OHE briefing
No 32 September 1995 £5.00

# HANDLING UNCERTAINTY IN THE RESULTS OF ECONOMIC EVALUATION

## Introduction

The recent increase in the number of published economic evaluations has been considerable [Wellcome, 1992; Udvarhelyi *et al*, 1992]. It is of some concern, however, that reviews of economic evaluations have highlighted a high degree of methodological shortcomings in many studies [Adams *et al*, 1992; Gerard, 1992]. Furthermore, the situation does not appear to have improved over time [Udvarhelyi *et al*, 1992]. In particular, the importance of dealing systematically and comprehensively with uncertainty appears to have been overlooked by many analysts. Udvarhelyi and colleagues note that, although authors frequently mentioned the limitations in their underlying assumptions, only 30% of studies used sensitivity analysis to explore the effect of changes in those assumptions [Udvarhelyi *et al*, 1992]. Adams and colleagues found that only 16% of studies had utilised sensitivity analysis in their review of economic analyses alongside randomised trials [Adams *et al*, 1992]. By contrast, Gerard found that 79% of cost utility analyses reviewed had conducted a sensitivity analysis, although just over half of these were judged to be limited in scope [Gerard, 1992]. In a recent review of economic evaluations focusing on methods employed to handle uncertainty, the concerns raised by the more general methodological reviews were found to be justified [Briggs & Sculpher, forthcoming]. A summary of this review is given in text box one.

The increasing use of the randomised controlled trial (RCT) as a vehicle for economic evaluation presents the opportunity to sample economic as well as clinical data and offers the potential for uncertainty to be quantified through conventional statistical techniques [O'Brien *et al*, 1994]. However, most economic evaluations are based largely on deterministic data (ie, estimates that have been taken from the literature or provided by experts) which have no intrinsic measure of variance, and therefore statistical analysis is impossible. Even where stochastic data (ie, data which have been sampled allowing estimation of both average values and associated variance) are collected from a clinical trial, there is a continuing role for sensitivity analysis in dealing with those parameters where uncertainty is not related to sampling error [Briggs *et al*, 1994]..

Many commentators in the economic evaluation methodology literature stress the importance of using sensitivity analysis to test the robustness of a study's conclusions [Weinstein *et al*, 1980; Weinstein, 1981; Drummond *et al*, 1987; Eisenberg, 1989; Luce & Elixhauser, 1990]. Perhaps more significantly, recent guidelines for conducting economic evaluation drawn up between the Department of Health and the Association of the British Pharmaceutical Industry [ABPI, 1994] stress that not only should economic evaluation include sensitivity analysis but that the results of that analysis should be quantitatively reported. The failure of many studies to use any sensitivity analysis or to present only a limited analysis highlights the significance of these guidelines. Despite the many recommendations to conduct sensitivity analysis, few details are offered as to how exactly the analysis should be carried out and how the results should be presented. Sensitivity analysis is not a single technique but encompasses a range of approaches designed to examine the effect of changing the underlying assumptions of a study. Many of the terms employed, such as 'robustness' and 'plausible range', are ill-defined and open to a good deal of interpretation.

The purpose of this paper is to examine uncertainty in economic evaluation and how sensitivity analysis can be employed to represent that uncertainty. This paper should be of interest to all those intending to undertake economic evaluation as well as those considering applying the results of completed evaluation studies.

**Box 1** *Results of a review of sensitivity analysis*

A structured methodological review of journal articles published in 1992 was undertaken to determine whether recently published economic evaluation studies deal systematically and comprehensively with uncertainty. Ninety three journal articles were identified from a range of searches including a computerised search of the MEDLINE CD-Rom database. Articles were reviewed to determine how they had handled uncertainty in: a) data sources; b) generalisability; c) extrapolation; and d) analytic method. Articles were subsequently assessed to determine how they had synthesised cost and outcome data whilst quantifying this uncertainty in terms of the overall results of their analysis. Finally, studies were rated on the basis of their overall performance with respect to dealing systematically and comprehensively with uncertainty.

The results were disappointing: 22 (24%) studies failed to consider uncertainty at all; 35 (38%) studies employed sensitivity analysis either in a manner judged as inadequate or they failed to give sufficient information for a judgement to be made. Only 36 (39%) studies were judged to have given at least an adequate account of uncertainty, with just 13 (14% overall) of those judged to have given a good account of uncertainty. Studies published in the general medical journals scored higher on average than those published in clinical sub-speciality journals.

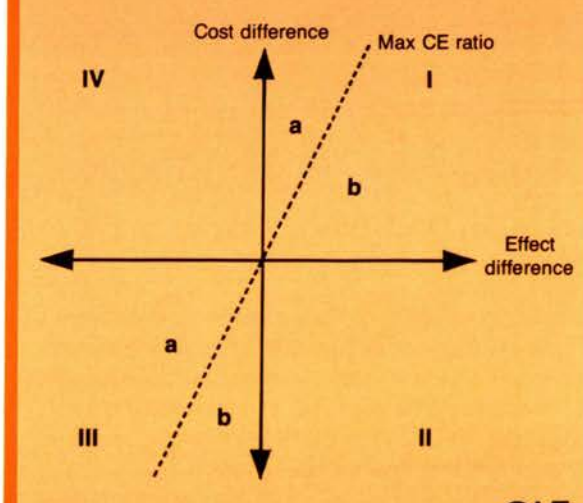
The reason for such disappointing results may be the general lack of detail in much of the methodology literature concerning exactly how the methods for handling uncertainty should be applied and how results should be presented. Journal editors and readers of economic evaluation articles should acquaint themselves with these methods in order that they can critically evaluate the extent to which authors have allowed for uncertainties inherent in their analysis (Briggs & Sculpher, forthcoming).

## Why quantify uncertainty?

By directly relating the costs and benefits of two (or more) alternative interventions, economic evaluation seeks to improve the efficiency of health care provision at two levels. Firstly, by identifying the least cost alternative for providing health care of a minimum standard within a particular area; and secondly, by identifying the appropriate allocation of resources between medical specialities. To achieve the former, it may be appropriate to measure the effectiveness of alternative interventions in units relevant to the particular medical area: for example, the number of episode free days in the treatment of asthmatics. However, where decisions are to be made relating to the allocation of resources between medical specialities, effectiveness must be measured in common units. One such measure which has received much attention is the Quality Adjusted Life Year (QALY); many studies now present their results in terms of costs per QALY, allowing direct comparison of cost per QALY figures between alternative interventions. A summary of the QALY approach can be found in text box two.

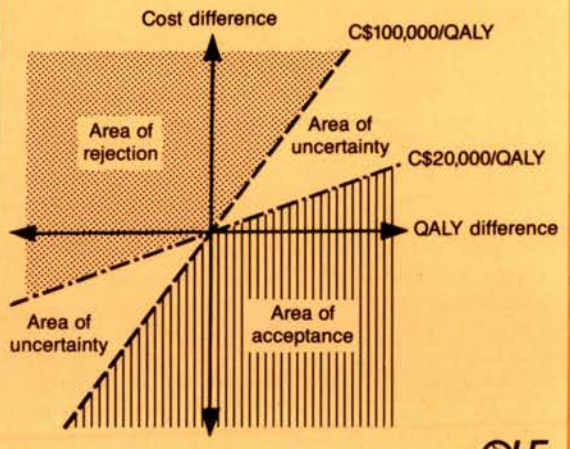
If the results of an economic evaluation demonstrate both cost-savings and increased benefits for a particular health care intervention then that intervention is said to *dominate* the comparator. In such a case the dominant strategy is clearly cost-effective. However, if one health care intervention is shown to be more costly but also more effective than a comparator it is impossible to say *a priori* whether that intervention is cost-effective. Instead, an incremental cost-effectiveness ratio can be calculated and compared to other cost-effectiveness ratios representing alternative uses of health care resources. Decisions relating to the cost-effectiveness of interventions can be presented diagrammatically within the cost-effectiveness plane shown in Figure 1 [Anderson *et al*, 1986; Black, 1990; Laupacis *et al.* 1992]. The horizontal axis represents the difference in effect between two alternative interventions and the vertical axis represents the difference in costs. If the results of an economic evaluation show the difference in costs and effects of alternatives such that a point in quadrant II or IV is indicated, a clear case of dominance exists; ie, one intervention has been shown as more cost-effective than the other. Where, however, a point in quadrants I or II is indicated, the judgment of cost-effectiveness will depend upon whether the additional health benefits are worth the additional costs [Doubilet *et al.* 1986]. The maximum acceptable cost-effectiveness ratio is defined as the point where the additional benefits are just worth the additional cost and is indicated in Figure 1 by the (slope of the) dotted line. An intervention falling in the Ia or IIIa areas is judged not cost-effective, whereas any intervention in the Ib or IIIb areas is cost-effective. How uncertainty affects the position of the base case analysis on the cost-effectiveness plane is crucial and is addressed toward the end of this section.

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



Where the measures of effectiveness chosen are specific to a given medical speciality, the maximum acceptable cost-effectiveness ratio will depend on the value placed upon the specific outcome measure; for example, an episode free day in the treatment of asthma or a symptom free day in the treatment of dyspepsia. The measurement of effectiveness in terms of a common outcome such as the QALY allows for inter-speciality comparison to be made without the need for individual specification of relative values between disease specific outcomes. Ultimately, the decision as to the maximum acceptable cost-effectiveness ratio for a health care intervention in terms of QALYs is a societal one and will depend upon the willingness of society to pay for additional health benefits, which in turn will depend upon the relative value society places on health care compared to alternative uses of public money, such as education, transport, housing etc. However, other factors important to the priority setting process, such as societal preferences for particular diseases, groups of patients and/or medical care, are not addressed by the QALY approach (see text box two). Consensus as to a maximum cost-effectiveness ratio has not therefore been achieved. Laupacis *et al* have suggested that, in Canada, any health care intervention with a cost-per-quality adjusted life year (QALY) ratio below C$20,000/QALY has strong evidence for adoption and any intervention costing more than C$100,000/QALY has only weak evidence for adoption [Laupacis *et al*, 1992]. Figure 2 presents this information on the cost-effectiveness plane. The implication is that they regard the maximum acceptable cost-effectiveness (QALY) ratio to lie somewhere between C$20,000 and C$100,000 per QALY [Battista, 1992]. Uncertainty associated with the results of economic evaluations has important implications for the decision making process. Wrong decisions are costly. The failure to implement a cost-effective strategy is, in principle, just as costly as the implementation of a non-cost-effective strategy in the sense that such decisions will result in a failure to maximise health benefit from available resources. Decisions cannot be said to be 'fully informed' unless they are taken with a knowledge of the implications of uncertainty. Where a situation of dominance exists for the base case parameters (ie, one intervention is both less costly and more effective - appearing on the cost-effectiveness plane as a point in either quadrant II or IV in Figure 1), uncertainty in the value of those parameters could

**Figure 2** *Cost-per-QALY limits proposed by Laupacis et al (1992)*




Adapted from Laupacis *et al* (1992)

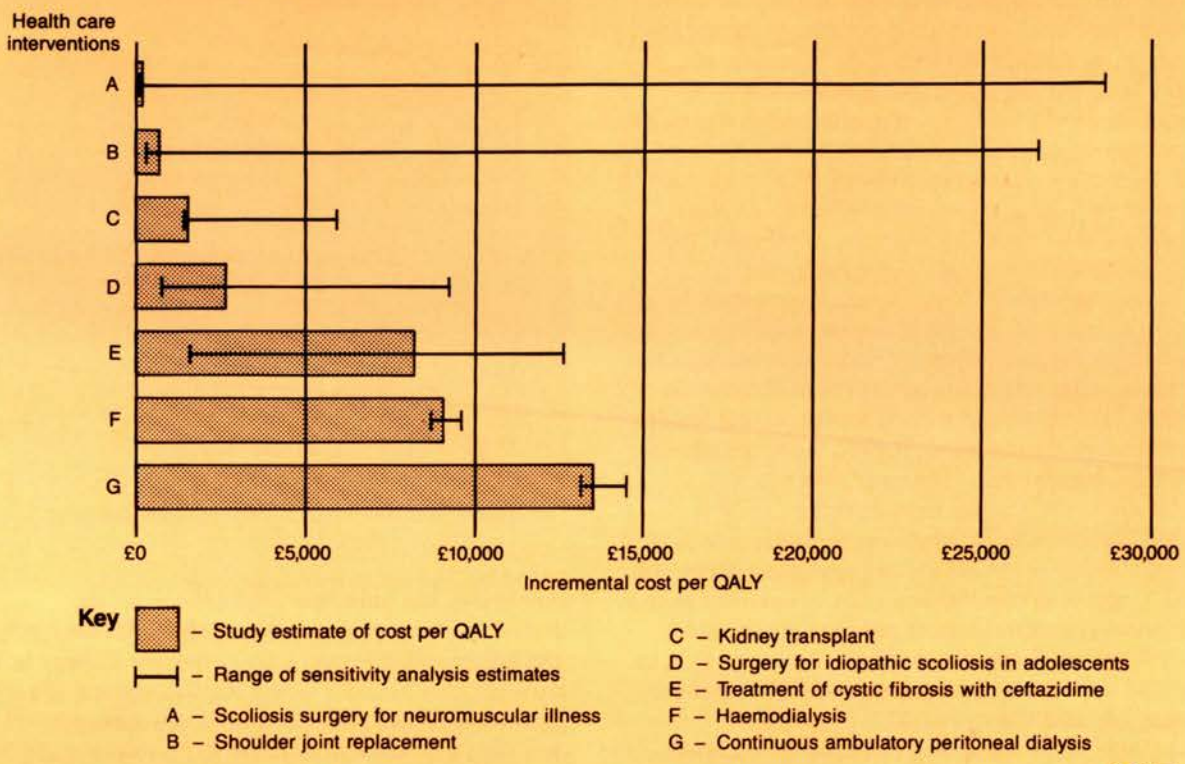
## Box 2 *The Quality Adjusted Life Year (QALY)*

Quality adjusted life years (QALYs), use the quality of life experienced by a patient in a particular state of health to weight the length of time spent in that state. In this way, QALY profiles can be constructed which show the quality of life of patients experiencing particular treatment regimens for a given disease over the course of their treatment and/or disease. The difference in areas between these profiles show the QALY gain associated with a particular therapy compared to another. By relating this gain to the additional costs associated with the therapy, an incremental cost-per-QALY ratio can be constructed. In principle, by comparing a number of alternative technologies in terms of their cost-per-QALY, they can be ranked in order of their cost-effectiveness. Ultimately, society could specify a maximum acceptable cost-per-QALY ratio, above which new technologies would be deemed to be not cost-effective and would therefore not receive funding. Since the QALY is a measure common to all health care interventions, comparisons would not be limited by disease speciality, but could inform resource allocation decisions between specialities.

In practice, there are a number of objections that have been raised to the QALY approach. Many of these centre around the fact that different methods employed in both the cost and QALY calculations and the different alternatives to which the assessed technologies were compared, may invalidate the subsequent comparison of cost-per-QALY ratios on the grounds that like is not being compared with like. It has also been suggested that the use of QALYs may result in forms of resource allocation which are unjust; for example, some people believe that QALYs are biased against the elderly since old people have relatively less time in which to accrue QALYs. Research has shown that the public value health at different levels according to the stage of a person's life; for example, the health of a pregnant woman with a young family is valued more highly than a single man's health. Since the fundamental principle of QALYs is that 'a QALY is a QALY no matter to whom it accrues', the application of cost-per-QALY type analysis to resource allocation could result in decisions which do not truly reflect the wishes of society.

It is important to recognise that some form of single outcome measure is required in order that explicit comparisons of the alternative use of scarce health care resources can be made. Although QALYs have their critics, they are a strong contender for that single outcome measure and cost-per-QALY figures represent a valuable input into the decision making process. However, a healthy scepticism should be maintained whenever cost-per-QALY comparisons are made.

**Figure 3** *Variability in point estimates of incremental cost-effectiveness following sensitivity analysis*




Adapted from Petrou *et al* (1993)

potentially lead to a situation where that intervention no longer dominates the other. The importance of uncertainty will depend on the extent to which it affects the appropriate decision. For example; although the baseline estimates of an economic evaluation may indicate that a particular intervention is cost-effective, inherent uncertainties in the analysis could conceivably mean that in reality the intervention is not cost-effective at all. Figure 3 shows that the application of a basic sensitivity analysis to a number of studies reporting point estimates of cost-effectiveness ratios (illustrated by the solid bars) can lead to wide estimates of the possible range of values around those ratios (represented by the 'I' bars) [Petrou, 1993]. More importantly, it is clear that within the bounds of uncertainty illustrated in Figure 3, the true incremental cost-effectiveness ratios may have a completely different ordering to that suggested by the point estimates. Suppose that the maximum acceptable cost-effectiveness ratio were £10,000 per QALY. It is clear that given such a scenario, a decision maker seeking to invest additional health care resources might prefer intervention **C** over either **A** or **B** due to the relative precision of its incremental cost-effectiveness ratio.

## Types of uncertainty

It is useful to consider separately four broad types of uncertainty in economic evaluation. These relate to:

- data requirements of the study,
- extrapolation of data or endpoints,
- generalisabilty of results,
- and the choice of analytic methods.

### *Uncertainty relating to data requirements*

The data required for any full economic evaluation are the resource consequences and non-resource consequences (health outcomes) of the technologies being compared, and the data necessary to value those consequences. Variability within the population of interest with respect to these parameters is a key source of uncertainty in economic evaluations. Uncertainty of this sort can be handled by sampling from the appropriate population and applying standard statistical methods to obtain an estimate representative of that population. Indeed, with the increasing use of the RCT as a vehicle for collecting economic as well as clinical data [Drummond and Davies, 1991] there has been recent interest relating to the calculation of confidence intervals for the results of economic evaluations [O'Brien *et al*, 1994]. They key elements of this so called 'stochastic' cost-effectiveness approach are summarised in text box three. Although the stochastic estimates of cost-effectiveness potentially allow the use of the standard statistical techniques of hypothesis testing to indicate whether a particular intervention is cost-effective, this approach can only quantify uncertainty in sampled data.

### *Uncertainty relating to extrapolation*

Extrapolation in economic evaluation can occur where economic evaluations have used an intermediate endpoint of clinical outcome and extrapolated that intermediate endpoint to a final health outcome. For example, when addressing the issue of the most cost-effective method for reducing blood cholesterol (an intermediate health outcome) it is assumed that

**Box 3** *Stochastic cost-effectiveness studies*

As more economic evaluations have become integrated into clinical trials, so interest has grown in the calculation of statistical confidence intervals for cost-effectiveness ratios. Clinical trials offer the potential to sample economic as well as clinical data allowing the use of standard inferential statistical techniques.

O'Brien *et al* describe a method for calculating the confidence intervals for a cost-effectiveness ratio in a hypothetical case. By treating the difference in costs (**ΔC**) and the difference in effects (**ΔE**) between two interventions as random variables, they are able to employ a Taylor Series expansion to estimate the variance associated with the ratio of incremental costs to incremental effects (**ΔC/ΔE**).

In principle, the advantage of such an approach would be that uncertainty in economic evaluation could be represented by confidence intervals which are a widely understood and accepted method for quantifying uncertainty. Difference in cost-effectiveness could then be tested by the accepted methods of inferential statistics.

In practice, O'Brien *et al* still have to overcome a number of simplifying assumptions before the approach they describe will become operational, for example, unit costs are treated as deterministic in their exposition whereas, in reality, unit costs are also random variables. Furthermore, it is not clear the extent to which the accepted conventions of statistical power and significance are relevant to economic evaluation [O'Brien and Drummond, 1994].

As the methods for stochastic cost-effectiveness studies are refined [van Hout *et al.* 1994; Willan and O'Brien, *forthcoming*], it is likely that more studies will attempt to quantify uncertainty in economic evaluation by the calculation of confidence intervals around point estimates of incremental cost-effectiveness. However, at present, the methods for calculating confidence intervals around cost-effectiveness ratios are sufficiently underdeveloped to make this approach generally inaccessible.

reducing blood cholesterol reduces heart disease (a final health outcome). Uncertainty is clearly introduced into this process since the relationship between the reduction of cholesterol and the reduction of heart disease is an estimated parameter in the evaluation [Schulman *et al*, 1990].

A second form of extrapolation occurs where data are extrapolated beyond the primary data source. Patients in clinical trials tend to be followed for short time periods, typically one or two years. However, the economic costs and benefits of a technology may continue for the lifetime of the patients. For example, in an economic evaluation of zidovudine therapy versus no therapy for asymptomatic HIV patients, the authors considered the cost-effectiveness of the drug on the basis of clinical trial results for one year's therapy, and then sought to extrapolate those results to patients' entire lifetimes by modelling the profile of the two survival curves [Schulman *et al*, 1991]. The inherent uncertainty in this process was largely responsible for the wide ranging estimates of cost-effectiveness in their study.

### *Uncertainty relating to generalisability*

Generalisability is concerned with the extent to which the results of a study, as they apply to a particular population/context hold true for another population or in a different context. The extent to which the results of a study are generalisable is another source of uncertainty in economic evaluation. A key form of this type of uncertainty concerns whether the results of a study conducted on one group of patients is also valid for another. Differences in relative prices/costs, demography and epidemiology of disease, availability of health care resources, incentives to health care professionals and institutions, and variations in clinical practice may all affect the relative cost-effectiveness of the same health care technology in different countries [Drummond, 1994]. For example, in a comparison of medical and surgical treatments for duodenal ulcer, Sonnenberg argued that the medical treatment of duodenal ulcer was more cost-effective since symptoms were well controlled and most patients avoided the cost and discomfort of surgery. However, he warned against the extrapolation of his USA based results to Europe due to the significantly lower cost of surgery in most European countries [Sonnenberg, 1989].

Another form of uncertainty relating to generalisability is the extent to which the cost-effectiveness observed in a trial would hold true in routine clinical practice. It is well known that experimental trial designs may impose atypical patterns of care on patients [Schwartz & Lellouch, 1967; MacRae, 1989], that is, a clinical trial may lack *external validity* and a technology shown to be cost-effective on the basis of data from a trial may no longer prove cost-effective when data based on routine clinical practice are considered. Data are said to show *efficacy* when the purpose of a trial was to demonstrate the clinical potential of an intervention; by contrast, *effectiveness* data show the effect that a technology will have when used in the 'real world' environment. Even economic evaluations based upon pragmatic clinical trials (designed to increase external validity by analysing on an intention-to-treat basis) may not truly represent the effectiveness that will be achieved after widespread dissemination of a technology [Evans and Robinson, 1980]. Monitoring will be more comprehensive due to the fact that a trial is in progress and compliance among patients may well be affected by the knowledge that they are taking part in a clinical trial.

### *Uncertainty relating to analytic method*

The analytic methods used in an economic evaluation consist of a range of techniques including methods of measuring and valuing resource consequences and health outcomes, and the choice of costs and benefits to include in an evaluation. In a number of areas economists disagree as to the appropriate methods [Drummond *et al*, 1993]. An example is the recent

debate in the literature concerning the appropriate way to include time preference in economic evaluations [Cairns, 1992; Coyle & Tolley, 1992; Parsonage & Neuberger, 1992; Sheldon, 1992]; specifically, should health benefits be subjected to discounting at the same rate as costs, if at all?

## Methods to handle uncertainty

There are two potential methods for dealing with uncertainty in economic evaluation: statistical analysis and sensitivity analysis. They differ in terms of the scope of uncertainty to which they can be applied and the potential for bias inherent in the approach.

### *Statistical analysis*

In clinical evaluation, statistical analysis is accepted as the appropriate method for representing uncertainty, with the RCT widely regarded as the appropriate vehicle for generating the sample data. In economic evaluation the role of statistical analysis for estimation and hypothesis testing may be limited. Despite the increased use of economic analysis alongside clinical trials, the number of technologies for which there are high quality sample data regarding costs and effects of all alternatives is relatively few. Although economic evaluations conducted alongside RCTs provide sample data on resource use, the unit costs of those resources are generally provided from a single setting, even when more than one clinical centre is involved in the study. There is clearly a potential for unit costs to be sampled from a number of institutions, although care must be taken to ensure that the sample of clinical centres is a representative sample of the appropriate population of centres.

Where suitable sample data do exist and as the methods for applying statistical methods in stochastic cost-effectiveness studies are continually refined [Van Hout *et al*, 1994; Willan & O'Brien, *forthcoming*], it may be that statistical analysis becomes the method of choice for dealing with uncertainty in the data sources of a study. A more detailed discussion of the issues surrounding uncertainty in stochastic cost-effectiveness studies can be found in text box three.

### *Sensitivity analysis*

The strength of statistical analysis lies in the randomisation process which controls for confounding factors, and in the potential to avoid bias by blinding trial participants to the treatment process wherever possible. However, statistical analysis is unable to deal with uncertainty associated with extrapolation, generalisability and analytic method. Furthermore, the majority of economic evaluations are not carried out alongside RCTs but involve deterministic estimates. In such cases, sensitivity analysis is the only available method for handling uncertainty and is therefore the focus of this briefing paper. Sensitivity analysis involves varying parameters of the evaluation across 'feasible' or 'plausible' ranges to examine the effect for the results of the study.

## Types of sensitivity analysis

Although sensitivity analysis is commonly referred to as if it were a single approach to handling uncertainty, there are a number of approaches that can be adopted when conducting a sensitivity analysis. It is useful to distinguish four main types which are detailed below.

### *Simple sensitivity analysis*

This is the most common form of sensitivity analysis. One or more parameters are varied across their plausible range (interpretation of 'plausible' will be considered below). A distinction can be made between one-way and multi-way analysis. In a one-way analysis, extreme values are taken for each parameter individually to examine the effect on the results of a study. For example, in a cost-effectiveness analysis of a screening programme for hepatitis B surface antigen in India, the authors carried out a one-way sensitivity analysis, which they illustrated graphically by plotting the range of cost and effectiveness which resulted from the variation of each parameter in their analysis [McNeil *et al*, 1981].

Multi-way analysis allows the variation of more than one parameter at a time. However, it becomes increasingly difficult to present the results of a multi-way analysis the greater the number of parameters varied. For example, in an economic evaluation of antenatal HIV testing for women of childbearing age, the authors presented a two-way sensitivity analysis relating the risk group of the women and the percentage reduction in adult contacts for women identified as HIV positive, to the dollar savings per woman screened [Brandeau *et al*, 1992]. By producing a series of these two-way sensitivity analyses, the authors were able to present the additional effect of varying assumptions concerning indirect costs, counselling costs, and level of infection transmission. The effect they achieved was to present three separate three-way sensitivity analyses where the third variable changed in each case.

### *Extreme scenario analysis*

Extreme scenario analysis is simply a special case of a multi-way simple sensitivity analysis where all the most favourable values for a given intervention are combined to give a 'best case' scenario and all the least favourable values are combined to give a 'worst case' scenario. The effect of applying these scenarios for the study results can then be determined. For example, in a study examining the cost-effectiveness of three alternative strategies (treat-all, test, wait-and-see) for the prevention of Lyme disease after tick bites, the authors found that a one-way sensitivity analysis of each parameter did not alter the relative ordering of the treatment strategies with respect to costs and outcomes [Magid *et al*, 1992]. They therefore constructed a worst-case scenario, relative to the least cost 'treat-all' strategy, by combining all of the least favourable estimates of the parameters used in the one-way sensitivity analysis. Since the 'test' strategy remained both more costly and less effective than the 'treat-all' strategy, their subsequent incremental cost-effectiveness analysis focussed on the relationship between the 'treat-all' and 'wait-and-see' strategies.

### *Threshold analysis*

Threshold analysis does not explicitly involve the specification of ranges for parameters. Rather, the critical value of a parameter relating to the decision

threshold is identified. The problem in economic evaluation is identifying the relevant decision rule. In theory, this decision rule is the maximum acceptable cost-effectiveness ratio; in practice however, it may be impossible to agree a universally acceptable value for such a ratio. A good example of the practical application of threshold analysis is in the economic evaluation of a new pharmaceutical product prior to price setting. In a preliminary analysis of ondansetron, Buxton and O'Brien examined two thresholds associated with pricing the drug by reanalysing original clinical trial data on efficacy and side-effects [Buxton & O'Brien, 1992]. The first threshold they calculated was that which equalised the net costs of each drug therapy. This threshold indicates the point at which the most effective intervention (ondansetron) just ceases to dominate the comparator intervention. The domination of one intervention over another gives a clear decision rule, hence the knowledge of such thresholds can be an extremely important part of a manufacturers' pricing policy as well as being highly desirable from the perspective of potential payers. Secondly they calculated the threshold which equalised average cost-effectiveness. The value of this second threshold is limited since it is impossible to judge *a priori* the maximum acceptable cost-effectiveness ratio relevant to a given decision maker in a given context. However, at this point of equal cost-effectiveness, the incremental cost-effectiveness of the new therapy is at least as good as the cost-effectiveness of the previous therapy.

In a similar way, threshold values can be identified for other parameters of an economic evaluation. Although the analyst does not control these parameters (in the way that a manufacturer can control the price of their own product), the threshold values can be presented to the decision maker in order that they can judge whether those values are likely to occur in reality. Thus they will be able to reach a judgement concerning the likelihood that a particular intervention will dominate another in terms of cost-effectiveness. For example, in a cost-effectiveness analysis of stenting compared to conventional angioplasty as a treatment for symptomatic coronary disease, the authors presented the threshold values of the stent and conventional angioplasty restenosis rates which showed the cost effectiveness of stenting to be $20 000, $40 000 and $60 000 per QALY. They then went on to examine the values of other parameters which kept the cost-effectiveness of stenting below $40 000 per QALY (they justify their use of $40 000 per QALY as similar to the reported cost-effectiveness of haemodialysis or treatments for mild hypertension) [Cohen *et al*, 1994].

### *Probabilistic sensitivity analysis*

One problem with the types of sensitivity analysis discussed above is that they do not contain information concerning the relative likelihood that the extreme values or scenarios under consideration will occur. Probabilistic sensitivity analysis attempts to overcome this problem by applying distributions to the specified ranges and sampling at random from these distributions to simulate uncertainty, thereby generating an empirical distribution of the cost-effectiveness ratio. Although probabilistic sensitivity analysis is a promising approach, further development is required before this approach is likely to be routinely adopted. Text box four summarises the approach and outlines its potential.

## Presenting sensitivity analysis results

The aim of conducting a sensitivity analysis is to present the consequences of the inherent uncertainty of an evaluation for the study results. In particular, potential users of economic evaluation will want to know how *robust* the results are to variation in the underlying parameters.

### *Preliminaries*

Two important steps must be taken before conducting a sensitivity analysis. Firstly, the parameters to be included in the analysis must be identified; and secondly, the range of values over which the parameters are to be varied must be specified.

**Box 4** *Probabilistic sensitivity analysis*

One of the major limitations of traditional sensitivity analysis is that it contains no information concerning the *likelihood* that a particular intervention may prove cost-effective. For example, an intervention may be shown to be cost-effective with base-line assumptions but may cease to be cost-effective if some key assumptions are changed. The decision maker will want to know how *likely* it is that the intervention will be cost-effective, ie, they need to know how much confidence they can place in the study results.

Probabilistic sensitivity analysis is an approach that has been used in the medical decision making literature to address this problem [Doubilet *et al*, 1985]. Uncertainty in the parameters of a decision analytic model is represented by a range of values, and distributions are then defined for these ranges. A process known as Monte-Carlo Simulation is then employed whereby a computer picks values, at random, from the ranges specified for each parameter according to the specified distribution. The result for each set of values is then recorded and the process repeated for a large number of runs. The proportion of times one option is preferred over another gives the 'confidence' that the decision maker can place on the results of the analysis.

A major problem with this approach is the lack of data on which to base decisions concerning the range and distribution of values for each parameter. Furthermore, when applying these methods to economic evaluation a number of problems occur concerning the appropriate way to deal with cost-effectiveness ratios. Probabilistic sensitivity analysis cannot deal with uncertainty related to the choice on analytic method so would require a separate analysis for each change in the methods employed. However, probabilistic sensitivity analysis is a promising approach, and although it is not easily accessible, as yet, its use in economic evaluation warrants further research.

## *Specifying parameters for inclusion*

Care should be taken to include all parameters in the sensitivity analysis, ie, not simply those relating to the data requirements of the evaluation, but also those relating to extrapolation and analytical method. Particularly where the evaluation includes a detailed costing, consideration should be given for combining data elements into a single parameter for the purposes of the sensitivity analysis. For example, where there are a large number of resource items costed, variations in individual unit costs are unlikely to affect the results of the analysis to any significant extent. Furthermore, individual unit costs are unlikely to vary in isolation. A more meaningful approach may be to consider variations in all unit costs simultaneously as a single parameter in the sensitivity analysis.

## *Defining ranges for each parameter*

The choice of range of values specified for these parameters will have important implications for the generalisability of the results. For example: within the confines of a particular hospital, patterns of resource use and the unit costs of those resources may not be expected to vary much. Hence, the ranges chosen for these variables for a study concerned directly with provision of a service within that provider unit may be simply reflect the degree of uncertainty associated with the measurement of that data. However, if authors attempt to generalise their results beyond their own institution, the ranges chosen for parameters must reflect the fact that patterns of resource use and unit costs are likely to vary between institutions. Attempts by authors to address the issue of generalisability would benefit those decision makers who are concerned with the application of the results of economic evaluations at their own local level.

Once the level of generalisability required from the results of a study has been determined, authors must then attempt to define a *plausible* range of values for their parameters. Although frequently used to describe the range of values included in sensitivity analyses, the term plausible has received little attention in the literature. What constitutes a plausible range and how should authors go about making sure the ranges they use are plausible? The problem is that the term plausible is subjective and is as such very difficult to define; however, it is possible to infer certain concepts of what is should be meant by a plausible range. Firstly, the term plausible embodies a notion of likelihood, ie, that ranges should include all probable values but not all possible ones. If all possible values of parameters were included in a given range, no matter how remote the chance that they might occur, it is likely that a wide range of possible cost-effectiveness ratios will be produced which will tell the decision maker very little about which intervention is likely to be most cost-effective. Secondly, the term 'plausible' implies that a level of assessment has been applied to the range rather than simply incorporating *ad hoc* values. It is important, therefore, that authors spend time assessing the appropriate range of values to include in any sensitivity analysis. For example: it may be possible to sample expert opinion in the absence of sample data from an RCT on the clinical effectiveness of an intervention. More frequently, perhaps, evidence on the appropriate ranges of values will be found from a wide ranging review of the literature. Although ultimately the ranges of values included in a sensitivity analysis are under the control of the analyst and as such will be open to the criticism of bias, an open attempt to justify the ranges chosen will make the selection process explicit and allow criticism and comment.

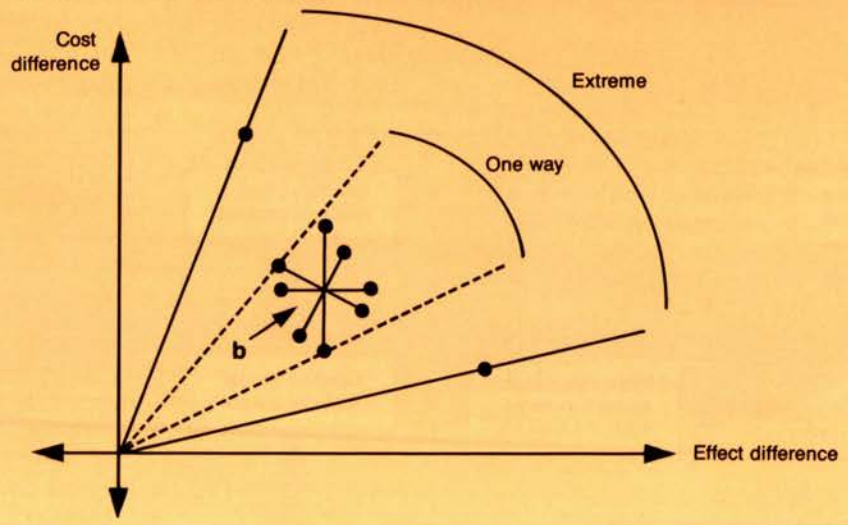
## *Presenting the results of sensitivity analysis*

The results of sensitivity analyses can be presented on the cost-effectiveness plane. The advantage of this form of presentation, compared to simply presenting incremental cost-effectiveness ratios, are two-fold. Firstly, a graphical approach allows the decision maker to compare the relative magnitude of cost and effect differences which are hidden within a single ratio. Secondly, the decision maker can clearly see the consequences of one intervention dominating another, either at baseline or as part of the sensitivity analysis, where incremental cost-effectiveness ratios are inappropriate. Figure 4 illustrates a typical sensitivity analysis on the cost-effectiveness plane. The horizontal axis shows the difference in effect between a new technology and conventional therapy and the vertical axis shows the difference in cost. The point **b** shows the baseline results; in this case the new intervention costs more and is associated with increased benefits. The uncertainty associated with the evaluation is shown using four separate one-way sensitivity analyses and their combined effect is shown using an extreme scenario analysis. The one-way sensitivity analysis is represented by a separate line for each parameter joining two extremes and passing through the baseline. The extreme scenario analysis is represented by two points representing best and worst case scenarios. The range of cost-effectiveness ratios implied by each type of analysis is indicated by the dotted lines.

The differing levels of robustness implied by extreme scenario analysis and simple sensitivity analysis are shown in Figure 4. Clearly, extreme scenario analysis will always give a wider representation of uncertainty than simple sensitivity analysis. Since the aim of presenting a sensitivity analysis is to improve decision makers' understanding of the implication of uncertainty for the results of the evaluation, the choice of sensitivity analysis to present should be governed by the information it imparts. The chance that all the extreme values of each parameter will occur simultaneously will be very small. It follows that the value of an extreme scenario analysis is greatest where it shows that the decision implied by the base case analysis does not change when extreme scenarios are considered. If a particular evaluation shows one intervention to dominate another at baseline, and this dominance is maintained under an extreme scenario analysis, then both the analyst and decision maker can be satisfied that the conclusions of the evaluation are robust. In such a case, the additional presentation of other types of sensitivity analysis are unlikely to provide the decision maker with additional information.

If on the other hand, the baseline results of the

**Figure 4** *A graphical presentation of sensitivity analysis*




analysis change from dominance to an incremental cost-effectiveness ratio, then (unless that ratio is clearly low compared to other reported ratios for alternative treatments) the cost-effectiveness decision is no longer clear and a simple sensitivity analysis should be carried out. If dominance is subsequently maintained for all parameters then the user can be sure that the results are robust to the simple sensitivity analysis. If, however, key parameters are identified which, when varied across their plausible range, result in the intervention of interest ceasing to dominate, a threshold analysis should be carried out to determine the critical values of those key parameters. Thus, the decision maker can clearly see that, although the intervention dominates at baseline, certain key parameters, were they to take the specified values, would result in the intervention ceasing to dominate. Additionally, extreme scenario analysis would give the upper bound on the incremental cost-effectiveness ratio, were the worst-case scenario to occur.

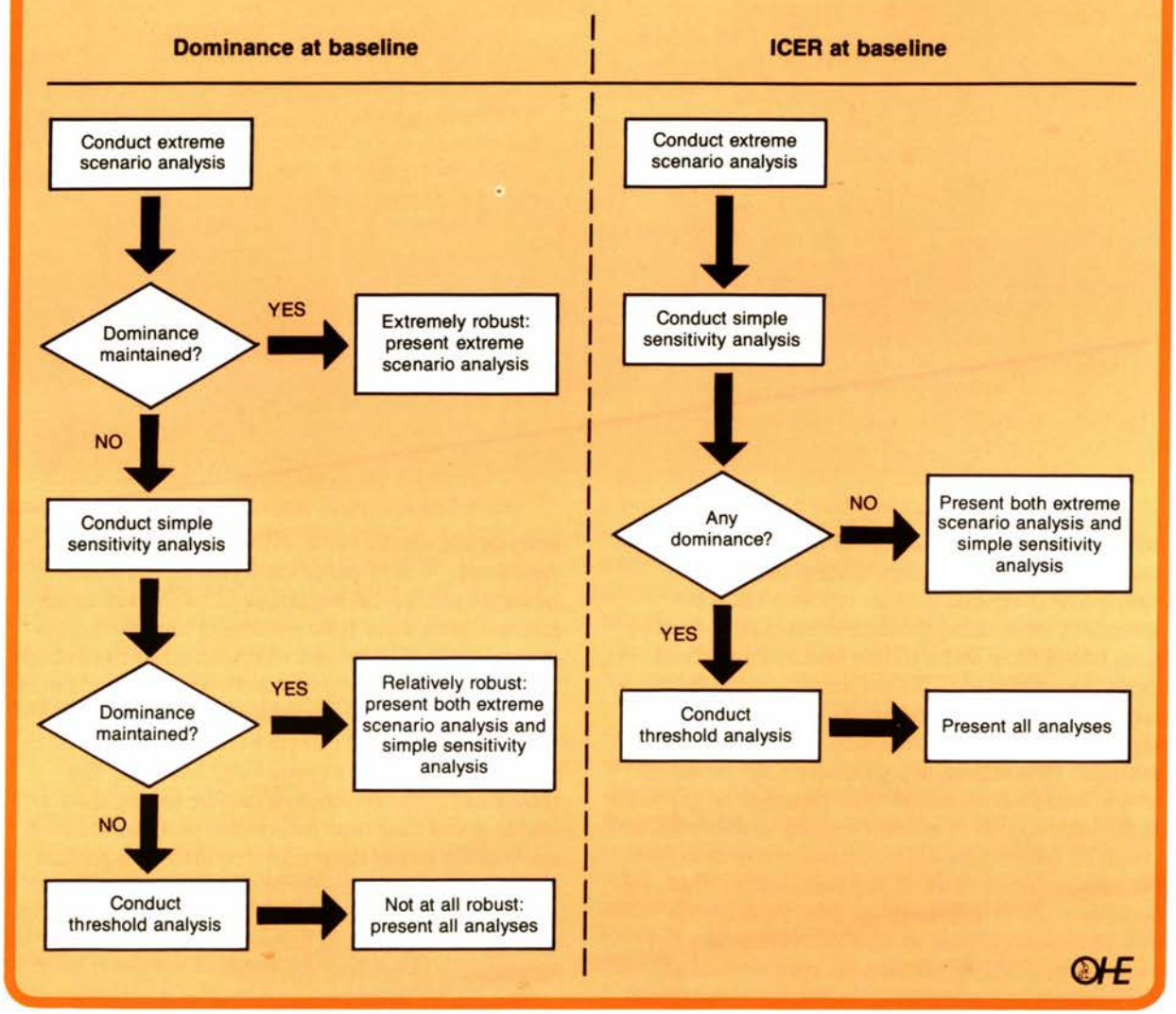
Where the baseline results indicate that an intervention is associated with both additional costs and additional benefits, both extreme scenario analysis and simple sensitivity analysis should be conducted. Where incremental cost-effectiveness ratios are concerned, the notion of robustness is not relevant since there is no maximum acceptable cost-effectiveness ratio against which the cost-effectiveness of individual interventions can be judged [Doubilet *et al.* 1986]. Users will, therefore, be interested in both the ranges of incremental cost-effectiveness ratios implied by each type of sensitivity analysis. The robustness of the results will then be for the users to decide given their own judgements as to the value of the benefits of the intervention in their own context. Threshold analysis should be presented (in a similar way to the case of dominance at baseline) when variation in a particular parameter causes the intervention to become dominant.

Sensitivity analyses undertaken in this fashion are only valid for the analytic methods employed in the economic evaluation. Handling uncertainty in analytic method is explored in text box five. The ideas presented in this section are illustrated in flow chart form in Figure 5.

**Box 5** *Uncertainty in analytic methods*

The analytical methods of a study are under the direct control of the analyst, therefore the uncertainty introduced by disagreement among economists as to the most appropriate methods is fundamentally different from other uncertainties discussed in this paper. In designing a study, the analysts will have chosen the relevant costs to include, the method by which utility data will have been calculated etc. Just as economists have differing opinions as to the most appropriate methods to employ it is likely that decision makers will also have different preferences for the methods they want to see employed. Where they hold different opinions to the authors of individual studies, it would be valuable for them to know the effect of using alternative methods for the results of that study. Potentially, at least, the evaluation could be conducted using any combination of the available methods. Where possible, analysts should recalculate the results of their evaluation using alternative methodologies. For example, studies could be presented with and without discounted benefits and including/excluding indirect costs. This would enable decision makers to decide for themselves the most appropriate methods and also to take into account that the conclusions of a study may/may not vary depending on the methodology adopted. If this practice were widespread, it would allow the comparison of like with like when examining incremental cost-effectiveness ratios from two different studies. Although ideally a complete sensitivity analysis should be conducted for each combination of methods employed, this approach is likely to be impractical given the publication constraints imposed by most journals. As guidelines for conducting economic evaluation are developed and implemented [CDHHCS, 1993; Detsky, 1993; ABPI, 1994] it is strongly advised that analysts use the recommended methods for their baseline analysis while presenting the effect of employing different methods in areas of methodological controversy.

**Figure 5** *Guidelines for presenting sensitivity analysis*




## Conclusions

Many economic evaluations simply present point estimates of cost-effectiveness ratios. In order that decision makers are fully informed analysts must consider the effects uncertainty may have for the results of their studies. In clinical evaluation, statistical analysis has been seen as the most appropriate method for handling uncertainty. However, in economic evaluation its role may be limited by the paucity of high quality data for the variables of interest, and also due to the fact that some areas of uncertainty simply do not involve sample data. For these reasons, this paper has concentrated on sensitivity analysis as a method for handling uncertainty in economic evaluation.

The methodology literature for economic evaluation has tended to imply that sensitivity analysis is a single technique for handling uncertainty. In fact, sensitivity analysis encompasses a number of different approaches. The problem is that these different techniques may imply differing levels of robustness for the conclusions of a study. It is therefore important that those reading and considering applying the results of economic evaluations are aware of the level of robustness implied by different methods. It is also important for analysts to be aware of the different approaches in order that they can impart the maximum amount of information to decision makers regarding the importance of uncertainty for the results of their study.

The results of any sensitivity analysis will depend crucially upon the ranges of values chosen for each variable. Too often in the past, such ranges have been chosen with little care. Unless the ranges chosen reflect a true estimate of the underlying uncertainty then the results of the sensitivity analysis may be, at best, misleading.

One of the major criticisms levelled at sensitivity analysis is that it is largely under the control of the analyst with the associated problems of bias and perverse incentives. In order to limit this criticism, analysts must strive to make the uncertainty in their work transparent and the users of such work must treat sceptically any results which have not been subjected to comprehensive and transparent examination using sensitivity analysis.

**Acknowledgements:** *The author would like to thank Adrian Towse, Martin Buxton and Bernie O'Brien for comments and suggestions on earlier drafts. The usual disclaimer applies.*

## References


ABPI. Pharmaceutical Industry and Department of Health agree guidelines for the economic analysis of medicines. Press Release 1994.

Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. Medical Care 1992;30(3): 231-243.

Anderson JP, Bush JW, Chen M, Dolenc M. Policy space areas and properties of benefit-cost/utility analysis. JAMA 1986; 255:794.

Black WC. The cost-effectiveness plane: a graphic representation of cost-effectiveness. Medical Decision Making 1990; 10: 212.

Brandeau ML, Owens DK, Sox C, Wachter RM. Screening women of childbearing age for human immunodefficiency virus: a cost-benefit analysis. Archives of Internal Medicine 1992;152:2229-2237.

Briggs, AH, Sculpher MJ. Sensitivity analysis in economic evaluation: a review of published studies. (Health Economics, forthcoming).

Briggs AH, Sculpher MJ, Buxton MJ. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health Economics 1994;3:95-104.

Buxton MJ, O'Brien BJ. Economic evaluation of odansetron: preliminary analysis using clinical trial data prior to price setting. British Journal of Cancer 1992; 66(supplement):S64-S67.

Cairns J. Discounting and health benefits: another perspective. Health Economics 1992;1:76-79.

CDHHCS (Commonwealth Department of Health Housing and Community Services). Guidelines for the pharmaceutical industry on preparation of submissions to the pharmaceutical benefits advisory committee. Canberra, Australia: Australian Government Publishing Service, 1992.

Cohen DJ, Breall JA, Ho KKL, Kuntz RE, Goldman L, Baim DS, Weinstein MC. Evaluating the potential cost-effectiveness of stenting as a treatment for symptomatic single-vessel coronary disease: use of a decision analytic model. Circulation 1994;89:1859-1874.

Coyle D, Tolley K. Discounting of health benefits in the pharmacoeconomic analysis of drug therapies: an issue for debate? PharmacoEconomics 1992;2(2):153-162.

Detsky A. Guidelines for preparation of economic evaluation of pharmaceutical products: a draft for Ontario and Canada. PharmacoEconomics 1993; 3: 354-361.

Doubilet P, Begg CB, Weinstein MC et al. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. Medical Decision Making 1985;5:157-77.

Doubilet P, Weinstein MC, McNeil BJ. Use and misuse of the term "cost-effective" in medicine. New England Journal of Medicine 1986; 314(4): 253-256.

Drummond MF. Comparing cost-effectiveness across countries: the model of acid-related disease. PharmacoEconomics 1994; 5(Suppl. 3): 60-67.

Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation in health care. International Journal of Technology Assessment in Health Care 1993;9:26-36.

Drummond MF, Davies L. Economic analysis alongside clinical trials. Revisiting the methodological issues. International Journal of Technology Assessment in Health Care 1991;7(4):561-573.

Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1987.

Eisenberg JM. Clinical economics. A guide to the economic analysis of clinical practices. Journal of the American Medical Association 1989;262(20):2879-2886.

Evans RG, Robinson GC. Surgical day care: measurements of the economic payoff. Canadian Medical Association Journal 1980;123:873-880.

Gerard K. Cost-utility in practice: A policy maker's guide to the state of the art. Health Policy 1992;21:249-279.

Katz DA, Welch HG. Discounting in cost-effectiveness analysis of health care programmes. PharmacoEconomics 1993;3(4):276-285.

Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. Canadian Medical Association Journal 1992;146(4):473-481.

Luce BR, Elixhauser A. Estimating costs in the economic evaluation of medical technologies. International Journal of Technology Assessment in Health Care 1990;6:57-75.

Magid D, Schwartz B, Craft J, Schwartz J. Prevention of lyme disease after tick bites: a cost-effectiveness analysis. New England Journal of Medicine 1992;327:534-541.

MacRae KD. Pragmatic versus explanatory trials. International Journal of Technology Assessment in Health Care 1989;5:333-339.

McNeil BJ, Dudley RA, Hoop B, et al. A cost-effectiveness analysis of screening for hepatitis B surface antigen in India. Medical Decision Making 1981;1:345-359.

O'Brien BJ, Drummond MF. Statistical versus quantitative significance in the socioeconomic evaluation of medicines. PharmacoEconomics 1994; 5(5): 389-398.

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. Medical Care 1994;32(2):150-163.

Parsonage M, Neuberger H. Discounting and health benefits. Health Economics 1992;1:71-79.

Petrou S, Malek M, Davey P.The reliability of cost-utility estimates on cost-per-QALY league tables. PharmacoEconomics 1993;3(5):354-361.

Schulman KA, Kinosian B, Jacobson TA et al. Reducing high blood cholestorol level with drugs: cost-effectiveness of pharmacologic management . Journal of the American Medical Association 1990; 264:3025-3033.

Schulman KA, Lynn LA, Glick HA, Eisenberg JM. Cost effectiveness of low-dose zidovudine therapy for asymptomatic patients with human immunodeficiency virus (HIV) infection. Annals of Internal Medicine 1991;114(9):798-801.

Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. Journal of Chronic Disease 1967;20:637-648.

Sheldon T. Discounting in health care decision making: time for a change? Journal of Public Health Medicine 1992;14:250-256.

Sonnenberg A. Costs of medical and surgical treatment of duodenal ulcer. Gastroenterology 1989;96:1445-1452.

Torrance GW. Measurement of health state utilities for economic appraisal. Journal of Health Economics 1986;5:1-30.

Udvarhelyi S, Colditz GA, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? Annals of Internal Medicine 1992;116:238-244.

Van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E ratios alongside a clinical trial. Health Economics 1994; 3:309-319.

Weinstein MC, Fineberg HV, Elstein AS, Frazier HS, Neuhauser D, Neutra RR, Mcneil BJ. Clinical Decision Analysis. Philidelphia: W.B. Saunders, 1980.

Weinstein MC. Economic assessments of medical practices and technologies. Medical Decision Making 1981;1(4):309-330.

Willan A, O'Brien B. Cost-effectiveness ratios in clinical trials: from deterministic to stochastic models. Proceedings of the American Statistical Society (forthcoming).

Wellcome Foundation Ltd. Economic Evaluation Bibliography. Health Economics 1992;1(Supplement).

## Office of Health Economics

The Office of Health Economics was founded in 1962 by the Association of the British Pharmaceutical Industry. Its terms of reference are:

- To undertake research on the economic aspects of medical care.
- To investigate other health and social problems.
- To collect data from other countries.
- To publish results, data and conclusions relevant to the above.


## Office of Health Economics

12 Whitehall London SW1A 2DY
Telephone: 0171-930 9203
Fax: 0171-747 1419

This briefing was prepared by Andrew Briggs
*Research Fellow, Health Economics Research Group (HERG), Brunel University, Uxbridge, Middlesex UB8 3PH.*


Printed by White Crescent Press Ltd