

# Rationale and methodology for trial-based economic evaluation

**Clin. Invest.** (2012) 2(12), 1191–1200

Economic evaluation provides a means of allocating finite healthcare resources in an efficient manner. It can inform decision-making processes at many levels, from national decision-making bodies, such as the NICE in England and Wales, to decisions by local healthcare providers. A common vehicle for the conduct of economic evaluation is the randomized controlled trial. This paper provides an overview of the methodologies underpinning economic evaluations based on randomized controlled trials. It covers broad design issues, approaches to measuring resource inputs, approaches to valuing resource inputs, approaches to the measurement and valuation of outcomes, the analysis of data inputs, the comparative analysis of costs and consequences, and methods for handling uncertainty and extrapolating cost–effectiveness. The strengths and limitations of trial-based economic evaluations in comparison with other designs are outlined.

## Stavros Petrou

Warwick Medical School, University of Warwick,  
Coventry, CV4 7AL, UK.  
Tel.: +44 2476 151124  
Fax: +44 2476 151586  
E-mail: s.petrou@warwick.ac.uk

**Keywords:** economic evaluation • costs • health utilities • methodology • trials

Economic evaluation provides a framework for assessing the costs and consequences of alternative programs or interventions [1]. Within the healthcare context, it aims to identify the combination of human and material inputs that maximize health benefits or other measures of social welfare [1]. Economic evaluation has increasingly been used to inform healthcare decision-making in the UK by bodies such as the NICE for England and Wales and the Scottish Medicines Consortium for Scotland [2,3]. Similarly, economic evaluation has increasingly been used to inform the regulatory and reimbursement decisions of government agencies in other nations [4,5]. A common vehicle for the conduct of economic evaluation is the randomized controlled trial (RCT). Indeed, many trial funders, such as the National Institute for Health Research Health Technology Assessment Program in the United Kingdom, routinely require assessments of cost–effectiveness to be incorporated into the design of RCTs. This is reflected in the prevalence of trial-based economic evaluations reported in the literature. Information extracted from the NHS Economic Evaluation Database revealed that over 30% of economic evaluations published in recent years were based on data from a single RCT, with this proportion increasing noticeably over the lifetime of the database [101].

There are several arguments in favor of conducting economic evaluations alongside RCTs. RCTs are generally commissioned because of a lack of existing evidence on treatment effect, and so provide an early opportunity to produce reliable estimates of cost–effectiveness. Trial-based economic evaluations tend to have a low marginal cost compared with alternative study designs. Importantly, they provide access to a wealth of individual patient data to which a wide range of statistical and econometric techniques examining relationships between clinical and economic parameters of interest can be applied. Some

**FUTURE  
SCIENCE**

part of

**fsg**

health economists counter that trial-based economic evaluations tend to be limited by their truncated time horizons, limited comparators, lack of relevance to the decision context in a specific country, and failure to incorporate all relevant evidence, for example, from other trials or observational studies [6]. This has manifested a rather sterile debate about the relative merits of trial-based economic evaluations versus economic evaluations using decision analytic modeling, when in truth the two are often complements rather than alternatives [7]. Indeed, trial-based economic evaluations often require the application of complex modeling techniques. This paper provides an overview of the methodologies underpinning economic evaluations based on RCTs.

### Design of trial-based economic evaluations

Designing a rigorous trial-based economic evaluation requires multidisciplinary input in much the same way that the design of a robust RCT is not restricted to the purview of one disciplinary group. Increasingly, this requires coordination by an accredited clinical trials unit. Indeed, the design requirements of trial-based economic evaluations are increasingly reflected in the standard operating procedures of clinical trials units and within the broader governance structures of individual trials [8]. The research instruments and procedures used to collect relevant health economic data should ideally be pilot tested for efficiency, clarity and ease of use. A good example is provided by a recent analysis of computer-aided self-exposure therapy for phobia or panic disorder where researchers pilot-tested the economic data collection instruments and procedures for subsequent use [9]. Two pertinent issues that should be considered at the design stage are, first, the implications of basing the economic evaluation on an explanatory rather than a pragmatic trial and, second, the appropriate sample size and statistical power for economic end points.

The pragmatic trial offers analysts an opportunity to evaluate the cost-effectiveness of a healthcare program or intervention under real-world conditions. This typically results in enrolling patients representative of the general clinical caseload from representative settings, comparing the intervention of interest with current practice, and following up patients under routine conditions. Where the economic evaluation is based on a less naturalistic trial design – for example, an explanatory trial designed primarily to address safety and efficacy questions – efforts should be made to increase the generalizability of the study findings. This might be achieved by, for example, relaxing stringent inclusion criteria [10], factoring

out the effects of protocol-driven costs in the final calculus [11] or by agreeing on the incorporation of a separate ‘usual-care’ arm into the trial [12].

Methods for estimating the appropriate sample size and statistical power for economic end points in RCTs have been developed on the basis of Fieller’s theorem (a statistical approach that allows the calculation of a confidence interval for the ratio of two means), the nonparametric bootstrap (the practice of estimating properties of an estimator [such as its variance] by measuring those properties when sampling with replacement from the observed data) and Bayes’ theorem (an approach that incorporates prior knowledge on parameters through specification of prior distributions at the design and analysis stage) [13]. Notably, however, sample size and statistical power calculations for RCTs have typically been based on the primary clinical outcomes alone. This is due, in part, to the complexities surrounding the estimation of the joint distribution of the difference in costs and consequences between the trial arms. It may also be due to the generally larger sample size requirements for satisfying economic end points, because of the large variability in healthcare resource use and cost measures [14]. The latter issue may mean that it is neither financially nor ethically feasible to conduct RCTs large enough to detect statistically significant differences in economic end points. In the absence of formal sample sizes for trial-based economic evaluations, health economists generally focus on estimation rather than hypothesis testing of economic end points [14]. This offers an advantage in that many health economists consider the traditional rules of statistical inference surrounding a single parameter, such as clinical effectiveness or cost, to be arbitrary, and may result in inferior healthcare outcomes compared with basing decisions on expected cost-effectiveness [15].

### Measurement & valuation of data inputs

#### ■ Measurement of resource use

Trial-based economic evaluations entail the collection of patient-level resource utilization data, for example, duration of hospital stays and types and quantities of community services, over the follow-up period of the trial. The perspective of the analysis affects the categories of resource use that are included in the study. The perspective of an economic evaluation usually falls into one of the following three categories: healthcare system, public sector or societal. The perspective adopted by an economic evaluation should ideally be informed by national methodological guidance (in England and Wales, for example, NICE recommends including

NHS and personal social services as a minimum [2]), as well as consideration of the clinical context of the evaluation and an assessment of where the foregone benefits are likely to fall [16]. Programs or interventions in some contexts are likely to result in resource consequences beyond the health sector. Many neonatal interventions, for example, are likely to have resource consequences for several sectors of the economy, as well as for individuals. Low-birth-weight babies being studied within a trial context may require support from social service departments upon their discharge from hospital. The parents of sick neonates may have to forego other productive activities (paid or unpaid work) in order to spend time with them; their transport costs to and from the neonatal unit may be considerable, and care for other children may have to be arranged. In contexts such as this, there would be considerable value in also adopting a broader societal perspective, at least as part of a sensitivity analysis.

The main items of resource utilization can normally be incorporated into the trial case report forms with little additional burden to the trial. However, trial-based economic evaluations sometimes require collection of additional resource-use data from other sources, for example, medical records, interviews with health professionals or separate patient questionnaires and diaries [17]. A recent trial-based economic evaluation of acupuncture care as an adjunct to exercise-based physical therapy for osteoarthritis of the knee largely relied on self-reported resource utilization in postal questionnaires completed at 6 weeks, and 6 and 12 months post randomization [18].

A particular problem that arises when asking patients or carers to complete resource-use questionnaires is deciding on the optimum recall period. Two types of recall error can be distinguished; simply forgetting an entire episode or incorrectly recalling when it occurred. If a study requires information for a specific period of time, for example if there is reason to believe that an intervention may influence the number of GP visits immediately following a hospital procedure, then a shorter recall period of a few weeks could be reasonable. However, if a study is trying to develop a picture of resource use over a much longer period, such as 6 months, then using a short recall period to provide a snapshot of typical resource use may be insufficient and misleading. Analysts therefore have to contend with a trade-off between recall bias and complete sampling information [19].

#### ■ Valuation of resource use

The total cost for an individual patient participating in a trial can be expressed as:

$$C_i = \sum_j Q_{ij} : UC_j$$

where  $C_i$  represents the cost for patient  $i$ ,  $Q_{ij}$  represents the quantity of resource item  $j$  by patient  $i$ , and  $UC_j$  represents the unit cost of resource item  $j$ . This requires the estimation of unit costs for each element of resource use consumed by the patient. Theoretically, unit costs should be based on the economic notion of opportunity cost, which represents the value of the resource in its most highly valued alternative use [1]. In the absence of competitive health markets, however, nationally representative healthcare tariffs, such as the Payment by Results tariffs [102], NHS reference costs for clinically similar treatments [103] and the compendia of unit costs covering hospital and community health and social care services [104], in England, are assumed to approximate to opportunity costs. In jurisdictions with systems of billing and fee-for-service payment of providers, market prices are deflated using cost-to-charge ratios to more accurately reflect opportunity costs [20]. There may be circumstances where unit-cost estimates for health resources are not readily available and have to be generated from first principles at the trial centers using accounting studies. These accounting studies may themselves use a number of methods including time and motion studies, diary methods, work sampling, interviews with key caregivers, case note analysis and analyses of patient-activity databases. Note that in trial-based economic evaluations, unit costs for each resource item ( $UC_j$ ) tend to be standardized across patients and trial participating centers. Use of unstandardized unit costs may be considered appropriate when the relative prices of factors that contribute towards costs, such as labor and equipment, vary between trial centers [21] and, in the case of multinational trials, between countries [22]. All costs should be valued at the same price date with healthcare-specific inflation indices available for the task [104]. Economic evaluations based on multinational trials should, where appropriate, convert costs into a common currency. Purchasing power parity adjustments are recommended for such conversions [105]. There may be circumstances, however, where country-specific unit costs are appropriately attached to country-specific resource-use values and the resulting economic estimates reported separately for each country.

#### ■ Measurement & valuation of outcomes

Outcome measures incorporated into trial-based economic evaluations range from biomedical markers for final health end points, for example, bone-mineral density as a maker for fracture outcomes, to

intermediate health outcomes, for example, hospital episodes avoided, to more final health outcomes, for example, life years gained [10]. However, many health economists, and indeed for many reimbursement agencies, recommend the quality-adjusted life year (QALY) as the primary measure of health consequence for trial-based economic evaluations. The QALY is a preference-based measure of health outcome that combines length of life and health-related quality of life in a single metric [23]. For reimbursement agencies, the QALY has the advantage of allowing cost-effectiveness comparisons to be made across different healthcare interventions for disparate health conditions. For health economists, it offers an additional advantage in that the techniques used to derive the health-related quality of life component of the QALY generate values that reflect, to varying degrees, people's preferences for health outcomes. It is worth noting, however, that less than 25% of economic evaluations included in the NHS Economic Evaluation Database have measured outcomes in terms of QALYs [101].

In order to estimate QALYs over the time horizon of a trial, either a multi-attribute utility measure or a direct preference elicitation technique is completed by patients at different time points. The preference-based health-related quality of life scores (or utility scores) that they generate are combined with survival data to generate QALY profiles. Multi-attribute utility measures are essentially generic health-related quality of life instruments with pre-existing preference weights that can be attached to each permutation of responses. The most commonly used multi-attribute utility measures are the EQ-5D [24], Health Utilities Index [25], SF-6D [26], Quality of Well-Being Scale [27] and Assessment of Quality of Life [28]. The underpinning preference weights for these measures are generally drawn from surveys of the general population. The alternative approach of using a direct preference elicitation technique, such as the standard gamble approach or time trade-off approach [29], is more expensive and time consuming and relies on patients to not only describe their health status, but also to value it using a complex scaling technique. For both approaches, the frequency and timing of assessments should be influenced by the disease severity of the study population, study duration, timing of trial visits and patient burden [30]. Proxy measurements may be considered when patients are either too ill or do not have the cognitive competencies to complete the measures [31].

There are several circumstances where the QALY metric itself is too restrictive and does not

capture the main outcomes of interest within a trial. In response to this, researchers are developing instruments based on Armatya Sen's Capability Theory that try to measure broader outcomes such as attachment, security, enjoyment, role and control for incorporation within a trial-based economic evaluation framework [32]. Discrete choice experiments have also been developed that describe healthcare interventions in terms of their attributes, including health outcomes, nonhealth outcomes and process attributes [33], although their incorporation into trial-based economic evaluations is still at an embryonic stage.

---

### Analysis of data inputs

Where trials require the measurement and valuation of costs and outcomes over several years of patient follow up, the costs and outcomes that occur after the first year of follow up are typically reduced by a discount factor to their equivalent present values. The discounted amounts, the equivalent present values, are totalled allowing all interventions to be fairly compared on the basis of their present values. The discount rates applied within trial-based economic evaluations have varied markedly across jurisdictions and over time. NICE has recommended that economic evaluations conducted in England and Wales should discount both costs and outcomes at an annual rate of 3.5% [2]. There is recognition, however, that the discounting of health outcomes is controversial. Furthermore, the empirical evidence on whether individuals discount future health in practice is disputed [34]. In the light of this, sensitivity analyses that apply differential discount rates to costs and outcomes are acceptable [2].

Further analyzing of data inputs into a trial-based economic evaluation are often required before estimation of cost-effectiveness can be made. For example it is not unusual for cost data inputs to be truncated at zero and right skewed due to the impossibility of incurring costs less than zero and the small numbers of high resource-use patients. This means that cost data do not usually conform to the assumptions of standard statistical tests for comparing differences in arithmetic means (the crucial cost variable for policy makers). Popular methods for dealing with this problem include using the nonparametric bootstrap method as the primary statistical test for making inferences about arithmetic means or generalized linear models to directly model the mean of the cost distribution [35]. Nevertheless, there are circumstances, particularly in very large samples where the near-normality of sample means is assured, where simple approaches

for analyzing cost data may be sufficient [36]. For QALYs, adjustments are often made to account for baseline differences in health status between the trial groups [37].

Another analytical challenge is how to deal with missing data that arises for some variables or patients. A recent paper by Sterne and colleagues highlights three missing data mechanisms:

- Missing completely at random – where missing data bear no relation to the value of any other factor in the study population;
- Missing at random – where missing data are correlated in an observable way with the mechanism that generates the outcome;
- Not missing at random – where missing values depend on unobserved variables [38].

The authors' summary of different approaches for handling missing data, including multiple imputations, can be applied to the data inputs for economic evaluation [38]. Censoring represents a particular form of missing data and arises where information on the costs and outcomes of some patients is truncated and not available for the full follow-up period of the trial. In the past, this problem was frequently ignored, or analysts deployed simple methods such as complete case analysis and available case analysis. Recently, the Kaplan Meier Sample Average estimator (an estimator that measures the survival function by computing the probabilities of occurrence of events at a certain point of time and multiplying these successive probabilities by any earlier computed probabilities to get the final estimate) and the Inverse Probability Weighting estimator (an estimator that in addition accounts for attrition through covariate adjustment) have become popular methods for dealing with censored data [36].

### Comparative analysis of costs & consequences

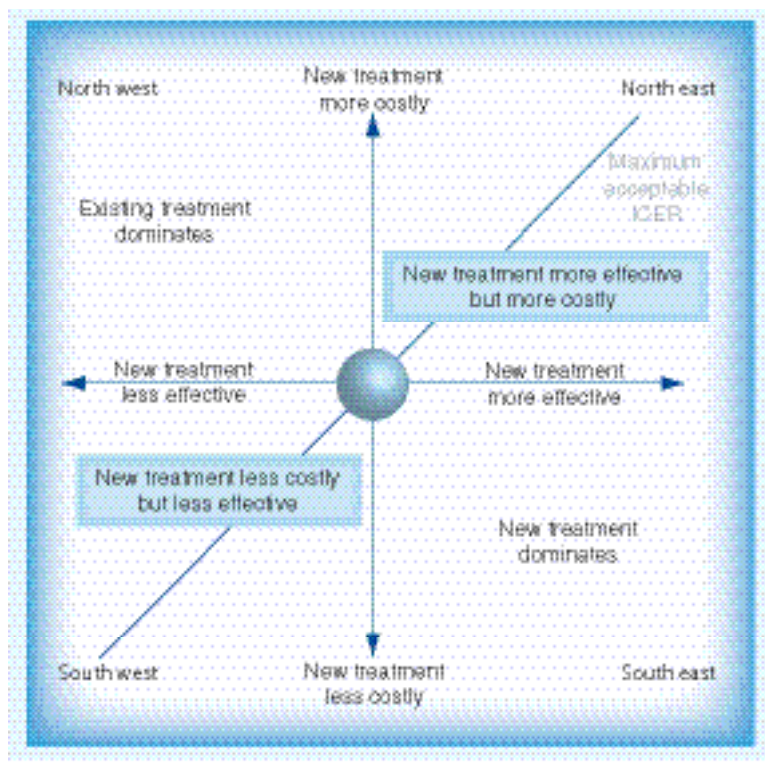
Economic evaluation synthesizes evidence on costs and consequences within an explicit framework, enabling decision makers to assess whether a new healthcare program or intervention offers good 'value for money'. Data on either costs or consequences, when viewed in isolation, do not provide decision makers with the information required for 'value-for-money' assessments. The simplest form of trial-based economic evaluation is cost-minimization analysis, which seeks to establish the least costly method of achieving given outcomes. It is only appropriate if all outcomes are found to be identical,

and the underpinning trial is sized around a safety or equivalence hypothesis [39]. In cost-minimization analysis, the decision rule for decision makers is straightforward. If health effects are equivalent, then the cheaper programme or intervention is preferable. Theoretically, trial-based economic evaluations can also take the form of cost-benefit analyses where the consequences of programs or interventions are measured and valued in monetary units, either by asking relevant individuals how much they would be willing to pay to obtain the observed consequences of the programme or intervention (contingent valuation), or by asking relevant individuals how much they would be willing to trade between observed health outcomes, nonhealth outcomes and process attributes and subsequently converting responses to willing to pay estimates for unit changes in attributes (discrete choice experiments). As with cost-minimization analysis, the decision rule for decision-makers is straightforward. If the monetary valuation of the consequences of the healthcare program or intervention exceeds its net costs, then the program or intervention should be provided since there is a net gain to society. Most trial-based economic evaluations, however, take the form of cost-effectiveness analyses (where consequences are measured in natural or physical units) or cost-utility analysis (where consequences are measured in terms of preference-based metrics such as QALYs). Here, the decision rule becomes rather more complex, since costs and consequences are expressed in different metrics.

For trial-based economic evaluations that take the form of cost-effectiveness analyses or cost utility analyses, costs and consequences can be averaged across all patients in the treatment (t) or the control (c) group to obtain mean cost (C) and mean effect (E) for each arm of the trial. The results are typically reported in terms of an incremental cost-effectiveness ratio (ICER), which is the difference in costs divided by the difference in effects:

$$\text{ICER} = \frac{C_t - C_c}{E_t - E_c} = \frac{DC}{DE}$$

The difference in costs and the difference in effects between the trial groups can also be depicted on the cost-effectiveness plane (Figure 1) [10,35]. Incremental effectiveness (relative to the comparator) is shown on the X-axis, while the Y-axis shows the incremental cost, and the origin of the graph (C) represents the point of comparison or control. In comparison with this central point, the intervention of interest can therefore be more effective or less effective, and can be more costly or less costly, and these combinations are represented by the four quadrants of



**Figure 1. The cost-effectiveness plane.**

ICER: Incremental Cost-Effectiveness Ratio.

Reproduced with permission from [35] Oxford University Press (2010).

the figure. If the new intervention is found to be less costly and more effective, the ICER will fall in the south-east quadrant and decision makers will have no difficulty opting to adopt it (assuming, at this stage, that no uncertainty surrounds the ICER); the new intervention can be said to dominate the comparator. If it turns out to be less effective and more costly, the ICER will fall in the north-west quadrant and a decision to reject can equally readily be made; in this case the new intervention is dominated by the comparator. More interesting and typically more common situations arise in the north-east and south-west quadrants, where the new intervention is more effective but also more costly (the north-east quadrant), or is less effective but also less costly (the south-west quadrant). In these areas of the figure, there is a trade-off between effect and cost: additional health benefit can be obtained but at higher cost (north east), or savings can be made but only by surrendering some health benefit (south west). In order to assess whether these trade-offs are acceptable, a maximum acceptable ICER or a maximum willingness to pay for a unit of effect ( $\lambda$ ) is required. The dashed diagonal line running through [Figure 1](#) depicts one possible maximum willingness to pay for a unit of effect ( $\lambda$ ). ICERs falling to the

right of this line would be considered cost-effective, whilst ICERs falling to the left would not. Note that a steeper diagonal line reflects a greater willingness to pay by decision makers for a unit of effect. There are a number of revealed and stated preference techniques for estimating the value of  $\lambda$  [35]. In many jurisdictions, however, the decision rules surrounding the value of  $\lambda$  have evolved historically and with little scientific basis [40,41].

### Handling uncertainty

A number of different types of uncertainty can arise in trial-based economic evaluations. Sampling (or stochastic) uncertainty depends on variation in both the numerator (incremental cost) and the denominator (incremental effectiveness) of the ICER. A common method for estimating confidence intervals for the ICER is the nonparametric bootstrap, which resamples with replacement cost-effect pairs from the trial data under the assumption that the trial population is a valid representation of the underlying population of interest [42]. The bootstrap replicates are plotted on the cost-effectiveness plane as a scatter of points representing sampling uncertainty. However, there are several circumstances where a meaningful ordering of the bootstrapped replicates, required to make the confidence interval surrounding the ICER interpretable, is very difficult. For example, a negative ICER might represent improved outcomes and lower costs (south-east quadrant of the cost-effectiveness plane) or worse outcomes and higher costs (north-west quadrant), two qualitatively different scenarios. One way of handling this problem is to place both costs and effects on a linear scale, either net monetary benefit, defined as  $\Delta E \times \lambda - \Delta C$ , or net health benefit, defined as  $\Delta E - \Delta C/\lambda$  [43,44]. Here, analysts need not worry about ambiguous interpretations of positive or negative ICERs; larger net monetary benefits or net health benefits are unambiguously better and smaller are unambiguously worse. Decision uncertainty can be dealt with by constructing a cost-effectiveness acceptability curve, which shows the probability that the new intervention is cost-effective across a range of values of  $\lambda$  [45]. For each value of  $\lambda$ , the cost-effectiveness acceptability curve shows the proportion of bootstrap replicates that fall to the right of the diagonal line running through the cost-effectiveness plane. Heterogeneity may be important if particular groups that differ with respect to observed or unobserved characteristics, such as age or sex, differ systematically in ways that affect the results of the economic evaluation; for example, through their treatment costs or their capacity to benefit from an intervention

[46]. If heterogeneity is important, efforts should be made to report differences in costs, outcomes or cost-effectiveness that can be explained by variations between subgroups of patients. Simply dividing the study population into different subgroups and estimating costs, outcomes or cost-effectiveness for each of those subgroups may reduce the power to detect significant differences between groups. For trial-based economic evaluations, regression modeling of the data may disentangle differences in costs, outcomes or cost-effectiveness that can be explained by variations between subgroups of patients. The study by Mihaylova *et al.* provides a template [46]. Recognizing the absence of evidence of heterogeneity in treatment effect across subgroups in the Heart Protection Study, these authors applied the same trial-wide relative-risk reduction to different subgroups defined in terms of absolute-risk levels at baseline, resulting in large but reliable differences in cost-effectiveness. Finally, methodological uncertainty, the uncertainty surrounding the value of key parameters in the economic evaluation, for example the discount rate, can be dealt with using sensitivity analysis.

### Extrapolation of cost-effectiveness

There are several circumstances where cost-effectiveness observed within a trial is substantially different to what would have been observed with continued patient follow up. Consequently, extrapolation of cost-effectiveness over an extended time horizon, often a lifetime horizon, is generally considered important. Simple extrapolation methods are sometimes used, such as adding remaining life expectancy at the end of a trial using life-table data. However, unbiased estimation of long-term cost-effectiveness may require complex modeling of epidemiological, clinical and economic variables. The individual-level data available within trials often permit the construction of these complex models that allow estimates of lifetime costs, utilities and cost-effectiveness to be made. A good example is provided by the long-term economic evaluation based on the UK Prospective Diabetes Study [47].

### Future perspective

Economic evaluation provides a means of allocating finite healthcare resources in an efficient manner. It can inform the decision-making processes at many levels, from national decision-making bodies, such as NICE, to decisions by local healthcare providers. A number of methodological advances in economic evaluation have been made in the last two decades surrounding sample size estimation, methods for

collecting data inputs, methods for valuation of health benefits, analysis of skewed, missing or censored data, ascertainment of the maximum willingness to pay for health benefits, handling of different forms of uncertainty and extrapolation of survival and health-related quality of life gains beyond the period of observation.

However, it is also important to mention some of the controversies surrounding economic evaluation that remain. Although health economists agree about the objectives of economic evaluation, they disagree about a number of methodological issues. For example, some health economists argue that cost-benefit analysis is the optimal form of economic evaluation because of its foundation in welfare economic theory, whilst others promote cost-utility analysis because of its broader acceptance by the research community. Some health economists argue that economic evaluation of health interventions should be limited to a healthcare system perspective, whilst others promote a broader societal perspective. There are also a number of methodological concerns that are common to economic evaluations based on RCTs: a single RCT might not compare all the relevant options available, might not provide evidence on all relevant inputs into an economic evaluation, might not be conducted over a long enough time horizon to capture differences in economic outcomes or even measure those outcomes, or might not provide evidence specific to a particular setting or patient group [6]. In addition, reliance on a single RCT as a vehicle for an economic evaluation may mean ignoring evidence from other trials, meta-analyses, and observational studies. Under these circumstances, an alternative framework for the conduct of economic evaluation is provided by decision analytic modeling.

Nevertheless, trial-based economic evaluations are likely to continue to provide an important strand of evidence that facilitates evidence-based decision-making. Trial-based economic evaluations will continue to offer a number of strengths and unique opportunities. In addition to producing unbiased estimates of effect, trials make available to researchers a wealth of data on individual participants: their characteristics and history at entry; their risk factors at entry and often over time; the frequency, timing and sequence of end points; the reliable clinical ascertainment of these end points; patient treatments, co-medications and adherence; their resource use; and their health-related quality of life. Since these are available for the same individuals over time, the full structure of covariance can be estimated.

Moreover, a strong move in recent years towards

Executive summary
<p><b>Background</b></p> <ul style="list-style-type: none"> <li>Economic evaluation provides a framework for assessing the costs and consequences of alternative healthcare programs or interventions.</li> <li>A common vehicle for the conduct of economic evaluation is the randomized controlled trial.</li> </ul>
<p><b>Design of trial-based economic evaluations</b></p> <ul style="list-style-type: none"> <li>Designing a rigorous trial-based economic evaluation requires multidisciplinary input. An accredited clinical trials unit should be involved in the process.</li> <li>Sample-size calculations for randomized clinical trials should consider economic end points.</li> </ul>
<p><b>Measurement of resource use</b></p> <ul style="list-style-type: none"> <li>The main items of resource utilization can normally be incorporated into trial case report forms with little additional burden to the trial. However, these data may have to be supplemented with data from other sources.</li> <li>Trade-off between recall bias and complete sampling information are common for resource utilization data.</li> </ul>
<p><b>Valuation of resource use</b></p> <ul style="list-style-type: none"> <li>The valuation of resource inputs should be underpinned by rigorous accounting procedures.</li> </ul>
<p><b>Measurement &amp; valuation of outcomes</b></p> <ul style="list-style-type: none"> <li>Analysts should consider approaches to outcomes measurement that are rooted in economic theory.</li> </ul>
<p><b>Analysis of data inputs</b></p> <ul style="list-style-type: none"> <li>Advanced statistical or econometric approaches may be required to handle skewed, missing and censored data inputs into trial-based economic evaluations.</li> </ul>
<p><b>Comparative analysis of costs &amp; consequences</b></p> <ul style="list-style-type: none"> <li>Trial-based economic evaluation aim to assess the incremental cost–effectiveness of a new health care program or intervention, and where the incremental cost–effectiveness ratio lies on the cost–effectiveness plane.</li> </ul>
<p><b>Handling uncertainty</b></p> <ul style="list-style-type: none"> <li>Attempts should be made to address the uncertainty that can arise in trial-based economic evaluations, including sampling (or stochastic) uncertainty, decision uncertainty, heterogeneity and methodological uncertainty.</li> </ul>
<p><b>Extrapolation of cost–effectiveness</b></p> <ul style="list-style-type: none"> <li>Extrapolation of cost–effectiveness beyond the follow-up period of the trial is often required.</li> </ul>
<p><b>Future perspective</b></p> <ul style="list-style-type: none"> <li>Trial-based economic evaluations are complements to, rather than substitutes for, economic evaluations that use decision analytic modeling.</li> </ul>

the use of randomized studies in other areas such as development economics is worth noting [48], although here, as in health economics, it is clear that it is not appropriate in all circumstances and needs to be seen alongside other designs [49]. Lack of space precludes a description of reporting guidelines for economic evaluation. The development of new reporting guidelines for health economic evaluation is the current line of enquiry for an International Society for Pharmacoeconomics and Outcomes Research task force that hopes to publish its results in the form of a new checklist in 2013.

### Financial & competing interests disclosure

The author has no relevant affiliations or

financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

### References

Papers of special note have been highlighted as:

- of interest

1 Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart G. *Methods for the Economic Evaluation of Health Care Programs, 3<sup>rd</sup> Edition*. Oxford University Press, Oxford, UK (2005).

- Seminal text on methods of economic evaluation.

2 National Institute for Health and Clinical

Excellence (NICE). *NICE Guide to the Methods of Technology Appraisal*. NICE, London, UK (2008).

- Technical guidance on the methods that should be applied within economic evaluation.

3 Scottish Medicines Consortium (SMC). *Guidance to Manufacturers for Completion of New Product Assessment Form (NPAF)*. SMC, Glasgow, UK (2007).

4 Pharmaceutical Benefits Advisory Committee (PBAC). *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee*. PBAC, Canberra, Australia (2002).

5 CDR Directorate at Canadian Coordinating Office for Health Technology Assessment (CCOHTA). *Canada Common Drug Review Submission Guidelines for manufacturers*. CCOHTA, Ottawa, Canada (2003).



- 6 Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for healthcare decision making? *Health Econ.* 15(7), 677–687 (2006).
- **Highlights the limitations of trial-based economic evaluations and emphasizes the relevance of modeling-based economic evaluations.**
- 7 Buxton MJ, Drummond MF, Van HB *et al.* Modelling in economic evaluation: an unavoidable fact of life. *Health Econ.* 6(3), 217–227 (1997).
- 8 Edwards RT, Hounsome B, Linck P, Russell IT. Economic evaluation alongside pragmatic randomised trials: developing a standard operating procedure for clinical trials units. *Trials* 9, 64 (2008).
- 9 McCrone P, Marks IM, Mataix-Cols D, Kenwright M, McDonough M. Computer-aided self-exposure therapy for phobia/panic disorder: a pilot economic evaluation. *Cogn. Behav. Ther.* 38(2), 91–99 (2009).
- 10 Glick HA, Joshi JA, Sonnad SS, Polsky D. *Economic Evaluation in Clinical Trials.* Oxford University Press, Oxford, UK (2007).
- 11 Coyle D, Lee LM. The problem of protocol driven costs in pharmacoeconomic analysis. *Pharmacoeconomics* 14, 357–363 (1998).
- 12 O’Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? *Value Health* 8(1), 67–79 (2005).
- 13 Gardiner JC, Sirbu CM, Rahbar MH. Updated on statistical power and sample size assessments for cost–effectiveness studies. *Expert Rev. Pharmacoeconomics Outcomes Res.* 4(1), 89–98 (2004).
- 14 Briggs A. Economic evaluation and clinical trials: size matters. *BMJ* 321, 1362 (2000).
- 15 Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of healthcare technologies. *J. Health Econ.* 18, 341–364 (1999).
- 16 Claxton K, Walker S, Palmer S, Sculpher M. Appropriate perspectives for healthcare decisions. In: *CHE Research Paper 54; Centre for Health Economics.* University of York, York, UK (2010).
- 17 Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. *Int. J. Technol. Assess.* 21(2), 165–171 (2005).
- 18 Whitehurst DG, Bryan S, Hay EM, Thomas E, Young J, Foster NE. Cost–effectiveness of acupuncture care as an adjunct to exercise-based physical therapy for osteoarthritis of the knee. *Phys. Ther.* 91(5), 630–641 (2011).
- 19 Clarke PM, Fiebig DG, Gerdtham UG. Optimal recall length in survey design. *J. Health Econ.* 27(5), 1275–1284 (2008).
- 20 Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost–effectiveness in Health and Medicine.* Oxford University Press, New York, NY, USA (1996).
- 21 Raikou M, Briggs A, Gray A, McGuire A. Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. *Health Econ.* 9(3), 191–198 (2000).
- 22 Torti FM, Reed SD, Schulman KA. Analytic considerations in economic evaluations of multinational cardiovascular clinical trials. *Value Health* 9(5), 281–291 (2006).
- 23 Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int. J. Technol. Assess. Health Care.* 5, 559–575 (1989).
- 24 Brooks R. EQ-5D, the current state of play. *Health. Pol.* 37, 53–72 (1996).
- **Introductory text on the EQ-5D, a standardized instrument for quality-adjusted life year measurement.**
- 25 Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions: health utilities index. *Pharmaco. Econ.* 7, 503–520 (1995).
- 26 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* 21(2), 271–292 (2002).
- 27 Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv. Res.* 11, 478–507 (1976).
- 28 Richardson J, Day N, Peacock S, Iezzi A. Measurement of the quality of life for economic evaluation and the Assessment of Quality of Life (AQoL) mark 2 instrument. *Aust. Econ. Rev.* 37(1), 62–88 (2004).
- 29 Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J. Health Econ.* 15, 209–231 (1996).
- 30 Ramsey S, Willke R, Briggs A *et al.* Good research practices for cost–effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA task force report. *Value Health* 8(5), 521–533 (2005).
- **Good research practice guidelines for trial-based economic evaluations.**
- 31 Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation.* Oxford University Press, Oxford, UK (2007).
- 32 Coast J, Flynn TN, Natarajan L *et al.* Valuing the ICECAP capability index for older people. *Soc. Sci. Med.* 67(5), 874–882 (2008).
- 33 de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ.* 21(2), 145–172 (2012).
- 34 Brouwer WBF, van Exel NJA. Discounting in decision making: the consistency argument revisited empirically. *Health Pol.* 67, 187–194 (2004).
- 35 Gray A, Clarke P, Wolstenholme J, Wordsworth S. *Applied Methods of Cost–effectiveness Analysis in Health Care.* Oxford University Press, Oxford, UK (2010).
- 36 Mihaylova B, Briggs A, O’Hagan A, Thompson S. Review of statistical methods for analysing healthcare resources and costs. *Health Econ.* 20(8), 897–916 (2011).
- 37 Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost–effectiveness analysis: the importance of controlling for baseline utility. *Health Econ.* 14(5), 487–496 (2005).
- 38 Sterne JAC, White IR, Carlin JB *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338, b2393 (2009).
- 39 Briggs AH, O’Brien BJ. The death of cost-minimization analysis. *Health Econ.* 10, 179–184 (2001).
- 40 Devlin N, Parkin D. Does NICE have a cost–effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ.* 13, 437–452 (2004).
- 41 Grosse S. Assessing cost–effectiveness in healthcare: history of the \$50,000 per QALY threshold. *Expert Rev. Pharmacoecon. Outcomes Res.* 8(2), 165–178 (2008).
- 42 Briggs AH, Wonderling DE, Mooney CZ. Pulling cost–effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ.* 6(4), 327–340 (1997).
- 43 Stinnett AA, Mullaly J. Net health benefits: a new framework for the analysis of uncertainty in cost–effectiveness analysis. *Med. Decis. Making* 18(Suppl. 2), S68–S80

- (1998).
- 44 Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *Int. J. Technol. Assess. Health Care* 14(3), 467–471 (1998).
- 45 Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves: facts, fallacies and frequently asked questions. *Health Econ.* 13, 405–415 (2004).
- **Introductory text on the use of cost-effectiveness acceptability curves for handling decision uncertainty.**
- 46 Mihaylova B, Briggs A, Armitage J, Parish S, Gray A, Collins R. Cost-effectiveness of simvastatin in people at different levels of vascular disease risk: economic analysis of a randomised trial in 20,536 individuals. *Lancet* 365 (9473), 1779–1785 (2005).
- **Template for disentangling cost-effectiveness by subgroup.**
- 47 UKPDS Group. Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes: UKPDS 40. *BMJ* 317, 720–726 (1998).
- 48 Duflo E, Kremer M. Use of randomization in the evaluation of development effectiveness. In: *Evaluating Development Effectiveness (World Bank Series on Development and Evaluation, Volume 7*. Feinstein O, Ingram GK, Pitman GK (Ed.). Transaction Publishers, Piscataway, NJ, USA, 205–232, (2004).
- 49 Deaton A. Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development. In: *NBER Working Paper 14690*. National bureau of economic research, Inc., MA, USA (2009).
- **Websites**
- 101 Centre for Reviews and Dissemination database. [www.york.ac.uk/inst/crd](http://www.york.ac.uk/inst/crd)
- 102 Department of Health, England. Payment by results. Department of Health, London, UK (2010). [www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/DH\\_900](http://www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/DH_900)
- 103 Department of Health, England. NHS Reference Costs 2010. Department of Health, London, UK (2010). [www.dh.gov.uk/en/Publicationsandstatistics](http://www.dh.gov.uk/en/Publicationsandstatistics)
- 104 Curtis L. Unit Costs of Health and Social Care 2010. Personal social services research unit, University of Kent, Canterbury, UK (2010). [www.pssru.ac.uk](http://www.pssru.ac.uk)
- 105 Global Purchasing Power Parities and Real Expenditures 2005. World Bank International Comparison Program, Washington, DC, USA (2009). <http://siteresources.worldbank.org/ICPINT/Resources/icp-final.pdf>