparent in a more pragmatic trial, may drive costs and outcomes.

2.4. Is the trial in question a suitable vehicle for an economic evaluation?

The decision as to whether a given trial is a suitable vehicle for an economic evaluation is based upon *good trial design* that is capable of providing unbiased answers to the clinical question; whether *current practice* (which may be best supportive care) is one of the alternatives being compared and that *the trial setting* allows for generalizable results. Factors to be considered include the trial type, relevance of the outcome measures and populations, expected effect size, external validity as well as logistical considerations for data collection [8-11].

With regards to trial type, explanatory trials seek to identify whether an intervention works under ideal conditions (a measure of efficacy) whilst pragmatic trials consider whether an intervention works in real-life conditions (a measure of effectiveness). Ideally, trials must be at least partly pragmatic and therefore relate to actual clinical practice if they are to be suitable for economic analysis. However, in the case of new medicines, most RCTs are explanatory, since they are concerned with efficacy (as demanded by regulatory authorities), are typically conducted within well-resourced settings, recruit a select sample of patients (e.g. defined by their adherence, co-morbidity, age) and often measure a short-term surrogate outcome. These do not always contain all the elements necessary for economic evaluation, since they may be limited by factors such as lack of appropriate comparators, insufficient duration of follow-up particularly of lifelong treatments for chronic diseases, inadequate powering for economic outcomes, and poor external validity due to limited generalizability to other populations or care settings. These factors may be mitigated by the use of modelling, for instance to synthesise an indirect comparison against a relevant comparator or to extrapolate treatment effects and costs over a longer time horizon.

Pragmatic trials on the other hand have broad inclusion criteria for trial participants, measure patient-relevant outcome measures and are thus more suited to inform clinical decision making. They are intended to reflect usual practice and are better suited as vehicles for economic evaluation, but again modelling is invariably necessary in order to meet the evidential requirements of decision makers. However in the absence of such studies (as is normally the case with new medicines) the analyst is faced with the challenge of how might effectiveness, and hence cost-effectiveness, be predicted from data pertaining to a treatment's efficacy. Neither explanatory nor pragmatic trials are typically sufficient for data on adverse events, especially if these are rare.

2.5. Trials of public health interventions: A special case?

Many public health interventions are best described as complex [12]. They have several components, and effects depend on the context of delivery, synergistic impact of other socio-economic factors and general societal trends that are outside the remit of those conducting a trial. This makes it difficult to attribute cause and, by contrast to trials of clinical interventions, trials of public health interventions are conducted as pragmatic, community-based, and often cluster RCTs. Cluster RCTs require appropriate statistical methods of analysis, such as the use of multilevel modelling to control for the effect of clustering on outcomes and costs, and other considerations required for the sample size calculation, the univariate analysis of incremental costs and outcomes, the correlation between individual costs and effects, and the distribution of costs and outcomes [13].

There is a growing literature pertaining to solutions to methodological challenges of conducting economic evaluations of public health interventions [14-17]. Fischer et al [18] question the "fit" of traditional methods of health technology assessment and preoccupation with evidence-based medical decision making to a public health context, arguing that this will inhibit spending decisions on public health interventions. They state that many population-focussed public health interventions seek to make small changes in behaviour across large populations, and large RCTs with sufficient power to detect such changes are impractical or prohibitively expensive. Consequently, many trials are underpowered and unable to demonstrate whether public health interventions are effective. However, even if underpowered, such trials can generate point estimates that can serve as inputs to economic models.

Fischer et al [18] propose instead a decision theory approach where decision makers use their prior beliefs and experience in a structured Bayesian approach. They suggest that this approach is suitable for health improvement programmes such as screening and immunization, behaviour change "nudges" like food labelling, government policy on health harming products e.g. taxation and pricing, and social and economic policies such as those relating to the environment, housing and education that have important health impacts. Such community-level interventions are associated with large population-health effects that would not be readily captured in RCTs without the application of extremely large sample sizes. Applying a decision-theory approach involves the use of prior belief evidence combined with RCT evidence, even if underpowered, with the intent of generating a set of informed posterior beliefs about the effectiveness of a public health intervention to be weighed up against costs. This approach differs from the standard evidence-based medicine approach which relies heavily on RCT evidence.

Trials with embedded “process and economic evaluations” help researchers understand mechanisms of change and implications for cost-effectiveness across the socio-economic gradient in society. Hence helping to address a key tenet in public health - that of reducing inequalities in health.

3. Methods for resource use estimation

Costs are determined by the perspective adopted, and this varies according to the payer, healthcare system, and by country. A US study, for example, might take an insurers’ perspective and measure direct costs such as hospital stays whereas the preference for a societal perspective in the Netherlands might lead to an economic evaluation to estimate indirect costs as well as the broader public sector costs of social services, educational and/or criminal justice services etc. In developing countries, patients may be asked about coping costs and what assets they have to sell in order to fund their treatment. Regardless of which items are to be costed, reliable estimation of resource use depends on availability of valid methods of measurement. Methods employed in trial-based economic evaluations typically include a combination of one or more of the following: abstraction of data from routine medical records (e.g. patient notes, electronic medical records or patient administration systems), data collated mainly for the purpose of payment (e.g. hospital episode statistics, medicines dispensing data, insurance claims), dedicated sections within case report forms and administration of questionnaires to patients [5,19] (Figure 1). The choice of data collection method is particularly important to minimise missing values and achieve high level of accuracy.

Insert Figure 1 here

3.1. Based on patient recall

Our experience with DIRUM, the Database of Instruments for Resource Use Measurement, (<http://www.dirum.org>) suggests that patient questionnaires are used extensively in many countries. For example, the European version of the widely-used Client Service Receipt Inventory (CSRI) [20] has been piloted in Spain, Italy, Denmark and the Netherlands, and found to be an acceptable method for collating service receipt and associated data alongside assessment of outcomes in patients with schizophrenia [21]. In the UK, methods that are reliant on patient recall are the most common for measuring resource use among NIHR-HTA programme-funded trials [5]. This is partly due to ease of use compared with the logistic challenges associated with obtaining electronic resource use data from local or national databases, and partly because there are no other means of collecting certain data such as time off work (productivity costs) or non-medical costs (such as transport to/from place of care). Data reliant on patient recall are subject to bias which may adversely affect the resulting cost estimates. It is generally accepted that there is an inverse relationship between length of recall period and accuracy of data [22] owing to patients forgetting specific events and /or their timings. It is difficult to define an optimum recall period, as this is context-specific, depending on factors such as the nature of the resource use being asked (which might define how memorable the event was), the ability of the respondent to understand what is being asked of them, the nature of the illness or disease and the frequency of measurement and /or follow-up in relation to patients’ clinical requirements.

While Brown et al [23] suggested that patients could recall their clinical encounters effectively up to 3 months after the event, Evans et al [24] suggest recall is in fact related to the saliency or the impact of the event on the patient’s life. For example, a recall of 6 months might suffice for hospitalisations whereas a shorter time frame would be required for repeated visits to a general

practitioner (GP) or community pharmacy. Questionnaires that require recall periods of greater than 12 months are not recommended [25]. Richards et al [26] highlight the association between age and recall bias, advising caution with older patients and suggesting there are disparities, even within shorter timeframes of 3 months, on high saliency items such as hospitalisations. It is somewhat reassuring that the median recall time frames employed in a sample of 100 UK NIHR-HTA programme-funded trials was 3 months (IQR, 0.5 to 6 months) [27].

Where the likelihood of comprehension is low (e.g. in children) or cognitive impairment is high, researchers may have no alternative but to rely on proxy recall, especially if collecting data on a broad spectrum of resource usage to a given costing perspective. Levels of agreement between patient and proxy or healthcare record and proxy are not well established [24] although, in palliative care, at least, there is some evidence to suggest information that relies on observable aspects of resource use can have a good level of agreement between patient and proxy [28].

3.2. Based on routinely collected data

In the UK, health and social care is provided by multiple agencies that use different database systems, currently with limited record linkage. Routinely collected data such as from hospital patient notes and GP records are nonetheless used widely alongside patient and proxy-reported healthcare use in trial-based economic evaluations [5]. GP records have traditionally been very diffuse though systems such as EMIS-web (<https://www.emishealth.com>) and the Clinical Practice Research Datalink (<http://www.cprd.com>) provide a basis for more accessible data capture. One limitation in the use of GP records relates to the trial setting. When studies recruit patients from a hospital setting, for instance, sourcing data from individual GPs can be challenging as patients will be served by many. Patients recruited via community pharmacies can be geographically disperse, meaning tracking to general practice or hospital can be logistically very challenging. However, the availability of national hospital datasets can mitigate this problem.

Acute inpatient care data for hospitals in England are collated for submission to the Secondary Uses Service (SUS) primarily for the purposes of Payment by Results (PbR), planning, commissioning and management, but also for audit and research [29]. Similar data are collected within the Patient Episode Database for Wales (PEDW) and the Scottish Morbidity Database in Scotland. All use Healthcare Resource Groups (HRGs) as a currency; however, whereas bundled HRGs which constitute all the elements of care associated with a particular intervention, are nationally agreed and can be costed using the national tariff, unbundled HRGs are more reliant on the average unit cost to the National Health Service as these are commissioned, priced and paid for at a local level [29]. The flow of electronic data is presented in Figure 2, which also highlights data amenable to costing purposes.

The advantages and disadvantages of alternative data sources are listed in Table 1. For example, Hospital Episode Statistics (HES) a data warehouse of details of all admissions, outpatient appointments and A&E attendances at NHS hospitals in England do not contain unbundled HRGs, but do contain operational procedure codes (OPCS) which can be used to generate some unbundled HRGs such as those associated with diagnostic imaging or expensive drugs. Perhaps the biggest limitation of HES data is that they cannot be used to generate unbundled HRGs for critical care which can vary by as much as several thousand pounds per day per HRG. Costing carried out by NHS Trusts can also provide patient spell level data using outputs from Patient Level Information and Costing

Systems (PLICS). These provide spell HRGs, admission dates and discharge dates, and can also identify clinical codes and total costs associated with critical care and expensive drugs although these costs tend not to be disaggregated into their unbundled HRGs, as seen in SUS PbR.

Insert Figure 2 and Table 1 here

Besides the perspective and methods of resource use data collection, there needs to be consideration of the scope of the data being collected to – whether only intervention-related resource use data should be collected, or all resource use, remains a moot point. Electronic databases offer the advantage of providing often complete and extensive resource-use data, allowing for cost effectiveness to be presented at various levels of inclusion of resource use. There are disadvantages, however, to accessing routine electronic data, which include the cost, the delay between patient exiting a trial and the data becoming available, potential for missing data, and the complexity of data management. The prospect of more extensive linked data sources will benefit health economists engaged in trial-based evaluations especially if they can ease the burden associated with accessing, extracting and interpreting the coding within the data systems in order to determine the relevant costs. In Wales there already exists the Secure Anonymised Information Linkage (SAIL) databank which uses a matching algorithm for GP care, secondary care and local authority social services [30]. A comparable initiative across England, Care.data, is currently under development.

4. Methods for measuring health outcomes

The conventional view from an economics perspective, that RCTs are mainly an opportunity for economic data collection, needs to be challenged. Economists need to be embedded within the trial management group and encouraged to lead on SWATs (Studies Within a Trial) to develop novel methodologies applicable to trial design, conduct or analysis. Besides costs, there are also opportunities for developing and refining health outcome measures. Key to this is the shift towards involving patients generally, and ascertaining their preferences, specifically, in patient-reported outcome measures of health.

4.1. Choice of outcome measure

The choice of economic outcome measure for a trial is dependent on the research question, the setting and characteristics of the trial participants. In the UK, NICE favours utilities derived from the European Quality of Life – 5 Dimension – 3 Levels (EQ-5D-3L) in adults, though accepts mapped EQ-5D-3L utilities from other health-related quality of life (HR-QoL) measures in situations where direct measurements are not available. A five level version of the EQ-5D (EQ-5D-5L) has been developed but as yet no scoring algorithm has been released for calculating utility values. Thus, the established EQ-5D-3L is currently more widely used in economic evaluation. In cases where the EQ-5D lacks content validity, which considers whether an instrument contains all the domains required to measure what it is attempting to measure, NICE suggests that alternative HR-QoL measures may be used, but does not explicitly refer to preference-based, disease-specific utility measures.

Challenges with eliciting patient-reported health outcome data can arise when the study population have cognitive limitations, for example dementia or learning difficulties, or when the study population involves children and young people. A systematic review of the use of the EQ-5D-3L with people with dementia found a satisfactory completion rate for people in the mild to moderate stage

of dementia due to its brevity [31]. To achieve a satisfactory completion rate in studies involving children and young people, the design of data collection materials e.g. questionnaire, form or diary, should be tailored for age. Satisfactory completion rates may be achieved, for example, by adapting questionnaire design in terms of colour and personalisation for different age categories [32,33]. Alternative validated preference-based utility measures validated for children include the HUI-2, CHU9D and AQoL-6D [34].

In cases where trial participants are unable to answer questions themselves, the use of a proxy to complete questionnaires on behalf of the participant may be considered. This in turn raises the question of whether the proxy should be a family member with direct experience of the participant's health, or a healthcare professional who has a broader experience of the condition. With reference to people with dementia, EQ-5D rated quality of life is consistently scored higher by patients than by their proxies [31]. Similarly, children and young people may rate their quality of life higher than their proxies, such as in the case of diabetes where they respond more positively to the PedsQL-General Module, diabetes self-efficacy (PedsQL-Diabetes Module) and their health utility (EQ-5D) than their proxies [32]. In the context of clinical trials, it is important to record who is completing proxy questionnaires to avoid any uncertainty over whether the questionnaire was completed by the family member, carer, healthcare professional or researcher. This may be facilitated by providing training and instructions to researchers on how to interview and how to administer questionnaires consistently. [35]

Healthcare interventions may have an HR-QoL impact beyond the index patient, for example on family carers. Al-Janabi et al [36] pose the argument that using a HR-QoL measure to assess carer quality of life imposes upon them a 'patient identity', misinterpreting their role. Their suggested solution was to measure care-related QoL instead.

Use of generic, preference-based utility measures allows for the consistent comparison of results across different settings, populations, and interventions; however disease-specific, preference-based measures allow a greater degree of sensitivity. By way of example, the epilepsy-specific quality adjusted life year (QALY) measure, NEWQOL-6D, was derived from the NEWQOL measure of health-related quality of life, firstly by psychometric and Rasch analyses to establish the dimension structure of NEWQOL and secondly, by valuing the health states generated by the classification system using Time Trade Off [37]. The results were modelled to generate a utility score for every health state defined by 6 dimensions (worry about attacks; depression; memory; concentration; stigma; control). A re-analysis of economic evaluation of an RCT comparing standard versus new antiepileptic drugs [38], revealed a different rank ordering of cost-effectiveness based on the NEWQOL-6D than with the EQ-5D, suggesting that selection of dimensions of health more relevant to epilepsy might affect policy choices. However the increased sensitivity of disease-specific preference-based outcome measures is traded against a loss of comparability of utility across different interventions and a potential insensitivity to side effects and /or comorbidities [39]. Further development of disease-specific utility measures, their comparison against generic instruments and impacts on decisions concerning technical and allocative efficiency is warranted.

Within a trial, QALYs are typically analysed according to treatment allocation; however trials are rarely powered to detect differences in QALYs. An alternative is to adopt an event-based approach, whereby clinical events (e.g. stroke, fracture, myocardial infarction) are included as covariates in the

regression equation [40]. Results allow a utility (decrement) to be attached to clinical events, and provide a coefficient (mediated by clinical events) which captures utility differences by trial arm. Briggs et al [41] demonstrated that for an RCT where no significant difference in utility was observed using a traditional approach, event-based methods resulted in a moderately significant difference between treatment groups. This approach is more closely aligned with model-based economic evaluations, where utilities are typically attached to health states.

4.2. Calculating minimally important differences

Preference surveys and ranking exercises can be used to inform the choice of outcome of most importance, however, methods that measure patients' willingness-to-trade enables researchers to profile the impact of differences in health technologies with common characteristics to ascertain the most acceptable threshold at which harms and benefits need to be considered (traded).

Minimally important differences are necessary for calculating the sample sizes of trials to ensure adequate powering; however these are typically determined by clinicians, whose views may not necessarily align with those of patients.

Consider an example of a drug which is effective in causing remission of disease symptoms, but is associated with dose-dependent toxicity that may reduce survival. A trial using a lower dose is to be designed, with survival as the primary outcome, and which requires a minimally important difference to calculate the sample size. Using a discrete choice experiment (DCE) to determine preferences in the absence of other data, with remission and survival as attributes, the relative importance of movements between levels of each attribute can be estimated as the marginal rate of substitution (MRS) [42]. This can be used to estimate patients' maximum acceptable decrease in survival, for a given increase in remission. The MRS is calculated as follows:

$$MRS_{remission} = \beta_{remission} / \beta_{survival}$$

where $\beta_{remission}$ and $\beta_{survival}$ are the coefficients for each attribute, respectively.

Using a known decrease in remission between low and standard dose ($\Delta_{remission} = remission_{low} - remission_{standard\ dose}$), the minimum percentage difference in overall survival ($= \Delta_{survival}$) which a DCE respondent would be willing to accept (where the utility associated with the lower dose is greater than or equal to the utility of standard dose), is estimated as:

$$-MRS_{remission} * \Delta_{remission} \leq \Delta_{survival}$$

If the difference in survival between treatment groups is less than $\Delta_{survival}$ then from a patient's perspective, the treatment option with lower recurrence would be chosen.

In the case of multiple attributes, this generalises to:

$$-\sum_{1}^N MRS_{attribute(N)} * \Delta_{attribute(N)} \leq \Delta_{survival}$$

This trade-off between different outcomes of interest provides the point of indifference from the respondent's perspective and therefore represents the minimally important difference in the primary outcome measure (probability of survival) between choosing the trial intervention or not.

Caveats are drawn from the type of study required to derive this result; sufficient qualitative research into the design of the DCE are required to ensure that attributes are meaningful both to the patient, and clinically. It is noted that the DCE are estimated with uncertainty, and responses vary depending upon an individual patients' situation and preferences. Indeed, for a given patient, the MRS may not be constant along the indifference curve. Notwithstanding these caveats, this method may provide a quantitative means of involving patients in the design stage of an RCT, resulting in an effect size that is relevant and meaningful to the patient, rather than chosen based solely on clinician consensus or in concordance with previous historic trials where the initial effect size may have been chosen somewhat more arbitrarily.

5. Handling data-related methodological challenges

5.1. Methods of extrapolation

Trial follow-up periods are typically shorter than the period during which differences in health effects and use of healthcare resources between interventions persist. Consequently, time horizon bias is a problematic feature in trial-based economic evaluations. Consider two time scales of importance; t_{RCT} which refers to the duration of an RCT (the observed period), and t_{TH} which refers to the time horizon of an economic evaluation (the unobserved period). In general, an economic analysis at t_{TH} requires extrapolation of the incomplete (censored) data from $t \leq t_{RCT}$ in order to generate predictions of health outcomes and costs after the RCT ceases. Multivariable regression-based statistical models fitted to RCT data allow simple extrapolation for trial participants beyond the observed period given information on individual-level characteristics such as age, sex, disease history, or the presence of other known risk factors that appear as predictor variables in the model. This approach, however, may be limited by the validity of assumptions such as the linear dependence of health outcomes on predictor variables or the appropriateness of extrapolating associations beyond the trial period, plus factors such as censored (or missing) individual-level data or bias arising from loss to follow-up, so such models should be used with caution.

Other common approaches include the use Markov models, which take the form of dynamic discrete-time models represented by difference equations (or, in continuous-time, differential equations) or, extrapolating survival/time-to-event data by fitting parametric models to RCT data and assuming proportional hazards (PHs) to determine intervention group survival. In the latter case, however, PHs becomes an increasingly unreliable assumption as t_{TH} increases towards lifetime duration due to additional factors such as age-dependent mortality [43]. Alternatives to parametric models for time-to-event data (such as spline-based methods, piecewise modelling, and other semi-parametric or non-parametric models) have also been proposed [44].

Whatever the method, robust extrapolation relies on thoughtful model development, reliable parameterisation, and intelligent fitting to data; uncertainty in extrapolation increases as the ratio t_{TH}/t_{RCT} increases, so the time horizon and time-discounting rates are critical. When fitting models to RCT data, fitting to only (stable) data in the observed period that most closely replicates expected conditions in the unobserved period, rather than the full dataset, may lead to more efficient use of the available evidence [45]. If models used for economic evaluation are underpinned by dynamic mechanistic disease models (such as in the case of infectious disease), consideration must be given to how the time scales of diseases processes relate to t_{RCT} in order to ensure that model fit is not

based on transient model behaviour that will not be representative of long-term disease dynamics reached within the unobserved period (Figure 3).

Insert Figure 3 here

Model design and parameterisation should be based on underlying assumptions informed by the problem (and disease) in hand, which should, in turn, be motivated by factors such as pathophysiological processes, clinical and epidemiological knowledge, previous modelling work, statistical evidence, and/or desired model complexity. Assimilation of this knowledge unavoidably introduces uncertainty in the model construction process, together with uncertainty arising from the data itself (such as quality, bias or missing data). Bias in model selection due to (subjective) modeller preference also represents a potential source of error.

The central methodological issue in extrapolation is to understand the robustness of (and confidence that should be attached to) long-term model predictions (and economic decisions) given the array of structural, parameter, and methodological uncertainties (as well as variability and heterogeneities) associated with the modelling process [46]. This question ideally necessitates (a) identifying and specifying all uncertainties arising from the modelling process, (b) quantifying (and representing) the resultant total uncertainty arising in the economic evaluation associated with extrapolation, and (c) assessing which sources of uncertainty most significantly affect the conclusions of the economic evaluation, although this 3-step procedure is currently rarely followed in full. Included in the process of quantifying uncertainty in model development should be a critique of whether structural assumptions in model design for $t \leq t_{RCT}$ are valid for (and can be trusted to generalise to) $t > t_{RCT}$; examples include consideration of whether new interventions may change in efficacy over time (e.g. new antimalarial drugs due to resistance) or whether temporal changes in disease parameters (e.g. due to seasonal fluctuations) are important. Structural uncertainty may be addressed using scenario analysis or model comparison/averaging [47], but it is important to recognise that good model fit for $t \leq t_{RCT}$ does not guarantee model reliability, accuracy or robustness of assumptions for $t > t_{RCT}$; if t_{RCT} is considerably less than t_{TH} , structurally different models with almost equivalent goodness-of-fit in the observed period can produce very different predictions for cost-effectiveness at t_{TH} (see Figure 4).

Insert Figure 4 here

Ultimately, model validation, assessing a model's accuracy in making predictions for the specific purpose of interest, is critical to confidence in the extrapolation process. As a minimum, this should include a careful critique of model structure, parameterisation, and problem formulation (in the specific context of $t > t_{RCT}$ conditions), but should ideally consider more rigorous validation principles including examination of different models addressing the same problem (and assessment of the implications for resultant economic evaluations and decision-making), comparison with relevant available data from sources beyond those from the RCT used for model fitting, and identification of opportunities to compare the predictions of models with observable events and outcomes that may occur for $t > t_{RCT}$ (arguably the most desirable form of validation) [48]. Dominant sources of uncertainty in the unobserved period may not be the same as those arising in the observed period, and the results may depend on the problem in hand, the type of outcomes of interest, and the nature of the model itself.

5.2. Dealing with missing data

Missing data is almost unavoidable in clinical trials and can impact on economic evaluations in terms of loss of precision and statistical power, which may bias the results such that any consequent conclusions made might lead to incorrect policy decisions. Resource use, cost or health outcome data derived from patient questionnaires or case report forms are prone to being incomplete, as are alternative electronic data sources, such as hospital episode statistics, which also have the potential to contain anomalies that are required to be treated as missing data.

The reporting of missing data in trial-based economic evaluations and the methods used to handle missing data are varied and unclear, with sensitivity analysis around the missing data rarely being performed [49]. Poor handling of missing data introduces not only the risk of imprecision, loss of power and bias, but in addition, missing baseline data becomes problematic when estimating effect size [50], leading to challenges in cost-effectiveness analysis when baseline adjustments are considered [51]. Sterne et al [52], offer guidance on the reporting of missing data, and the ISPOR Task Force report on conducting cost-effectiveness analysis alongside RCTs [53] make explicit recommendations on both the analysis and reporting of missing data. The Consolidated Standards of Reporting Trials (CONSORT) guidelines require that RCTs report the flow of patients through the trial, and the numbers included for analysis [54], however no such clause is included for reporting how missing data is analysed [55]. Neither the NICE reference case [7] nor the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [56] explicitly require the reporting of missing data, or give guidance as to how missing data ought to be analysed, although it may be assumed that missing data contributes to uncertainty, which is covered by the guidelines [49].

The potential for missing data should be realised at the trial design stage, with procedures implemented to minimise wherever possible [57]. Patient questionnaires require careful design and piloting, with consideration given as to how, when and where they are administered, as described above. With regards to cost data, it may be prudent to explore multiple sources of data, electronic in combination with case report forms, so as to have overlap and a secondary source from which to replace missing data.

The mechanism of missing data can be classified into three types, missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [58]. Complete case analysis, effectively ignoring incomplete data, remains popular for trial data [49,57], but is valid only when data are assumed to be MCAR. When the MCAR assumption does not hold, however, the data may no longer be representative of the target population, thus compromising external validity [59]. Ad-hoc methods, such as mean imputation, attempt to address missing data, but are also only valid under MCAR. Whilst preserving point estimates, mean imputation attenuates variance, leading to artificially small confidence intervals.

Maximum-likelihood estimation (MLE) methods such as expectation maximisation are valid under the less restrictive MAR assumption, however separate MLE models are required for each different analysis. Additionally, MLE can be challenging for certain approaches such as generalised linear modelling [60]. Multiple imputation (MI) [50] is also valid under MAR, and requires only a single model, however other challenges are introduced, as multiple imputed data sets require to be analysed, and methodology for pooling results is not always intuitive, often with transforms required prior to pooling [61]. Nevertheless, MI has been shown to be robust to departures from normality, in

cases of low sample size, and when the proportion of missing data is high [62]. MNAR specific methods include the Heckman selection model [63] and pattern mixture models [64], and a novel weighting approach adapts MI specifically for data which are MNAR [65]. It is noted however that for data which are MNAR, MLE and MI analysed using information relevant covariates often produces accurate unbiased results [66,67]. Methods such as these should be applied, where indicated, in trial-based economic evaluations.

5.3. Accounting for non-adherence

A key issue when generalising the findings of controlled trials to clinical practice is whether efficacy can translate to effectiveness; that is, whether the intervention which evidently works, works in practice. While adherence encompasses healthcare professionals' delivery of an intervention, as well as following clinical guidelines, a major factor that can impact on a treatment's effectiveness is the effect of patient non-adherence. It is generally recognised that there exists significant differences between adherence to (or persistence with) treatments in efficacy trials and routine care. In the case of 3-hydroxy-3-methylglutarate-CoA reductase inhibitors (statins) for instance, adherence following 4-6 years' follow-up of patients in pivotal trials of secondary prevention ranges from 81% to 99%. This contrasts with pragmatic studies conducted in community settings, where adherence ranges from 21% to 87% [68]. Differences may be due to intensity of monitoring, frequency of follow-up clinic visits and selection of patients, either through protocol specification or through self-selection of interested (and hence more adherent) patients. Failure to consider patients' non-adherence to treatments or interventions in economic evaluations can consequently lead to biased estimates of cost-effectiveness, with treatments potentially not being as cost-effective as anticipated. Reviews of economic evaluations indicate that only a minority assess the impact of non-adherence explicitly [69].

Methods for adjusting effect and cost data for non-adherence and /or premature discontinuation are mainly reliant on conventional Markov or decision analytic approach, with assumed reductions in benefit and a change in cost, for reductions in adherence [70,71]. Patients who discontinue treatment are commonly assumed to experience the benefits and incur the costs associated with a placebo group. These strong assumptions may not be valid given that patients may rationally discontinue treatment if they experience adverse reactions or perceive no benefit from treatments for symptomatic diseases. Placebo data, representing no treatment, may not be available in many trials. Alternative approaches require knowledge of baseline covariates that predict adherence for each treatment group of the trial. Fischer et al [72] propose a structural mean modelling approach to obtain adherence-adjusted estimates for treatment effects in an RCT comparing two active treatments; Kadambi et al [69], propose discrete event simulation for economic analyses, as this may be more amendable to modelling individuals' risk of non-adherence, and the link between reduced efficacy or increased adverse events and treatment discontinuation. The use of mechanism-based modelling, in which individual doses are simulated according to known patterns of adherence, and the pharmacokinetic and pharmacodynamics consequences are propagated to estimate the reduction in treatment effect, is an alternative approach that may assist in predicting the dilution of effect from perfect to typical behaviour of re-medicating [73]. Further research is warranted for methods of estimating the impact of variable adherence on the efficacy, and hence cost-effectiveness, of treatments.

6. Multinational trials

Drug trials are conducted increasingly at multiple centres across many nations. The decision to include more than one country for a given trial is typically driven by regulatory requirements, achieving the required sample size (e.g. for treatments of a rare disease), or a desire to complete within a shorter time frame [74]. Trials conducted in multiple countries may also enhance the generalizability of the findings to the participating countries [3,75] and serve to facilitate immediate adoption following regulatory approval. However, multinational trials are faced by methodological challenges that need to be considered at each stage.

While multinational trials increase the heterogeneity in the sample in terms of patient characteristics and disease severity [3,74,76], it remains common practice for most regulatory authorities to accept pooled estimates of clinical efficacy across countries. This is usually based on the assumption that the clinical response is unlikely to vary significantly and can therefore be considered homogenous [76]. However, difficulties arise when economic evaluations are conducted alongside multinational trials, where other sources of heterogeneity become pertinent. Moreover, an aggregate estimate of cost-effectiveness (or economic value) would be of limited, if any, value for each participating country's decision makers, where the use of country-specific estimates is usually a requirement for reimbursement authorities.

Costs of care vary widely from one country to another. This variation originates from differences in demographics, underlying disease epidemiology and patient mix. However, it is usually the differences in the price weights (both absolute and relative) and clinical practice that are the most influential [75,76], and these must be taken into account when analysing trial data. Prices may themselves influence clinical practice and resource use, for instance with substitution of more costly (or less subsidised) resources by cheaper ones. Differences in health care financing mechanisms also affect the choice of the perspective from which an economic evaluation is undertaken, be it a national health service, private insurer or the patient [35].

Due consideration should also be given to the choice of outcome measures to be used in the trial, the availability of translated and validated versions [77], and the selection of tariffs for preference-based measures. Differences in clinical practice between countries typically mean that the comparator may be less relevant, necessitating alternative or supplementary analysis e.g. using indirect comparisons.

The methods for multinational trial data analysis have been reviewed by Reed et al [74], who developed a classification system that combines both the type of analysis used for clinical effectiveness and resource use (fully pooled, partially split or fully split) and the costing method (one country or multi-country costing). Random-effects/multi-level modelling has been identified as a promising approach that offers an intermediate solution between fully split and fully pooled analysis and offers country specific estimates without compromising the power of the analysis [78]. In choosing an approach, Reed et al [74] recommend health economists need to balance a number of issues including, generalisability and transferability of data between countries, awareness of the statistical power and consideration of uncertainty within the analysis.

Researchers engaged in the design and analysis of economic evaluations based on multinational trials, are advised to consult relevant guidelines [76].

7. Conclusions

This review highlights selective key challenges in the design, conduct, analysis and reporting of trial-based economic evaluations. It unapologetically focuses on the research interests of the authors, and consequently adopts a UK-focus, but through this introduces some concepts and methods that have not yet gained mainstream adoption in trial-based economic evaluations. It should serve to stimulate debate around the appropriate approaches to best serve the evidential needs of decision makers.

8. References

1. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ* 2006;15(7):677-687.
2. Glick H, Doshi J, Sonad S, Polsky D. *Economic evaluation in clinical trials*. 1st ed. Oxford: Oxford University Press; 2007.
3. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ* 2011;342:d1548 doi:10.1136/bmj.d1548.
4. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. *Value Health* 2015;18(2):161-172.
5. 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):867-872.
6. Ahonkallio S, Santala M, Valtonen H, Martikainen H. Cost-minimisation analysis of endometrial thermal ablation in a day case or outpatient setting under different anaesthesia regimens. *Eur J Obstet Gynecol Reprod Biol* 2012;162(1):102-104.
7. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. <http://publications.nice.org.uk/pmg9>. Accessed 15 June 2015.
8. Evans WK, Coyle D, Gafni A, Walker H, National Cancer Institute of Canada Clinical Trials Group-Working Group on Economic Analysis. Which cancer clinical trials should be considered for economic evaluation? Selection criteria from the National Cancer Institute of Canada's Working Group on Economic Analysis. *Chronic Dis Can* 2003;24(4):102-107.
9. Revicki DA, Paramore C, Rentz AM. Incorporating pharmacoeconomic and health outcomes into randomized clinical trials. *Expert Rev Pharmacoecon Outcomes Res* 2005;5(6):695-703.
10. Marshall DA, Hux M. Design and analysis issues for economic analysis alongside clinical trials. *Med Care* 2009;47(7 Suppl 1):S14-20.
11. Ramsey SD, McIntosh M, Sullivan SD. Design issues for conducting cost-effectiveness analyses alongside clinical trials. *Annu Rev Public Health* 2001;22:129-141.
12. Medical Research Council. *Guidance on the development, evaluation and implementation of complex interventions to improve health*. 2008.

<http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>. Accessed 15 June 2015.

13. Gomes M, Grieve R, Nixon R, Edmunds WJ. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. *Med Decis Making* 2012;32(1):209-220.
14. 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;13:1001 doi:10.1186/1741-7015-11-21.
15. M. P. Kelly, D. McDaid, A. Ludbrook and J. Powell. Economic appraisal of public health interventions. NHS Health Development Agency Briefing Document. 2005. http://www.cawt.com/Site/11/Documents/Publications/Population%20Health/Economics%20o f%20Health%20Improvement/Economic_appraisal_of_public_health_interventions.pdf.
16. Payne K, McAllister M, Davies LM. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Econ* 2013;22(3):258-271.
17. 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.
18. Fischer AJ, Threlfall A, Meah S, Cookson R, Rutter H, Kelly MP. The appraisal of public health interventions: an overview. *J Public Health (Oxf)* 2013;35(4):488-494.
19. Martin A, Jones A, Mugford M, Shemilt I, Hancock R, Wittenberg R. Methods used to identify and measure resource use in economic evaluations: a systematic review of questionnaires for older people. *Health Econ* 2012;21(8):1017-1022.
20. J. Beecham and M. Knapp. Costing psychiatric interventions. Discussion Paper 1536. PSSRU, University of Kent at Canterbury. 1999. <http://www.pssru.ac.uk/pdf/dp1536.pdf>. Accessed 28 October 2013.
21. Chisholm D, Knapp MR, Knudsen HC, Amaddeo F, Gaité L, van Wijngaarden B. Client Socio-Demographic and Service Receipt Inventory--European Version: development of an instrument for international research. EPSILON Study 5. *European Psychiatric Services: Inputs Linked to Outcome Domains and Needs*. *Br J Psychiatry Suppl* 2000;177(39):s28-33.
22. Clarke PM, Fiebig DG, Gerdtham UG. Optimal recall length in survey design. *J Health Econ* 2008;27(5):1275-1284.
23. Brown JB, Adams ME. Patients as reliable reporters of medical care process. Recall of ambulatory encounter events. *Med Care* 1992;30(5):400-411.
24. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value Health* 2000;3(4):277-286.
25. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev* 2006;63(2):217-235.

26. Richards SH, Coast J, Peters TJ. Patient-reported use of health service resources compared with information from health providers. *Health Soc Care Community* 2003;11(6):510-518.
27. Ridyard CH, Hughes DA, DIRUM Team. Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. *Value Health* 2012;15(5):650-655.
28. McPherson CJ, Addington-Hall JM. Judging the quality of care at the end of life: can proxies provide reliable information? *Soc Sci Med* 2003;56(1):95-109.
29. Department of Health Payment by Results team. A simple guide to payment by results. November 2012.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf.
30. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009;9:3 doi:10.1186/1472-6947-9-3.
31. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. *Value Health* 2011;14(2):390-399.
32. Noyes JP, Williams A, Allen D, Brocklehurst P, Carter C, Gregory JW, et al. Evidence into practice: evaluating a child-centred intervention for diabetes medicine management. The EPIC Project. *BMC Pediatr* 2010;10:70 doi:10.1186/1471-2431-10-70.
33. Noyes JP, Lowes L, Whitaker R, Allen D, Carter C, Edwards RT, et al. Developing and evaluating a child-centred intervention for diabetes medicine management using mixed methods and a multicentre randomised controlled trial. *Health Services and Delivery Research* 2014;2(8) doi:10.3310/hsdr02080.
34. Thorrington D, Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. *PLoS One* 2015;10:8:e0135672 doi:10.1371/journal.pone.0135672.
35. Drummond MF, Sculpher M, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press; 2005.
36. Al-Janabi H, Flynn TN, Coast J. QALYs and carers. *Pharmacoeconomics* 2011;29(12):1015-1023.
37. Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Hughes D, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. *Epilepsy Behav* 2012;24(1):36-43.
38. Marson AG, Appleton R, Baker GA, Chadwick DW, Doughty J, Eaton B, et al. A randomised controlled trial examining the longer-term outcomes of standard versus new antiepileptic drugs. The SANAD trial. *Health Technol Assess* 2007;11(37):iii-iv, ix-x, 1-134.
39. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? *Value Health* 2012;15(3):504-513.
40. Clarke P, Gray A, Holman R. Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). *Med Decis Making* 2002;22(4):340-349.

41. Briggs AH, Bousquet J, Wallace MV, Busse WW, Clark TJ, Pedersen SE, et al. Cost-effectiveness of asthma control: an economic appraisal of the GOAL study. *Allergy* 2006;61(5):531-536.
42. Ryan M, Gerard K, Amaya-Amaya M. *Using Discrete Choice Experiments to Value Health and Health Care*. (Vol. 11). Dordrecht, The Netherlands: Springer; 2008.
43. Latimer NR. Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making* 2013;33(6):743-754.
44. Grieve R, Hawkins N, Pennington M. Extrapolation of survival data in cost-effectiveness analyses: improving the current state of play. *Med Decis Making* 2013;33(6):740-742.
45. Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Med Decis Making* 2014;34(3):343-351.
46. Bilcke J, Beutels P, Brisson M, Jit M. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. *Med Decis Making* 2011;31(4):675-692.
47. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural uncertainty in decision models. *Med Decis Making* 2011;31(4):662-674.
48. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value Health* 2012;15(6):843-850.
49. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Econ* 2012;21(2):187-200.
50. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30(4):377-399.
51. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med* 2005;24(7):993-1007.
52. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393 doi:10.1136/bmj.b2393.
53. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* 2005;8(5):521-533.
54. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332 doi:10.1136/bmj.c332.
55. Sylvestre Y. CONSORT: Missing missing data, guidelines, the effects on HTA monograph reporting. *Trials* 2011;12(Suppl 1): A61 doi:10.1186/1745-6215-12-S1-A61.

56. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Pharmacoeconomics* 2013;31(5):361-367.
57. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004;1(4):368-376.
58. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, NJ: Wiley; 2002.
59. Little RJA. Missing-Data Adjustments in Large Surveys. *J Bus Econ Stat* 1988;6(3):287-296.
60. Sinharay S, Stern HS, Russell D. The use of multiple imputation for the analysis of missing data. *Psychol Methods* 2001;6(4):317-329.
61. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;9:57 doi:10.1186/1471-2288-9-57.
62. Wayman J. Multiple imputation for missing data: What is it and how can I use it? Paper presented at the Proceedings of the Annual meeting of the American Educational Research Association. Chicago, Illinois, USA. 2003.
63. Heckman J. Sample selection bias as a specification error. *Econometrica* 1979;47(1):153-61.
64. Little R. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993;88(421):125-134.
65. Carpenter JR, Kenward MG, White IR. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Stat Methods Med Res* 2007;16(3):259-275.
66. Rubin DB, Stern HS, Vehovar V. Handling “don’t know” survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association* 1995;90:822–828.
67. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002;7(2):147-177.
68. Shroufi A, Powles JW. Adherence and chemoprevention in major cardiovascular disease: a simulation study of the benefits of additional use of statins. *J Epidemiol Community Health* 2010;64(2):109-113.
69. Kadambi A, Leipold RJ, Kansal AR, Sorensen S, Getsios D. Inclusion of compliance and persistence in economic models: past, present and future. *Appl Health Econ Health Policy* 2012;10(6):365-379.
70. Hughes DA, Bagust A, Haycox A, Walley T. Accounting for noncompliance in pharmacoeconomic evaluations. *Pharmacoeconomics* 2001;19(12):1185-1197.
71. Hughes D, Cowell W, Koncz T, Cramer J, International Society for Pharmacoeconomics & Outcomes Research Economics of Medication Compliance Working Group. Methods for

- integrating medication compliance and persistence in pharmacoeconomic evaluations. *Value Health* 2007;10(6):498-509.
72. Fischer K, Goetghebeur E, Vrijens B, White IR. A structural mean model to allow for noncompliance in a randomized trial comparing 2 active treatments. *Biostatistics* 2011;12(2):247-257.
 73. Hughes DA, Walley T. Predicting "real world" effectiveness by integrating adherence with pharmacodynamic modeling. *Clin Pharmacol Ther* 2003;74(1):1-8.
 74. Reed SD, Anstrom KJ, Bakhai A, Briggs AH, Califf RM, Cohen DJ, et al. Conducting economic evaluations alongside multinational clinical trials: toward a research consensus. *Am Heart J* 2005;149(3):434-443.
 75. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess* 2004;8(49):iii-iv, 1-192.
 76. 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):409-418.
 77. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8(2):94-104.
 78. Manca A, Sculpher MJ, Goeree R. The analysis of multinational cost-effectiveness data for reimbursement decisions: a critical appraisal of recent methodological developments. *Pharmacoeconomics* 2010;28(12):1079-1096.
 79. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. 1st ed. Princeton, NJ: Princeton University Press; 2007.

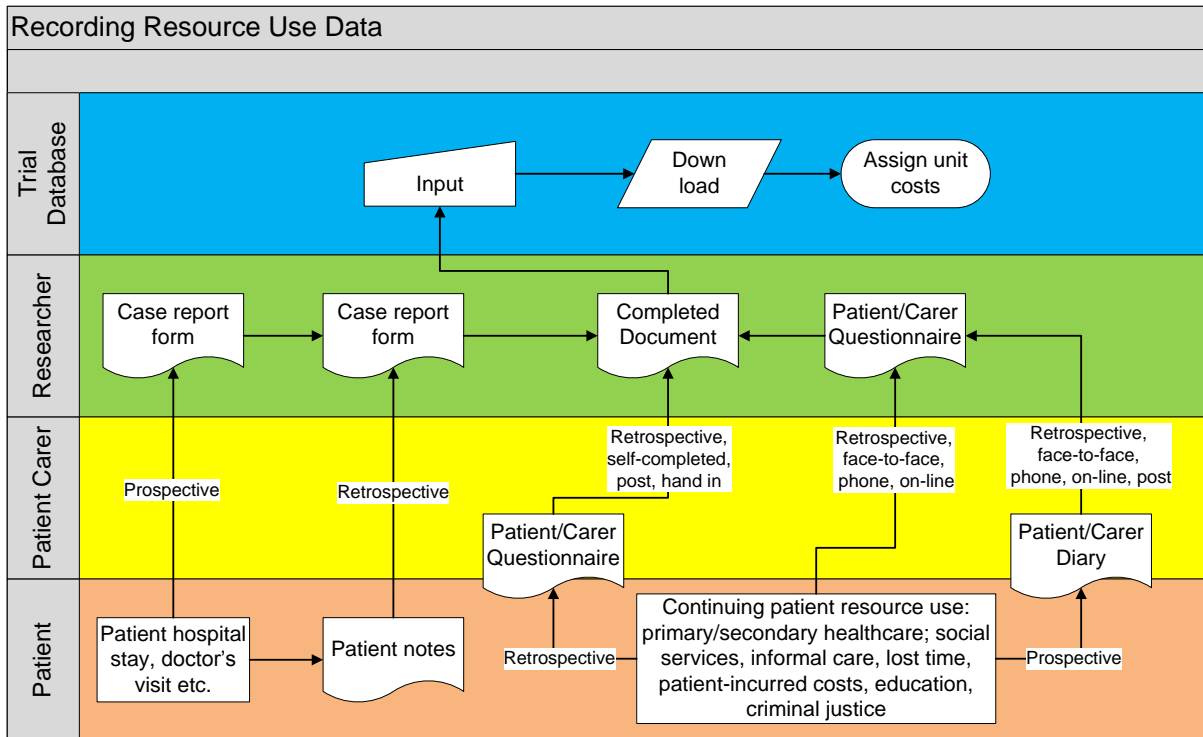


Figure 1: Methods for measuring resource use and assigning costs within trial-based economic evaluations

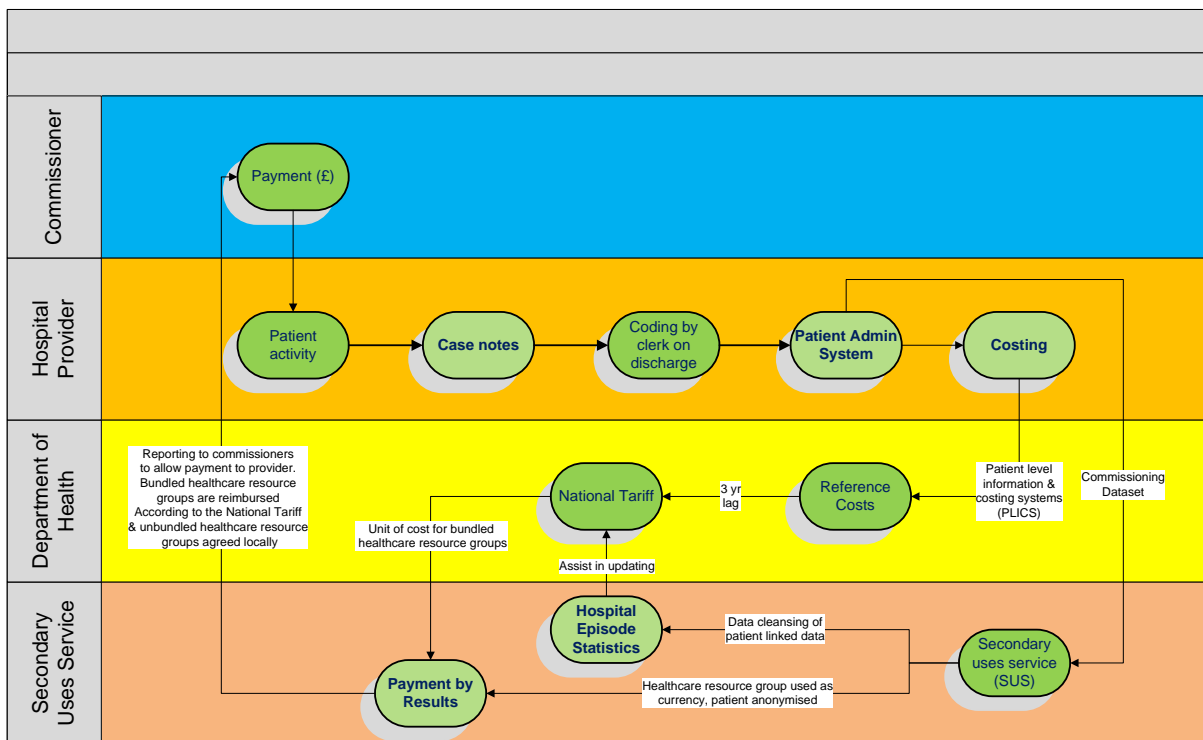


Figure 2: Flow of electronic data in secondary health care in the English National Health Service [29]. Cells in paler green represent the key areas where routinely collected data could be utilised for trial-based economic evaluations.

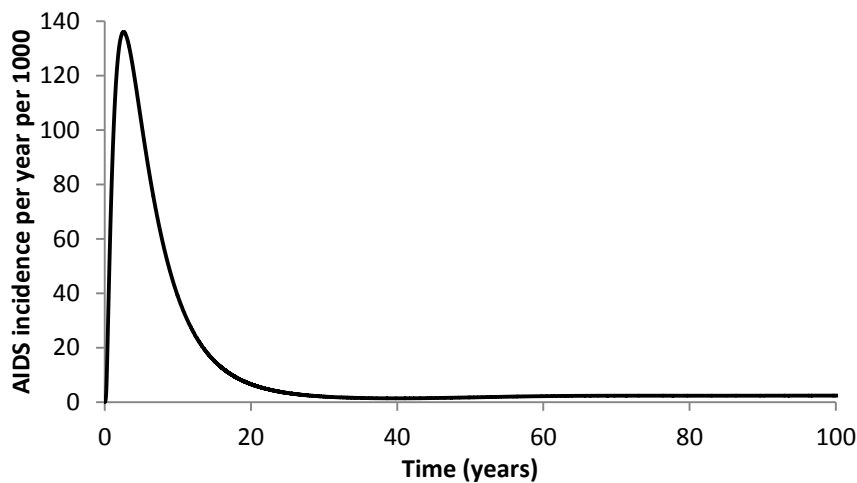


Figure 3: HIV infection dynamics based on the dynamic mechanistic transmission model (and parameters) of Keeling and Rohani [79] showing the total AIDS incidence (per year per 1000 population) across all sexual activity classes. Calculating the cost-effectiveness of a new intervention against HIV, based on extrapolating data obtained from an RCT (which typically take place over a relatively short timescale), will depend strongly on what phase of the epidemic curve is used for disease predictions and the time horizon of interest. If the RCT takes place near the start of a new HIV epidemic (when incidence increases approximately exponentially), this will give very different estimates of long-term cost-effectiveness compared to estimation in a population where the disease has already become endemic (here, after around 60 years). In the context of endemic infections, this early rapid growth and subsequent decline in the number of cases is an example of transient model behaviour, which differs significantly from the long-term disease dynamics, which, here, reach a steady-state.

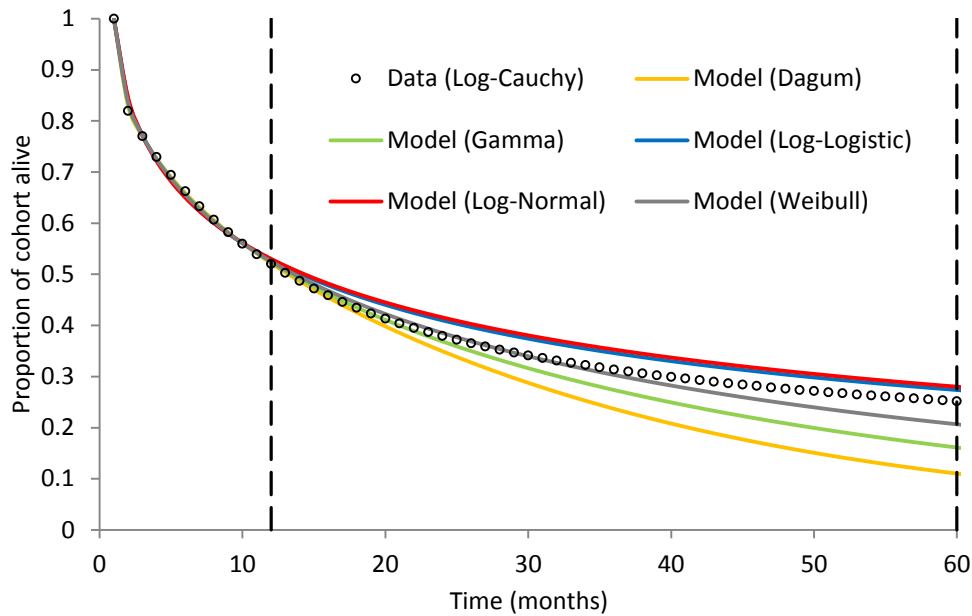


Figure 4: Illustrative example of extrapolating RCT survival data from an observed period (here, one year) to a follow-up period of four years using a simple 2-state Markov model, after which an economic evaluation of the intervention cost-effectiveness is undertaken. In this case, the Markov model consists of two states ('Well' and 'Dead', with the probability of moving from the former to the latter within a time-step given by the survival distribution of interest). Here, monthly simulated RCT data is generated from a log-Cauchy distribution and five parametric survival models (gamma, log-normal, Dagum, log-logistic and Weibull) are fitted to the RCT data using least-squares and then extrapolated using the Markov model to a time horizon of five years. Despite a near identical fit to the trial data, estimation of the cost-effectiveness over the follow-up period differs significantly depending on the parametric model adopted.

Source	Advantages	Disadvantages
Case Notes	<ul style="list-style-type: none"> • Close to source • Patient linkable • Can use local costs such as cost per bed day 	<ul style="list-style-type: none"> • Limited data • Hard to collect if require data from many hospitals • Have to be transcribed to CRF or trial database
Patient Administration System	<ul style="list-style-type: none"> • Close to source • Patient linkable • Can use local costs such as cost per bed day 	<ul style="list-style-type: none"> • Limited data • Hard to collect if require data from many hospitals • Practical issues in obtaining data from individual Trusts
Patient-level information and costing systems (PLICS)	<ul style="list-style-type: none"> • Detailed patient-level information (e.g. HRGs, clinical codes) • Patient linkable • Overall cost per patient per spell from hospital perspective 	<ul style="list-style-type: none"> • Hard to collect if require data from many hospitals • Delays in data generation due to detailed costing being undertaken at the end of a financial year
Hospital Episode Statistics (HES)	<ul style="list-style-type: none"> • Detailed patient-level information (e.g. HRGs, clinical codes) • Patient linkable • Shows attendances at other hospitals • Shows finished consultant episodes (FCEs) 	<ul style="list-style-type: none"> • Does not contain unbundled HRGs • Unsuitable for critical care • Long delays in data availability due to data cleansing • Tariff information such as market forces factors need to be linked in • Due to changes in management and the disbandment of the NHS Information Centre, information has become harder to obtain.
Secondary Uses Service (SUS) Payment by Results (PbR)	<ul style="list-style-type: none"> • Detailed patient-level information (e.g. HRGs, clinical codes) • Contains unbundled HRGs • Pseudonymised patient level • Tariff information such as market forces factors included 	<ul style="list-style-type: none"> • Not patient linkable

Table 1: Advantages and disadvantages of routinely available data in the National Health Service in the UK