

# Good Research Practices for Cost-Effectiveness Analysis Alongside Clinical Trials: The ISPOR RCT-CEA Task Force Report

Scott Ramsey, MD, PhD (cochair),<sup>1</sup> Richard Willke, PhD (cochair),<sup>2</sup> Andrew Briggs, DPhil,<sup>3</sup> Ruth Brown, MS,<sup>4</sup> Martin Buxton, PhD,<sup>5</sup> Anita Chawla, PhD,<sup>6</sup> John Cook, PhD,<sup>7</sup> Henry Glick, PhD,<sup>8</sup> Bengt Liljas, PhD,<sup>9</sup> Diana Petitti, MD,<sup>10</sup> Shelby Reed, PhD<sup>11</sup>

<sup>1</sup>Fred Hutchinson Cancer Research Center, Seattle, WA, USA; <sup>2</sup>Pfizer, Inc., Bridgewater, NJ, USA; <sup>3</sup>University of Oxford, Oxford, UK; <sup>4</sup>MEDTAP International, London, UK; <sup>5</sup>Brunel University, Uxbridge, Middlesex, UK; <sup>6</sup>Genentech, San Francisco, CA, USA; <sup>7</sup>Merck & Co., Inc, Blue Bell, PA, USA; <sup>8</sup>University of Pennsylvania, Philadelphia, PA, USA; <sup>9</sup>AstraZeneca, Lund, Sweden; <sup>10</sup>Kaiser Permanente, Pasadena, CA, USA; <sup>11</sup>Duke Clinical Research Institute, Durham, NC, USA

## ABSTRACT

**Objectives:** A growing number of prospective clinical trials include economic end points. Recognizing the variation in methodology and reporting of these studies, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) chartered the Task Force on Good Research Practices: Randomized Clinical Trials—Cost-Effectiveness Analysis. Its goal was to develop a guidance document for designing, conducting, and reporting cost-effectiveness analyses conducted as a part of clinical trials.

**Methods:** Task force cochairs were selected by the ISPOR Board of Directors. Cochairs invited panel members to participate. Panel members included representatives from academia, the pharmaceutical industry, and health insurance plans. An outline and a draft report developed by the panel were presented at the 2004 International and European ISPOR meetings, respectively. The manuscript was then submitted to a reference group for review and comment.

**Results:** The report addresses issues related to trial design, selecting data elements, database design and man-

agement, analysis, and reporting of results. Task force members agreed that trials should be designed to evaluate effectiveness (rather than efficacy), should include clinical outcome measures, and should obtain health resource use and health state utilities directly from study subjects. Collection of economic data should be fully integrated into the study. Analyses should be guided by an analysis plan and hypotheses. An incremental analysis should be conducted with an intention-to-treat approach. Uncertainty should be characterized. Manuscripts should adhere to established standards for reporting results of cost-effectiveness analyses.

**Conclusions:** Trial-based cost-effectiveness studies have appeal because of their high internal validity and timeliness. Improving the quality and uniformity of these studies will increase their value to decision makers who consider evidence of economic value along with clinical efficacy when making resource allocation decisions.

**Keywords:** cost-effectiveness, economic, guidelines, randomized clinical trial.

## Introduction

Clinical trials evaluating medicines, medical devices, and procedures now commonly assess economic value of these interventions. The growing number of prospective clinical/economic trials reflects both widespread interest in economic information for new technologies and the regulatory and

reimbursement requirements of many countries that now consider evidence of economic value along with clinical efficacy. In recent years, research has also improved the methods for the design, conduct, and analysis of economic data collected alongside clinical trials. Despite these advances, the literature reveals a great deal of variation in methodology and reporting of these studies. Improving the quality of these studies will enhance the credibility and usefulness of cost-effectiveness analyses to decision makers worldwide.

To foster improvements in the conduct and reporting of trial-based economic analyses, the

*Address correspondence to:* Scott Ramsey, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, PO Box 19024, Seattle, WA 98109, USA. E-mail: [sramsey@fhcrc.org](mailto:sramsey@fhcrc.org)

10.1111/j.1524-4733.2005.00045.x

International Society for Pharmacoeconomics and Outcomes Research (ISPOR) chartered the ISPOR Task Force on Good Research Practices: Randomized Clinical Trials-Cost-Effectiveness Analysis (RCT-CEA). The task force cochairs were selected by the ISPOR Board of Directors, and the cochairs invited the other panel members to participate. The panel was first assembled in January 2004, communicated monthly, and agreed on an outline and preliminary content that was presented for comment by ISPOR membership at the May 2004 annual meeting. The draft report was then written and presented to the ISPOR membership at the October 2004 European meeting. A volunteer reference group of ISPOR members provided valuable comments on the draft report, which supported the completion of the final report in February 2005.

The purpose of this report is to state the consensus position of this Task Force (Table 1). The goal for the panel was to develop a guidance document for the design, conduct, and reporting of cost-effectiveness analyses alongside clinical trials. The intended audiences are researchers in academics, industry, and government who design and implement these studies, decision makers who evaluate clinical and economic evidence for formulary and insurance coverage policies, and students of the area. The panel recognizes that advances in methodology for joint clinical/economic analyses will continue and that clinical/economic trials are heterogeneous in nature. Therefore, the report highlights areas of consensus, emerging methodologies with a diversity of professional opinions, and issues where further research and development are needed.

The focus of this report is cost-effectiveness analysis conducted alongside randomized clinical trials designed to test the efficacy or effectiveness for drugs, devices, surgical procedures, or screening interventions, including pragmatic trials. Clinical trials are artificial treatment environments and do not provide all the economic information needed by decision makers. Trial populations do not commonly reflect patient groups treated in clinical practice, and the time horizon for trials often does not reflect the duration of impact of the intervention. These issues are commonly addressed with modeling. The reader is referred to an earlier guidance document addressing these issues [1].

There are also some common issues in cost-effectiveness analysis that are fundamental to all studies of this nature that will not be addressed in this article. These include study perspective, choice of discount rate for costs and outcomes, type of analysis (e.g., cost-utility, cost-benefit), types of

costs that will be included (direct medical, nonmedical, etc.), and marginal versus average costing. These issues apply to all economic analyses, not just economic studies within clinical trials, and are well described in the literature.

### **Initial Trial Design Issues**

The quality of economic information that is derived from trials depends on the attributes of the trial's design. Those economic analyses often described as being conducted "alongside" clinical trials are indicative of an important practical design issue. Economic analysis is rarely the primary purpose of an experimental study. Nevertheless, it is important that the analyst contributes to the design of the study to ensure that the structure of the trial will provide the data necessary for a high quality economic study.

#### *Appropriate Trial Design*

The distinction between pragmatic and exploratory trials and the corresponding distinction between effectiveness and efficacy is well understood [2,3]. It is generally acknowledged that pragmatic effectiveness trials are the best vehicle for economic studies; however, it is usually necessary to undertake economic evaluations earlier in the development cycle where the focus is on efficacy, including phase III or even phase II drug trials, to provide timely information for pricing and reimbursement. Our report is meant to apply to both types of trials.

Large simple trials [4] are efficient for addressing clinical questions because they capture the main effects of treatments that have small to moderate impacts on large potential populations. They will also be efficient for answering economic questions for diseases or treatments where the bulk of costs derive from primary outcomes that are measured in the trial and for which the quality of life impacts are persistent, and thus can be measured infrequently.

An ideal follow-up period for an economic study is independent of the occurrence of clinical events, whether study-related or not. All patients should be followed for a common length of time or the full duration of the trial. Discontinuation of data collection because of a clinical event will fail to capture the important aspects of the disease under study: the adverse effects of the event on quality of life, resource use, and cost.

#### *Sample Size and Power*

In an ideal world, the economic appraisal would be factored into sample-size calculations using stand-

**Table 1** Core recommendations for conducting economic analyses alongside clinical trials*Trial design*

Trial design should reflect effectiveness rather than efficacy when possible.

Full follow-up of all patients is encouraged.

Describe power and ability to test hypotheses, given the trial sample size.

Clinical end points used in economic evaluations should be disaggregated.

Direct measures of outcome are preferred to use of intermediate end points.

*Data elements*

Obtain information to derive health state utilities directly from the study population.

Collect all resources that may substantially influence overall costs; these include those related and unrelated to the intervention.

*Database design and management*

Collection and management of the economic data should be fully integrated into the clinical data.

Consent forms should include wording permitting the collection of economic data, particularly when it will be gathered from third-party databases and may include pre- and/or post-trial records.

*Analysis*

The analysis of economic measures should be guided by a data analysis plan and hypotheses that are drafted prior to the onset of the study. All cost-effectiveness analyses should include the following: an intention-to-treat analysis; common time horizon(s) for accumulating costs and outcomes; a within-trial assessment of costs and outcomes; an assessment of uncertainty; a common discount rate applied to future costs and outcomes; an accounting for missing and/or censored data

Incremental costs and outcomes should be measured as differences in arithmetic means, with statistical testing accounting for issues specific to these data (e.g., skewness, mass at zero, censoring, construction of QALYs).

Imputation is desirable if there is a substantial amount of missing data. Censoring, if present, should also be addressed.

One or more summary measures should be used to characterize the relative value of the intervention.

Examples include ratio measures, difference measures, and probability measures (e.g., cost-effectiveness acceptability curves).

Uncertainty should be characterized. Account for uncertainty that stems from sampling, fixed parameters such as unit costs and the discount rate, and methods to address missing data.

Threats to external validity—including protocol-driven resource use, unrepresentative recruiting centers, restrictive inclusion and exclusion criteria, and artificially enhanced compliance—are best addressed at the design phase.

Multinational trials require special consideration to address intercountry differences in population characteristics and treatment patterns.

When models are used to estimate costs and outcomes beyond the time horizon of the trial, good modeling practices should be followed.

Models should reflect the expected duration of the intervention on costs and outcomes.

Subgroup analyses based on prespecified clinical and economic interactions, when found to be significant *ex post*, are appropriate. Ad hoc subgroup analysis is discouraged.

*Reporting the results*

Minimum reporting standards for cost-effectiveness analyses should be adhered to for those conducted alongside clinical trials.

The cost-effectiveness report should include a general description of the clinical trial and key clinical findings.

Reporting should distinguish economic data collected as part of the trial vs. data not collected as part of the trial.

The amount of missing data should be reported. If imputation methods are used, the method should be described.

Methods used to construct and compare costs and outcomes, and to project costs and outcomes beyond the trial period should be described.

The results section should include summaries of resource use, costs, and outcome measures, including point estimates and measures of uncertainty. Results should be reported for the time horizon of the trial, and for projections beyond the trial (if conducted).

Graphical displays are recommended for results not easily reported in tabular form (e.g., cost-effectiveness acceptability curves, joint density of incremental costs and outcomes).

QALYs, quality-adjusted life-years.

ard methods [5,6] based on asymptotic normality, or by simulation [7]. Nevertheless, it is common for the sample size of the trial to be based on the primary clinical outcomes alone. As a consequence, it is possible that the economic comparisons will be underpowered. Analysts should calculate the likely power of the study at the design stage to ascertain whether, given the proposed sample size, it makes sense to undertake an economic appraisal. In many cases, sample-size restrictions will necessitate focus on estimation rather than hypothesis testing of the economic outcomes. In cases where researchers wish to set up formal hypotheses for the economic analyses, these should be stated *a priori* including the thresholds (e.g., \$50,000 or \$100,000/quality-adjusted life-year [QALY]) and the power to detect when incremental analysis meets or exceeds those thresholds [8].

**Study End Points**

The choice of primary end point in a clinical study may not correspond with the ideal end point for economic evaluation. For example, the use of composite clinical end points is common in clinical trials (e.g., fatal events and nonfatal events combined) to provide greater statistical power. Nevertheless, cost per composite clinical end point is often an unsatisfactory summary measure for an economic analysis, in part because the different outcomes are rarely of equal importance. It is recommended that clinical end points used in economic evaluations be presented in disaggregated form. We recommend weighting end points (e.g., by utilities) so that they yield a measure of QALYs in the case of cost-effectiveness analysis, or a monetary benefit measure in the case of cost-benefit analysis. Alternatively, qual-

ity of life values may be obtained within the trial at regular intervals and the QALYs estimated as one of the outcomes of the trial.

If possible, one should avoid using intermediate end points (e.g., percent low-density lipoprotein reduction) as the measure of benefit; however, intermediate outcome measures are often employed when the costs of conducting a long-term trial are prohibitive. When use of intermediate outcomes is unavoidable, additional evidence is needed to link them with long-term costs and outcomes. If such a link is not reliable or is unavailable, the analyst should argue for follow-up sufficient to include clinically meaningful disease end points.

### *Appropriate Follow-up*

Economic analyses ideally include lifetime costs and outcomes of treatment. Nevertheless, clinical trials rarely extend beyond a few years and are often conducted over much shorter periods. In practice, consideration of the follow-up period for the trial involves the relationship between intermediate end points gathered in the short run and long-term disease outcomes—the stronger that relationship, the more a reliance on intermediate end points can be justified.

### **Data Elements**

The design issues discussed above will impact decisions about which resource use and outcome measures to collect, how to collect them, and how to value them. To begin, we recommend developing a description of the clinical processes for the intervention and how the intervention may impact resource use in the short and long term [9]. In this process, study perspective affects the types of resource use—both medical and nonmedical—that should be considered for inclusion in the study. For example, the societal perspective might include patients' costs for transportation; time spent undergoing treatment, caregiver time, and nonmedical goods and services attributable to the disease or treatment.

After resources have been identified, logistical and cost considerations often require prioritization as to which data elements will be collected. We recommend that analysts focus on “big ticket” items as well as resources that are expected to differ between treatment arms [10]. For the items chosen, the study should collect information on all resource use not just that considered to be disease- or intervention-related [11,12]. If necessary, the distinction between costs related and unrelated to the disease can be attempted at the analysis phase.

For each resource, the level of aggregation desired should be prospectively determined. As an example, hospitalizations could be considered in disaggregated units, such as nursing time, operating-room time, and supplies, or in highly aggregated units, such as numbers of hospitalizations or days in the hospital. The decision is typically driven by the characteristics of the intervention under study, resource use patterns expected, and availability of unit costs, also called unit prices or price weights. For practical reasons, the level of aggregation may be varied by whether the resource use is thought to be disease- or intervention-related.

In some trial settings, secondary data such as hospital bills or claims data are available. These data sources can provide an inexpensive, detailed accounting of some resources consumed by patients, and should be used when available [13].

### *Valuing Resource Use*

Unit costs should be consistent with measured resource use, the study's perspective, and its time horizon. For example, if the level of resource aggregation is hospital days that include intensive care unit (ICU) and non-ICU days, the unit costs should reflect the costs of each type of service [13,14]; if the study is conducted from the societal perspective, unit costs should reflect social opportunity cost. In selecting a costing approach, analysts should weigh issues of accuracy/bias, cost, feasibility, and generalizability [15]. For a more thorough discussion on costing issues, we refer readers to Drummond et al. [16] and Luce et al. [9].

Unit cost estimates are rarely derived via direct observation of patients in trials. Most often they are derived from substudies that are divorced from the trial itself. Sometimes, unit costs will be estimated from trial centers, but more commonly they are derived from national data [17–20]. If a reliable method for cost imputation exists (e.g., diagnostic group weights), one can combine the two methods by collecting a limited set of unit costs in a number of countries and imputing the remainder of costs [21]. Ideally, unit costs to be used for resource costing should be finalized prior to unblinding the trial data.

Because relative costs can affect resource use, in general one should use unit cost estimates that are specific to the precise intervention under study and specific events of interest in the trial but which are generalizable to the population that the study is intended to inform.

When unit costs from more than one country are used in an analysis (e.g., the pooled analysis, when

resource use in a country is multiplied by unit costs from that country), results have to be converted to a common currency if they are to be compared appropriately. Purchasing power parity adjustments are recommended for such conversion [22,23].

### Selecting and Tracking Measures of Outcomes

Because cost–utility analyses are widely accepted, we recommend that analysts collect preference weights as part of clinical trials. The most common method of assessing preferences is the use of a preference-weighted health state classification system such as the EuroQol-5D [24,25], one of the three versions of the Health Utilities Index [26–29], or the Quality of Well-Being Scale [30]. Analysts may also consider the inclusion of a rating scale to measure patient-based preferences [31]. Frequency and timing of these assessments should capture changes in patients' quality of life that may be affected by the treatment but will be influenced by the disease severity of the study population, the study duration, the timing of trial visits, and patient burden [32].

Other options for collecting preference data include direct-elicitation methods such as standard gamble or time-tradeoff exercises. These methods have certain theoretical advantages; however, their use in clinical trials is often difficult. Trained interviewers or computerized applications are routinely used to conduct such exercises [16,33]. Also, many respondents have difficulty understanding and completing the exercises [34–36]. Finally, there is some evidence that these measures are generally unresponsive to changes in health status [37–40]. At present, the balance between the feasibility and desirability of using direct-elicitation methods in clinical trials remains an issue to be decided on a case-by-case basis.

### Database Design and Management

Ideally, collection and management of the economic data should be fully integrated into the clinical data. As such, there should be no distinction between the clinical and economic data elements. As with any prospective study, there should be a plan for ongoing data quality monitoring to address missing and poor quality data issues immediately. Queries should be managed on an ongoing basis rather than at the end of the trial to maximize data completeness and quality, and the timeliness of final analysis.

Informed consent for clinical studies does not routinely include provisions for collection of economic data, particularly from third-party databases. Therefore, explicit language should be

included in trial consent documents. The consent forms should allow for capture of pre- and post-trial economic data if such data are necessary for the economic analysis.

Collection of economic data may reveal adverse events, such as hospitalization, not otherwise found in the clinical data. Data handling procedures are necessary to maintain consistency between economic and safety databases.

In reality, the clinical analysts are usually separate from the economic analysts. The clinical data elements and data formatting procedures needed for the economic analysis should be prespecified such that transfer of all necessary data for the economics study is timely and complete.

## Analysis

### Guiding Principles

The analysis of economic measures should be guided by a data analysis plan. A prespecified plan is particularly important if formal tests of hypotheses are to be performed. Any tests of hypotheses that are not specified within the plan should be reported as exploratory. In addition, the plan should specify whether regression or other multi-variable analyses will be used to improve precision and to adjust for treatment group imbalances. The plan should also identify any selected subgroups and state whether a no intention-to-treat analysis will be conducted.

Although the specific analytic methods used in the analysis of resource use, cost, and cost-effectiveness are likely to differ, there are several analysis features that should be common to all economic analyses alongside clinical trials:

- The *intention-to-treat* population should be used for the primary analysis.
- A common *time horizon(s)* should be used for accumulating costs and outcomes; a within-trial assessment of costs and outcomes should be conducted, even when modeling or projecting beyond the time horizon of the trial.
- An *assessment of uncertainty* is necessary for each measure (standard errors or confidence intervals for point estimates; *P* values for hypothesis tests).
- A (common) real *discount rate* should be applied to future costs and, when used in a cost-effectiveness analysis, to future outcomes.
- If data for some subjects are *missing* and/or *censored*, the analytic approach should address this issue consistently in the various analyses affected by missing data.



### *Trial Costs*

The purpose of clinical trial cost analysis is to estimate costs, cost differences associated with treatment, the variability of differences, and whether the differences occurred by chance.

Once resources have been identified and valued, differences between groups must be summarized. Arithmetic mean cost differences are generally considered the most appropriate and robust measure [41]. Nevertheless, cost data often do not conform to the assumptions for standard statistical tests for comparing differences in arithmetic means [42–44]. They are usually right-skewed and truncated at zero because of small numbers of high-resource-use patients, many patients who incur no costs, and the impossibility of incurring costs less than zero. In most cases, the nonparametric bootstrap is an appropriate method to compare means and calculate confidence intervals [45,46]. Other common nonparametric tests (e.g., Wilcoxon) compare medians and not means and thus are not appropriate [47–49]. Transformations to normalize the distribution are not straightforward and are often sensitive to departures from distributional assumptions. Retransformation to the original scale of costs must include transformation of error terms [50–54].

The same distributional issues that affect univariate tests of costs also affect use of costs as a dependent variable in a multivariate regression analysis. The underlying distribution of costs should be carefully assessed to determine the most appropriate approach to conduct statistical inference on the costs between treatment groups [55]. The choice of the multivariate model requires careful consideration: ordinary least squares and generalized linear models perform differently in terms of bias and efficiency of estimation, depending upon the underlying data distribution [51]. If differences in resource use or subsets of costs are to be estimated, similar considerations regarding the appropriateness of statistical tests based on distributional assumptions should be applied.

When study participants use large amounts of medical services that are unrelated to the disease or treatment under study, it may be difficult to detect the influence of the treatment on total health-care costs. One approach to addressing this problem is to conduct secondary analyses that evaluate costs that are considered related to the disease or treatment under study. If such analyses are performed, it is important to identify services that were deemed “disease-related” versus those deemed “unrelated,”

and to display costs for each component in the treatment and control arms.

### *Outcomes*

When one of the trial’s clinical end points is also used as the outcome for the cost-effectiveness analysis (e.g., in-trial mortality), it is generally most transparent to adopt the methods used in the clinical analysis for the primary analysis plan, particularly if the clinical result is cited in product labeling or a publication. In some cases, the clinical analysis methods are not appropriate for economic analysis (e.g., the clinical analysis may focus on relative treatment differences, while the economic analysis needs absolute treatment differences); if other outcomes are used for the economic analysis, the linkage between the clinical and economic measures should be clearly specified. Analyses of outcomes for the cost-effectiveness study may employ multivariable or other methods that are consistent with the cost analysis or otherwise appropriate for the data [56–60]. Cost-effectiveness analysis should still be performed if the clinical study fails to demonstrate a statistically significant difference in clinical end points. In situations where cost-minimization analysis is conducted, the analyst should also conduct joint analysis of costs and outcomes to convey information about the likelihood of an intervention being cost-effective.

Using nonclinical effectiveness end points, such as QALYs, involves both construction and analysis. Health state utilities, either collected directly from trial patients or imputed based on observed health states, can be transformed into QALYs using standard area-under-the-curve methods [16,61]; a recent consideration involves adjusting for changes in health [62]. Simple analysis of means is the usual starting point; refinements may include adjusting for ceiling effects [63] and modeling of longitudinal effects [64,65].

### *Missing and Censored Data*

Missing data are inevitable in economic analyses conducted alongside trials. Such data can include item-level missingness and missingness because of censoring. In analyzing data sets with missing data, one must determine the nature of the missing data and then define an approach for dealing with the missing data. Missing data may bear no relation to observed or unobserved factors in the population (missing completely at random), may have a relationship to observed variables (missing at random), or may be related to unobserved factors (not missing at random) [66]. Eliminating cases with missing

data is not recommended because it may introduce bias or severely reduce power to test hypotheses. Nevertheless, ignoring small amounts of missing data is acceptable if a reasonable case can be made that doing so is unlikely to bias treatment group comparisons.

Imputation refers to replacing missing fields with estimates. If one chooses to impute missing data, most experts recommend multiple imputation approaches, as they reflect the uncertainty that is inherent when replacing missing data [67–69]. Most commonly used statistical software packages include programs for imputation of missing data. A review of these programs can be found at <http://www.multiple-imputation.com> [70].

Censoring can be addressed with a number of approaches. Most assume that censoring is either completely at random [71] or at random [72–76]. Nevertheless, nonrandom censoring is common, and external data sources for similar patients may be required to both identify and address it.

### Summary Measures

One or more summary measures should be used to characterize the relative value of the treatments in the clinical trial. Three general classes of summary measures are available that differ in how the incremental costs and outcomes are combined into a single metric:

- *Ratio measures* (e.g., incremental cost-effectiveness ratios) are obtained by dividing the incremental cost by the incremental health benefit.
- *Difference measures* (e.g., net monetary benefits) rely on the ability to define a common metric (such as monetary units) by which both costs and outcomes can be measured [77–79].
- *Probability measures* (e.g., acceptability curves) characterize the likelihood the new treatment will be deemed cost-effective based on the incremental costs and outcomes [80,81].

The difference measures and probability measures are calculated for specific values of “willingness to pay” or cost-effectiveness thresholds. Because these values may not be known and/or vary among health-care decision makers, one should evaluate the summary measure over a reasonable range of values.

### Uncertainty

Results of economic assessments in trials are subject to a number of sources of uncertainty, including sampling uncertainty, uncertainty in parameters such as unit costs and the discount rate, and—when

missing data are present—imputation-related uncertainty.

*Sampling uncertainty.* Because economic outcomes in trials are the result of a single sample drawn from the population, one should report the variability in these outcomes that arises from such sampling. Variability should be reported for within-group estimates of costs and outcomes, between-group differences in costs and outcomes, and the comparison of costs and outcomes. One of the most common measures of this variability is the confidence interval.

Policy inferences about adoption of a therapy should be based on one’s level of confidence that its cost for a unit of outcome, for example, a QALY, is less than one’s maximum willingness to pay. Thus, one should report ranges of ceiling ratios for which one: 1) is confident that the therapy is good value for the cost; 2) is confident that the therapy is not good value; and 3) cannot be confident that the therapies differ from one another. Policymakers can then draw inferences by identifying their maximum willingness to pay and determining into which of the ranges it falls.

These ranges of ceiling ratios where one can and cannot be confident about a therapy’s value can be calculated by use of confidence intervals for cost-effectiveness ratios [82,83], confidence intervals for net monetary benefit, or the acceptability curve. One advantage of the confidence interval for the cost-effectiveness ratio is that its limits directly define the boundaries between these ranges. One advantage of the acceptability curve is that it defines the boundaries between these ranges for varying levels of confidence that range from 0 to 100%.

*Parameter uncertainty.* Uncertainty related to parameter estimates such as unit costs and the discount rate should be assessed by use of sensitivity analysis. For example, if one uses a discount rate of 3%, one may want to assess the impact of this assumption by repeating the analysis but using a 1% or a 5% rate. Analysts should evaluate all parameters that, when varied, have the potential to influence policy decisions. Measures of stochastic uncertainty and sensitivity analysis for parameter uncertainty are complements, not substitutes. Thus, when conducting sensitivity analysis, one should report both the revised point estimate and revised 95% confidence intervals that result from the sensitivity analysis.

*Imputation uncertainty.* Finally, some methods employed to address missing or censored data (e.g.,

use of an imputed mean) may artificially reduce estimates of stochastic uncertainty. One should make efforts to address this shrinkage when reporting stochastic uncertainty, for example, by bootstrapping the entire imputation and estimation process.

#### *Identifying and Addressing Threats to External Validity/Generalizability*

Because of the “artificiality” of most clinical trials, they have high internal, but may have low external validity. The threats to external validity come from:

- protocol-driven resource use (which could bias costs in each treatment arm upwards if included and downwards if excluded, but it is generally difficult to know how this will bias the difference between treatments);
- unrepresentative recruiting centers (e.g., large, urban, academic hospitals);
- inclusion of study sites from countries with varying access and availability of health-care services (e.g., rehabilitation, home care, or emergency services);
- restrictive inclusion and exclusion criteria (patient population, disease severity, comorbidities);
- artificially enhanced compliance.

The external validity can best be increased by making the trial more naturalistic during the design phase of the trial [16,84].

Additional threats arise with international trials, as treatment pathways, patient and health-care provider behavior, supply and financing of health care, and unit costs (prices) can differ tremendously between countries [85–92]. Pooled results may not be representative of any one country, but the sample size is usually not large enough to analyze countries separately.

It is common to apply country-specific unit costs for pooled trial resource use to estimate country-specific costs. In practice, this approach yields few qualitative differences in summary measures of cost-effectiveness among countries of similar levels of economic development but may not adjust for important country-specific differences [93,94]. Rather, intercountry differences in population characteristics and treatment patterns are more likely to influence summary measures between countries rather than differences in unit costs. Recommended approaches to address this issue include [93,95–97]:

- hypothesis tests of homogeneity of results across countries (and adjusting the resource use in other countries to better match those seen in country X);

- multivariate cost or outcome regressions to adjust for country effects (e.g., include country dummies or adjusted gross national product per capita as covariates);
- multilevel random effects model with shrinkage estimators.

#### *Modeling Beyond the Time Horizon of the Trial*

The cost-effectiveness observed within the trial may be substantially different from what would have been observed with continued follow-up. Modeling is used to estimate costs and outcomes that would have been observed had observation been prolonged. When modeling beyond the follow-up period for the trial, it is important to project costs and outcomes over the expected duration of treatment and its effects.

Direct modeling of long-term costs and outcomes is feasible when the trial period is long enough, or if at least a subset of patients are observed for a longer time and provide a basis for estimating other patients’ outcomes. Parametric survival models estimated on trial data are recommended for such projections. In cases where such direct modeling is not feasible, it may be possible to “marry” trial data to long-term observational data in a model. In either case, good modeling practices should be followed. The reader is referred to the consensus position of the ISPOR Task Force on Good Research Practices—Modeling Studies for discussion of modeling issues [1].

Cost-effectiveness ratios should be calculated at various time horizons (e.g., 2, 5, 10 years, or as appropriate for the disease), both to accommodate the needs of decision makers and to provide a “trajectory” of summary measures over time. The effects of long-term health-care costs not directly related to treatment should be taken into account as well as possible [98]. As always, assumptions used must be described and justified, and the uncertainty associated with projections must be taken into account.

#### *Subgroup Analysis*

The dangers of spurious subgroup effects are well-known. For example, the probability of finding a difference due solely to random variation increases with the number of differences examined unless the alpha-level is scrupulously adjusted. Yet economics requires a marginal approach, so proper subgroup analysis can be vital to decision makers. The focus should be on testing treatment interactions on the absolute scale, with a justification for choice of scale used. In cases where prespecified clinical interac-



tions are significant, subgroup analyses may be justified. Subgroup analysis based on prespecified economic interactions should also be reported.

### Reporting the Methods and Results

We anticipate that the results of an economic analysis will have a variety of audiences. Correspondingly, detailed and comprehensive information on the methods and results should be readily available to any interested reader. Journal word limits often necessitate parsimony in reporting; therefore, we recommend that detailed technical reports be made available on the World Wide Web.

A number of organizations have developed minimum reporting standards for economic analyses (e.g., study perspective, discount rate, marginal vs. average outcomes and analyses) [99,100]. The principles in these should be adhered to in all economic studies. Here, we highlight issues particular to economic studies conducted alongside clinical trials.

The report should include these elements:

#### *Trial-Related Issues*

- General description of the clinical trial, including patient demographics, trial setting (e.g., country, tertiary care hospital), inclusion and exclusion criteria and protocol-driven procedures that influence external validity, intervention and control arms, and time horizon for the intervention and follow-up;
- Key clinical findings.

#### *Data for the Economic Study*

- Clear delineation between patient-level data collected as part of the trial versus data not collected as part of the trial:
  - Trial: health related quality of life survey instruments, data sources, collection schedule (including the follow-up period), etc.
  - Nontrial: unit costs, published utility weights, etc.
- Amount of missing and censored data.

#### *Methods of Analysis*

- Construction of costs and outcomes;
- In cases where the main clinical end point is used in the denominator of the incremental cost-effectiveness ratio and different methods were used to analyze this end point in the clinical and economic analyses, any differences in the point estimates should be explained;

- Methods for addressing missing and censored data;
- Statistical methods used to compare resource use, costs, and outcomes;
- Methods and assumptions used to project costs and outcomes beyond the trial period;
- Any deviations from the prespecified analysis plan and justification for these changes.

#### *Results*

- Resource use, costs, outcome measures, including point estimates and measures of uncertainty;
- Results within the time horizon of the trial;
- Results with projections beyond the trial (if conducted);
- Graphical displays of results not easily reported in tabular form (e.g., cost-effectiveness acceptability curves, joint density of incremental costs and outcomes).

When there are economic analyses alongside several clinical trials for a given intervention, attempts may be made to estimate a summary cost-effectiveness ratio across trials (although the methods for this are not perfectly straightforward, i.e., clearly not a simple average of the individual incremental cost-effectiveness ratios). Data from economic analyses performed in the context of trials may also be used in independent cost-effectiveness models based on decision analysis or meta-analyses [84]. To facilitate synthesis of economic information from multiple trials, authors should report means and standard errors for the incremental costs and outcomes and their correlation.

#### **Conclusions**

As decision makers increasingly demand evidence of economic value for health-care interventions, conducting high quality economic analyses alongside clinical studies is desirable because they provide timely information with high internal validity. This ISPOR RCT-CEA Task Force Report is intended to provide guidance to improve their quality and consistency. The task force recognizes that there are areas where future methodological research could further improve the quality and usefulness of these studies. Examples here include (among many): new approaches for pooling and analyzing data from multinational trials; issues related to multiple trial analysis, such as Bayesian learning designs, pooling of clinical trial data, or meta-analysis; design and analysis in trials where outcomes are valued in mon-

etary units (i.e., willingness-to-pay studies); methods for projecting trial findings; appropriate methods for a priori selection of items of resource use to be included in trial protocols (e.g., whether to include outpatient services, nonstudy drugs); and selecting levels of aggregation of resources necessary for discriminating between intervention and control (e.g., counts of hospitalizations vs. length of stay). As these methods are identified and validated, they will be included in future versions of this guidance document.

Source of financial support: Support for this project was provided by ISPOR.

## References

- 1 Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision-analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value Health* 2003;6:9–17.
- 2 Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967;20:637–48.
- 3 O'Brien B. Economic evaluation of pharmaceuticals: Frankenstein's monster or vampire of trials? *Med Care* 1996;34:DS99–108.
- 4 Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995;48:23–40.
- 5 Laska EM, Meisner M, Siegel C. Power and sample size in cost-effectiveness analysis. *Med Decis Making* 1999;19:339–43.
- 6 Briggs A, Tambour M. The design and analysis of stochastic cost-effectiveness studies for the evaluation of health care interventions. *Drug Inf J* 2001;35:1455–68.
- 7 Al MJ, van Hout B, Michel BC, Rutten FF. Sample size calculation in economic evaluations. *Health Econ* 1998;7:327–35.
- 8 Gardner MJ, Altman DG. Estimation rather than hypothesis testing: confidence intervals rather than P values. In: Gardner MJ, Altman DG, eds., *Statistics with Confidence*. London: BMJ, 1989.
- 9 Luce BR, Manning WG, Siegel JE, et al. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, et al., eds., *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- 10 Schulman KA, Glick H, Buxton M, et al. The economic evaluation of the FIRST study: design of a prospective analysis alongside a multinational phase III clinical trial. *Flolan International Randomized Survival Trial*. *Control Clin Trials* 1996;17:304–15.
- 11 Garber AM, Weinstein MC, Torrance GW, Kamlet MS. Theoretical foundations of cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, et al., eds., *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- 12 Ramsey SD, McIntosh M, Sullivan SD. Design issues for conducting cost-effectiveness analyses alongside clinical trials. *Annu Rev Public Health* 2001;22:129–41.
- 13 Raikou M, Briggs A, Gray A, McGuire A. Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. *Health Econ* 2000;9:191–8.
- 14 Reed SD, Friedman JY, Gnanasakthy A, Schulman KA. Comparison of hospital costing methods in an economic evaluation of a multinational clinical trial. *Int J Technol Assess Health Care* 2003;19:396–406.
- 15 Coyle D, Drummond MF. Analyzing differences in the costs of treatment across centers within economic evaluations. *Int J Technol Assess Health Care* 2001;17:155–63.
- 16 Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes* (2nd ed.). Oxford: Oxford University Press, 1997.
- 17 U.S. Diagnosis Related Group Weights. Available from: <http://www.cms.hhs.gov/providers/hipps/ipspufs.asp> [Accessed February 1, 2005].
- 18 U.S. National Physician Fee Schedule. Available from: <http://cms.hhs.gov/providers/pufdownload/rvudown.asp> [Accessed February 1, 2005].
- 19 NHS Reference Costs 2003 and National Tariff 2004 ("Payment by Results Core Tools 2004"). Available from: [http://www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsPolicyAndGuidance/PublicationsPolicyAndGuidanceArticle/fs/en?CONTENT\\_ID=4070195&chk=UzhHA3](http://www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsPolicyAndGuidance/PublicationsPolicyAndGuidanceArticle/fs/en?CONTENT_ID=4070195&chk=UzhHA3) [Accessed February 1, 2005].
- 20 Australian Hospital Information, Performance Information Program. Round 6 National Hospital Cost Data Collection Cost Weights. Available from: [http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/health-casemix-costing-fc\\_r6.htm](http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/health-casemix-costing-fc_r6.htm) [Accessed February 1, 2005].
- 21 Glick HA, Orzol SM, Tooley JF, et al. Design and analysis of unit cost estimation studies: how many hospital diagnoses? How many countries? *Health Econ* 2003;12:517–27.
- 22 Drummond MF, Bloom BS, Carrin G, et al. Issues in the cross-national assessment of health technology. *Int J Technol Assess Health Care* 1992;8:671–82.
- 23 Schulman KA, Buxton M, Glick H, et al. Results of the economic evaluation of the FIRST study: a multinational prospective economic evaluation. *Int J Technol Assess Health Care* 1996;12:698–713.

- 24 Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095–108.
- 25 Euroqol Group. EuroQol—a new facility for the measurement of health-related quality of life. The Euroqol Group. *Health Policy* 1990, 1992;16:199–208.
- 26 Torrance GW, Zhang Y, Feeny D, et al. Multi-attribute preference functions for a comprehensive health status classification system. Working Paper No. 92–18. Hamilton, Ontario: McMaster University Centre for Health Economics and Policy Analysis.
- 27 Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res* 1982;30:1043–69.
- 28 Feeny D, Furlong W, Torrance GW, et al. Multi-attribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002;40:113–28.
- 29 Torrance GW, Feeny DH, Furlong WJ, et al. Multi-attribute utility function for a comprehensive health status classification system. *Health Utilities Index Mark 2*. *Med Care* 1996;34:702–22.
- 30 Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res* 1988;23:203–35.
- 31 Gold MR, Patrick DL, Torrance GW, et al. Identifying and valuing outcomes in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, et al., eds., *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- 32 Calvert MJ, Freemantle N. Use of health-related quality of life in prescribing research. Part 2: methodological considerations for the assessment of health-related quality of life in clinical trials. *J Clin Pharm Ther* 2004;29:85–94.
- 33 Sumner W, Nease R, Littenberg B. U-titer: a utility assessment tool. In: Clayton PD, ed., *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*, pp. 701–705. Washington, DC: McGraw Hill, 1991.
- 34 Hollingworth W, Deyo RA, Sullivan SD, et al. The practicality and validity of directly elicited and SF-36 derived health state preferences in patients with low back pain. *Health Econ* 2002; 11:71–85.
- 35 Soucek J, Stacks JR, Brody B, et al. A trial for comparing methods for eliciting treatment preferences from men with advanced prostate cancer: results from the initial visit. *Med Care* 2000; 38:1040–50.
- 36 Woloshin S, Schwartz LM, Moncur M, et al. Assessing values for health: numeracy matters. *Med Decis Making* 2001;21:382–90.
- 37 Katz JN, Phillips CB, Fossel AH, Liang MH. Stability and responsiveness of utility measures. *Med Care* 1994;32:183–8.
- 38 Salaffi F, Stancati A, Carotti M. Responsiveness of health status measures and utility-based methods in patients with rheumatoid arthritis. *Clin Rheumatol* 2002;21:478–87.
- 39 Tsevat J, Goldman L, Soukup JR, et al. Stability of time-tradeoff utilities in survivors of myocardial infarction. *Med Decis Making* 1993;13:161–5.
- 40 Verhoeven AC, Boers M, van Der Linden S. Responsiveness of the core set, response criteria, and utilities in early rheumatoid arthritis. *Ann Rheum Dis* 2000;59:966–74.
- 41 Barber JA, Thompson SG. Analysis and interpretation of cost data in randomised controlled trials: review of published studies. *BMJ* 1998; 317:1195–200.
- 42 O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Econ* 2003;12:33–49.
- 43 Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *BMJ* 2000;320:1197–9.
- 44 Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy* 1998;3:233–45.
- 45 Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
- 46 Desgagne A, Castilloux AM, Angers JF, et al. The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *Pharmacoeconomics* 1998;13:487–97.
- 47 Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of health care interventions. *Health Technol Assess* 1999; 3:1–134.
- 48 Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med* 2000;19:3219–36.
- 49 Briggs A, Clarke P, Polsky D, Glick H. Modelling the cost of health care interventions. Paper prepared for DEEM III: Costing Methods for Economic Evaluation. University of Aberdeen, April 15–16, 2003.
- 50 Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *J Health Econ* 1998;17:283–95.
- 51 Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ* 2001;20:461–94.
- 52 Bland JM, Altman DG. The use of transformation when comparing two means. *BMJ* 1996;312: 1153.
- 53 Bland JM, Altman DG. Transformations, means, and confidence intervals. *BMJ* 1996;312:1079.
- 54 Bland JM, Altman DG. Transforming data. *BMJ* 1996;312:770.

- 55 White IR, Thompson SG. Choice of test for comparing two groups, with particular application to skewed outcomes. *Stat Med* 2003;22:1205–15.
- 56 Hosmer DW Jr, Lemeshow S. *Applied Survival Analysis*. New York: John Wiley & Sons, 1999.
- 57 Kalbfleish JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons, 1980.
- 58 Davidson R, MacKinnon JG. *Estimation and Inference in Econometrics*. New York: Oxford University Press, 1993.
- 59 Hosmer DW Jr, Lemeshow S. *Applied Logistic Regression* (2nd ed.). New York: John Wiley & Sons, 2003.
- 60 Winer BJ, Brown DR, Torrie JH. *Statistical Principles in Experimental Design* (3rd ed.). New York: McGraw-Hill, 1991.
- 61 Patrick DL, Erickson P. Chapter 11, Ranking costs and outcomes of health care alternatives. In: *Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation*. New York: Oxford University Press, 1993.
- 62 Sharma R, Stano M, Hass M. Adjusting to changes in health: implications for cost-effectiveness analysis. *J Health Econ* 2004;23:335–51.
- 63 Austin PC. A comparison of methods for analyzing health-related quality of life measures. *Value Health* 2002;4:329–37.
- 64 Gray SM, Brookmeyer R. Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* 1998;54:976–88.
- 65 Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. *Psychol Methods* 1997;2:64–78.
- 66 Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, 1987.
- 67 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley, 1987.
- 68 Briggs A, Clark T, Wolstenholme J, Clarke P. Missing . . . presumed at random: cost-analysis of incomplete data. *Health Econ* 2003;12:377–92.
- 69 Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3–15.
- 70 Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. *Am Statistician* 2001;55:244–54.
- 71 Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997;53:419–34.
- 72 Bang H, Tsiatis AA. Median regression with censored cost data. *Biometrics* 2002;58:643–9.
- 73 Etzioni RD, Feuer EJ, Sullivan SD, et al. On the use of survival analysis techniques to estimate medical care costs. *J Health Econ* 1999;18:365–80.
- 74 Carides GW, Heyse JF, Iglewicz B. A regression-based method for estimating mean treatment cost in the presence of right-censoring. *Biostatistics* 2000;1:299–313.
- 75 O'Hagan A, Stevens JW. On estimators of medical costs with censored data. *J Health Econ* 2004;23:615–25.
- 76 Raikou M, McGuire A. Estimating medical care costs under conditions of censoring. *J Health Econ* 2004;23:443–70.
- 77 Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;18(Suppl. 2):S68–80.
- 78 Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *Int J Technol Assess Health Care* 1998;14:467–71.
- 79 Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Stat Med* 2001;20:1563–74.
- 80 Van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ* 1994;3:309–19.
- 81 Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;10:779–87.
- 82 O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care* 1994;32:150–63.
- 83 Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Stat Med* 1996;15:1447–8.
- 84 Gold MR, Siegel JE, Russell LB, Weinstein M. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- 85 O'Brien BJ. A tale of two or more cities: geographic transferability of pharmaco-economic data. *Am J Man Care* 1997;3(Suppl.):S33–9.
- 86 Schulman K, Burke J, Drummond M, et al. Resource costing for multinational neurologic clinical trials: methods and results. *Health Econ* 1998;7:629–38.
- 87 Koopmanschap MA, Touw KCR, Rutten FFH. Analysis of costs and cost-effectiveness in multinational trials. *Health Policy* 2001;58:175–86.
- 88 Gosden TB, Torgerson DJ. Converting international cost-effectiveness data to UK prices. *BMJ* 2002;325:275–6.
- 89 Sullivan SD, Liljas B, Buxton MJ, et al. Design and analytic considerations in determining the cost-effectiveness of early intervention in asthma from a multinational clinical trial. *Control Clin Trials* 2001;22:420–37.
- 90 Buxton MJ, Sullivan SD, Andersson FL, et al. Country-specific cost-effectiveness of early inter-

- vention with budesonide in mild asthma. *Eur Respir J* 2004;24:568–74.
- 91 Gerdtham UG, Jönsson B. Conversion factor instability in international comparisons of health care expenditure. *J Health Econ* 1991;10:227–34.
- 92 World Bank. *World Development Indicators 2001*. Washington, DC: The World Bank, 2001.
- 93 Willke RJ, Glick HA, Polsky D, Schulman K. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Econ* 1998;7:481–93.
- 94 Barbieri M, Drummond M, Willke R, et al. Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: lessons for inferring generalizability. *Value Health* 2005; 8:10–23.
- 95 Jönsson B, Weinstein M. Economic evaluation alongside multinational clinical trials: study considerations for GUSTO IIb. *Int J Technol Assess Health Care* 1997;13:49–58.
- 96 Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. *Stat Med* 2003;22:1955–76.
- 97 Wordsworth S, Ludbrook A. Comparing costing results in across country evaluations: the use of technology specific purchasing power parity estimates. *Health Econ* 2005;14:93–9.
- 98 Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ* 1997; 16:33–64.
- 99 Drummond MF, Jefferson TO, for the BMJ Working Party on Guidelines for Authors and Peer-Reviewers of Economic Submissions to the British Medical Journal. Guidelines for authors and peer-reviewers of economic submissions to the British Medical Journal. *BMJ* 1996;313:275–383.
- 100 Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996;276:1339–4.