The design and analysis of stochastic cost-effectiveness studies for the evaluation of health care interventions

Andrew Briggs
Health Economic Research Centre
Institute of Health Sciences
University of Oxford

Magnus Tambour
Centre for Health Economics
Department of Economics
Stockholm School of Economics

Stockholm School of Economics
Working Paper Series in Economics and Finance No. 234
April 1998

Abstract

Many clinical trials are in progress which involve the collection of patient-level data on both the health outcome and resource use consequences of the health care interventions under evaluation. The overall aim of many such evaluations will be to undertake a cost-effectiveness analysis, which will often result in a cost-effectiveness ratio summarising the value for money of the intervention in question. In this paper, we explore the issues surrounding the design and analysis of such studies. At the design stage of an analysis, we propose an improved sample size formula for cost-effectiveness analysis that allows for covariance between cost and effect differences. This approach is based on the 'net benefits' approach to the analysis of uncertainty in cost-effectiveness analysis. At the analysis stage of an evaluation, we explore the differences and similarities of the 'net benefit' approach to analysing cost-effectiveness information and the traditional approach based on cost-effectiveness ratios. Despite the apparent differences, we show that the two approaches are exactly equivalent when it comes to estimating the probability that the intervention is cost-effective under alternative values of the ceiling cost-effectiveness ratio appropriate for decision-making purposes.

Keywords: Clinical trials, Cost-effectiveness, Economic evaluation, Sample size.

JEL-codes: C12, C90, I10.

Introduction

Clinical trials of alternative health care interventions are increasingly collecting patient-level data on both the health outcome and resource consequences of the interventions under evaluation. Where a new intervention is both more costly and more effective than an alternative, it is appropriate to consider whether the additional cost is justified by the additional health benefits produced. In such situations, it is the incremental cost effectiveness ratio (ICER) that summarises the value for money of the new intervention. Much attention in the recent literature has been devoted to problems of estimating confidence limits for ICERs at the data analysis stage of an evaluation. A number of commentators have highlighted that, at the design stage, sample size calculations should be based not only on the clinical endpoints of the trial but also on the economic endpoints (Drummond and O'Brien, 1993; O'Brien et al. 1994; Sacristan et al. 1995; Coyle, 1996). In a recent paper, it was demonstrated how a sample size formula for cost-effectiveness analysis can be derived from the independent confidence intervals on cost and effect differences (Briggs and Gray, 1998). One of the acknowledged weaknesses of the approach is that it fails to allow for the covariance between the cost and effect differences.

In this paper, we demonstrate how a sample size formula, based on the net-benefit approach to interpreting cost-effectiveness results (Tambour et al. 1998; Stinnett and Mullahy, 1998) allows for covariance and is therefore more efficient than that originally proposed by Briggs and Gray. It is shown that the sample size formula proposed by Briggs and Gray is a special case of the general formula presented in this paper. We then go on to demonstrate how, at the analysis stage, the net benefit approach to decision making is formally equivalent to decision-making based on the cost-effectiveness plane. We suggest that the appropriate way to represent uncertainty in the decision-making process is through cost-effectiveness acceptability curves (van Hout et al. 1994), rather than by confidence intervals around either net-benefit values or cost-effectiveness ratios.

Cost-effectiveness and decision rules

Suppose that two treatments are to be compared in a clinical trial setting. One treatment represents the currently provided (or control) therapy and the other represents an experimental (or treatment) therapy. The true (but unobservable) health outcome effects of the two therapies are denoted by μ_{ET} and μ_{EC} for the treatment and control therapies respectively. Similarly, the true (but unobservable) resource costs associated with the two treatments are denoted μ_{CT} and μ_{CC}. O'Brien and colleagues identify four situations that can arise in relation to the incremental cost and effectiveness of two therapies:(O'Brien et al. 1994)

1. $\mu_{CT} - \mu_{CC} < 0$; $\mu_{ET} - \mu_{EC} > 0$; *dominance* - accept new treatment as it is both cheaper and more effective than the existing therapy.
2. $\mu_{CT} - \mu_{CC} > 0$; $\mu_{ET} - \mu_{EC} < 0$; *dominance* - reject new treatment as it is both more expensive and less effective than the existing therapy.
3. $\mu_{CT} - \mu_{CC} > 0$; $\mu_{ET} - \mu_{EC} > 0$; trade-off - consider magnitude of the additional cost of the new therapy relative to its additional effectiveness.
4. $\mu_{CT} - \mu_{CC} < 0$; $\mu_{ET} - \mu_{EC} < 0$; trade-off - consider magnitude of the cost-saving of the new therapy relative to its reduced effectiveness.

These four situations correspond to the four quadrants of the cost-effectiveness (CE) plane, which has been proposed for presenting cost-effectiveness results (Anderson et al. 1986; Black, 1990). The CE plane is presented in Figure 1. Where one intervention is simultaneously cheaper and more effective than the other (situations 1 & 2 above and quadrants II & IV on the CE plane) it is clearly the treatment of choice since it dominates the alternative intervention. However, where one intervention is both more effective and more costly (situations 3 & 4 above and quadrants I & III on the cost-effectiveness plane) then the decision is no longer clear. Rather, a judgement must be made concerning whether the difference in effectiveness resulting from a switch in therapy is justified given the difference in costs that such a switch would bring about. In order to aid such judgement, an incremental cost-effectiveness ratio (ICER) should be calculated which provides a summary of the cost-effectiveness of one intervention relative to the other:

$$ICER = \frac{\mu_{CT} - \mu_{CC}}{\mu_{ET} - \mu_{EC}}$$

In the absence of the simple case of dominance, the decision over which is the appropriate treatment to implement cannot be determined unless a cut-off value, or maximal willingness to pay for additional health effect (R_c), has been specified (Karlsson and Johannesson, 1996). This ceiling value of the ICER can be represented by the (slope of the) dashed line on the CE plane of Figure 1. If the incremental costs and effects are plotted to the right of this line on the CE plane then the treatment is considered cost-effective, while points to the left of this line represent cost-ineffective interventions.[1]

In terms of the ICER, the decision rule is that the new treatment should be implemented instead of the standard treatment if:

$$\text{ICER} = \frac{\mu_{CT} - \mu_{CC}}{\mu_{ET} - \mu_{EC}} = \frac{\mu_{\Delta C}}{\mu_{\Delta E}} < R_c. \tag{1}$$

In a trial situation, on the basis of data collected from two groups of patients receiving the alternative therapies, the ICER can be estimated by:

$$\hat{R} = \frac{\overline{C}_T - \overline{C}_C}{\overline{E}_T - \overline{E}_C} = \frac{\Delta\overline{C}}{\Delta\overline{E}} \tag{2}$$

where $\overline{C}_T$ and $\overline{C}_C$ are the mean costs in the treatment and control arms of the trial respectively and $\overline{E}_T$ and $\overline{E}_C$ are the mean effects.

Due to the uncertainty associated with sampling variation, many would argue that the decision rule to implement the new treatment should be based on the requirement that the observed ICER from Equation 2 be significantly less than the ceiling cost-effectiveness ratio, R_c, appropriate for decision making purposes. In other words, it is appropriate to test the null hypothesis $H_0 : \mu_{\Delta C}/\mu_{\Delta E} > R_c$ against the alternative hypothesis $H_1 : \mu_{\Delta C}/\mu_{\Delta E} < R_c$.

Although the underlying cost and effect data may not follow a well-behaved distribution in general, the *central limit theorem* states that the sample mean costs and effects in Equation 2 will approach a normal distribution with increasing sample size. Hence, with sufficient sample size, the distribution of the relevant estimators can be approximated by normal distributions with the following mean and variance:

---

[1] In principle, a new treatment might be introduced if it was less effective than the existing treatment, but was sufficiently less costly to compensate for that loss of effect (i.e. to the right of the line representing the ceiling ratio in the third quadrant of the CE plane). In reality, such a policy might raise ethical objections from those who believe it unjust to withdraw a treatment that has formerly been available. Since, at a design stage, it is unlikely that a trial will be undertaken to show an experimental therapy to be less effective than an existing therapy, we assume that only new treatments that are thought to be more effective than existing treatments will progress to evaluation.

$$
\begin{aligned}
\overline{C}_C &\sim N(\mu_{CC}, \sigma^2_{CC} / n_C) \\
\overline{E}_C &\sim N(\mu_{EC}, \sigma^2_{EC} / n_C) \\
\overline{C}_T &\sim N(\mu_{CT}, \sigma^2_{CT} / n_T) \\
\overline{E}_T &\sim N(\mu_{ET}, \sigma^2_{ET} / n_T)
\end{aligned}
$$

where σ^2 represents the true population variance for (subscripts) costs (C) and effects (E) for the treatment (T) and control (C) groups.

As it is known that the sum of two normal variables is itself normally distributed, we can assume that the estimates of the incremental costs and effects, the numerator and denominator of the estimated ICER from Equation 2, are distributed:

$$
\begin{aligned}
\Delta\overline{E} &\sim N(\mu_{ET} - \mu_{EC}, \sigma^2_{\Delta E}) \\
\Delta\overline{C} &\sim N(\mu_{CT} - \mu_{CC}, \sigma^2_{\Delta C}).
\end{aligned}
$$

where:[2]

$$
\begin{aligned}
\sigma^2_{\Delta E} &= \frac{\sigma^2_{ET}}{n_T} + \frac{\sigma^2_{EC}}{n_C} \\
\sigma^2_{\Delta C} &= \frac{\sigma^2_{CT}}{n_T} + \frac{\sigma^2_{CC}}{n_C}.
\end{aligned} \qquad (3)
$$

As a ratio of two asymptotically normal variables, the ICER suffers from the problem that its moments may not be defined, due to the non-negligible probability that the denominator of the ratio could take a zero value. Given the intractable nature of the variance of the ratio, many papers have focussed on possible of estimating confidence limits for the ICER which do not involve on a measure of variance (O'Brien et al. 1994; Wakker and Klaassen, 1995; Chaudhary and Stearns, 1996; Willan and O'Brien, 1996).

More recently, two papers have highlighted the 'net-benefit' approach to handling uncertainty in cost-effectiveness analysis (Tambour et al. 1998; Stinnett and Mullahy, 1998). The decision rule of Equation 1 can be rearranged to give a measure of 'net-benefit' and an associated decision rule that the new therapy should be implemented only if the net-benefits are positive. Two alternative formulations of net-benefit have been suggested based on a

[2] Note that implicit in this assumption of the variance of the cost and effect differences in Equation 3 is that the data where generated from independent groups in a randomised trial. However, for non-randomised trial designs, such as, for example, a before and after study, the independence assumption may not be justified and the above expressions should incorporate a covariance term.

simple rearrangement of Equation 1, such that the new therapy should be implemented over the existing treatment if:

$$\mathrm{NB}_C = R_c \mu_{\Delta E} - \mu_{\Delta C} > 0 \tag{4}$$

or, equivalently, if:

$$\mathrm{NB}_H = \mu_{\Delta E} - \mu_{\Delta C} / R_c > 0 \tag{5}$$

In Equation 4, the net-benefits are measured in monetary terms (Tambour et al. 1998), while in Equation 5, the net-benefits are measured in terms of health (Stinnett and Mullahy, 1998).

The advantage of the net-benefits approach is that the $(1-\alpha)\%$ confidence interval for net-benefits can be easily determined in the standard fashion, as:

$$N\hat{B} \pm z_{\alpha/2}\sqrt{\sigma^2_{NB}}$$

where $N\hat{B}$, is the estimated net-benefit measure, with variance σ^2_{NB}, and $z_{\alpha/2}$ is the critical value from the standard normal distribution. In contrast to a ratio, whose variance may not be defined, the variance of net-benefits is simply a linear combination of two asymptotically normal variables and can therefore be defined as:

$$var(N\hat{B}_C) = R_c^2\, var(\Delta\bar{E}) + var(\Delta\bar{C}) - 2R_c\, cov(\Delta\bar{E}, \Delta\bar{C}) \tag{6}$$

in terms of the monetary net-benefit measure of equation 4, or:

$$var(N\hat{B}_H) = var(\Delta\bar{E}) + \frac{1}{R_c^2} var(\Delta\bar{C}) - \frac{2}{R_c} cov(\Delta\bar{E}, \Delta\bar{C}) \tag{7}$$

for the net health benefit measure of Equation 5.

Sample size for cost-effectiveness analysis

We now show how the net-benefits approach can be used to generate a sample size formula for cost-effectiveness analysis. In this section, the monetary version of net-benefits from equations 4 and 6 is employed. However, an exactly equivalent formula, which generates the same sample size can be derived using the net health benefit approach from Equations 5 and 7.

An observed net benefit is significantly positive providing:

$$(R_c\Delta\bar{E} - \Delta\bar{C}) - z_{\alpha/2}\sqrt{\sigma^2_{NB}} > 0 .$$

Although it is tempting to base the sample size calculation on the numbers of patients required to show an observed difference as significant sample size calculations should be based on the hypothesised cost and effect differences (denoted $\Delta\tilde{E}, \Delta\tilde{C}$) such that the study has the appropriate power to detect the net benefit as different from zero. In algebraic terms:

$$(R_c\Delta\tilde{E} - \Delta\tilde{C}) - z_\beta\sqrt{\sigma^2_{NB}} > z_{\alpha/2}\sqrt{\sigma^2_{NB}} \quad (8)$$

where z_β is the critical value from the standard normal distribution corresponding to a required power of $1-\beta$.

Noting that the $cov(x, y) = \rho\sqrt{var(x)var(y)}$ (where ρ is the correlation coefficient) the variance of net benefits from Equation 6 can be expressed in terms of the variance of the cost and effect differences from Equation 3:

$$var(N\hat{B}_C) = R_c^2\left(\frac{\sigma^2_{ET}}{n_T} + \frac{\sigma^2_{EC}}{n_C}\right) + \left(\frac{\sigma^2_{CT}}{n_T} + \frac{\sigma^2_{CC}}{n_C}\right) - 2R_c\rho\sqrt{\left(\frac{\sigma^2_{ET}}{n_T} + \frac{\sigma^2_{EC}}{n_C}\right)\cdot\left(\frac{\sigma^2_{CT}}{n_T} + \frac{\sigma^2_{CC}}{n_C}\right)}$$

Substituting the above expression for the variance of net benefits into Equation 8 and assuming equal sample sizes for each arm of the trial gives the sample size formula (after rearranging on n) as:

$$n > \frac{(z_{\alpha/2} + z_\beta)^2\left[R_c^2(\sigma^2_{ET} + \sigma^2_{EC}) + (\sigma^2_{CT} + \sigma^2_{CC}) - 2R_c\rho\sqrt{(\sigma^2_{ET} + \sigma^2_{EC})(\sigma^2_{CT} + \sigma^2_{CC})}\right]}{(R_c\Delta\tilde{E} - \Delta\tilde{C})^2} \quad (9)$$

It is straightforward to show (see appendix) that the sample size formula generated by Briggs and Gray (1998) is simply a special case of the general formula in equation 9 above when the correlation between cost and effect differences is set to -1. This is because the sample size formula in their paper was based on the 'confidence box' approach to confidence interval estimation for an ICER (O'Brien et al. 1994) which implicitly assumes that cost and effect differences are perfectly negatively correlated such that the combination of the 95% confidence limits on cost and effect differences generates a 95% confidence limit for the ICER. Where cost and effect differences are not perfectly correlated, the confidence box method will produce a confidence interval for the ratio that is greater than the nominal 95% level. Moreover, the required sample size will be overstated when using the confidence box method if $\rho > -1$, since it can be shown that the sample size formula is a decreasing function of the correlation (see appendix).

A numerical example: the design stage

In order to illustrate the sample size calculations outlined above, we present a numerical example based on the same as that employed by Briggs and Gray (1998). Their example assumed that a randomised trial was planned, based on an effect difference in favour of a new therapy of 0.8 life-years at an additional cost over standard therapy of £1,200 (i.e., a hypothesised ICER of £1,200/0.8 = £1,500 per life year gained). The assumed standard deviations for the effects and costs (assuming they were the same in each arm) were 4.04 and 8,700 respectively. First we assume that cost and effect differences are independent, such that the correlation coefficient is zero and the correlation term from Equation 9 can be ignored. In order to employ the power calculation formula of Equation 9, we have only to estimate the appropriate cost-effectiveness ceiling ratio from which we want to have the power to be able to detect the observed cost-effectiveness ratio as significantly different from. Since no such universally accepted ceiling ratio exists for decision-making purposes we suggest that analysts plot the sample size requirements as a function of the maximum cost effectiveness ratio, for different levels of study power (and perhaps even significance). Figure 2 shows such a presentation for the example detailed above and is directly comparable to Figure 7 of the Briggs and Gray paper. Sample size requirements are shown for values of the maximum cost-effectiveness ratio between approximately £3,000 per life year and £30,000 per life year and for study power of 50-90%, assuming conventional 5% significance. Also shown is the sample size based only on the effectiveness outcome (the horizontal line at n=536) and the equivalent 90% power sample size calculation (dotted line) based on the Briggs and Gray sample size formula (i.e., assuming that the correlation is equal to -1).

From Figure 2 it appears that the sample size required for a trial is a decreasing function of R_c. Intuitively, this seems reasonable since, a higher ceiling ratio indicates a greater willingness to pay for health benefits which in turn implies a greater net-benefit suggesting less observations are required to establish the net-benefits as significant. However, this is not always true. Consider Figure 3 which shows the sample size requirements based on different assumptions concerning the correlation between cost and effect differences for the example, based on 90% power and 5% significance. The figure shows that very strong positive correlation between cost and effect differences may actually cause the sample size to be an increasing function of R_c in the example. In general, whether sample size is increasing or decreasing in R_c will depend on both the sign and the magnitude of the covariance. It can be shown that there will be a critical value of the correlation between cost and effect that will mean that the sample size requirement for the clinical study and the sample size requirement for an economic study based on the same effectiveness measure will be equal (see appendix)[3]. For example, it is clear from Figure 3 that the sample size for economic evaluation will exceed that for clinical evaluation unless the correlation coefficient is very high. In general, it is likely that sample sizes for economic analysis will exceed those for clinical analysis where a new treatment is hypothesised to be both more effective, but also more expensive. The extremely high correlation coefficients required to make sample size increasing in R_c are unlikely to be observed in practical application.

Cost-effectiveness ratios or net-benefits?

Using the net-benefit approach at the design stage would seem to imply that at an analysis stage the outcome of interest in the economic analysis is the net-benefit measure rather than the cost-effectiveness ratio. While some authors have suggested the net-benefit approach as a way to avoid the problems associated with confidence intervals around ratios (Tambour et al. 1998) others have gone further, suggesting that net (health) benefits themselves may be a more appropriate framework in which to examine the results of cost-effectiveness analyses (Stinnett and Mullahy, 1998). They list the main advantages of the net (health) benefits approach over cost-effectiveness ratios as improved interpretability, more straightforward statistical inference, and clearer analysis of multiple comparisons. It is not our purpose to

[3] Since the correlation coefficient must lie between -1 and +1 by definition, critical values of the coefficient outside of this range indicate that there is no covariance between cost and effect differences that would lead to the sample size requirements for the clinical and economic evaluations in question being equal.

rehearse those arguments here, indeed, in the main we agree with the authors that there can be significant problems associated with the unthinking interpretation of the cost-effectiveness ratio. However, few would argue the results of cost-effectiveness analysis be presented simply in terms of ICER. Indeed, many commentators have advocated the presentation of the results of economic evaluations on the cost-effectiveness plane (Anderson et al. 1986; Black, 1990) as outlined above and shown in Figure 1.

The presentation of the net (health) benefits by Stinnett and Mullahy leaves the impression that the approach is different from cost-effectiveness ratios in some important respect. However, we would argue that the only practical way to interpret net benefit measures (whether cost or health based) is in terms of the cost-effectiveness ratio. We show how, in terms of decision-making, a net benefit approach is equivalent to decision making on the cost-effectiveness plane where the summary results are presented in terms of ratios, by means of an example.

A numerical example: the analysis stage

Consider Table 1, which presents the results from a (hypothetical) clinical trial in terms of the cost and effects of two alternative interventions. As the new treatment is both more effective and more costly it is necessary to calculate an incremental cost-effectiveness ratio. This shows the cost-effectiveness of the new treatment relative to the control treatment as £4,836 per unit of health effect. Whether the new treatment should be implemented depends on the maximum willingness to pay for additional health effect, R_c. Suppose this is known to be £10,000. It would appear that the new treatment should be implemented since its cost-effectiveness is below the ceiling ratio. Table 1 also shows the net benefit values estimated using the willingness to pay figure of £10,000 per unit of health effect. This puts the net-benefit of implementing the new treatment at 0.38 units of health effect (from Equation 5) or, equivalently, £3,806 (from Equation 4).

However, this analysis is based purely on point estimates of observed cost and effect. In order to allow for sampling variation, confidence limits for the ICER and net-benefit statistics could be calculated and the new treatment only implemented if the treatment is found to be significantly cost-effective. An undoubted advantage of the net-benefit statistics are that they have a mathematically tractable variance and have asymptotically normal sampling distribution, which means that for sufficient sample sizes, the confidence intervals are straightforward to calculate using standard methods. By contrast, the variance of a ratio statistic is intractable and its sampling distribution is undefined. Although parametric methods, such as Fieller's theorem, have been suggested for estimating confidence limits for ratio statistics, in this example we employ the non-parametric method of bootstrapping since this allows us to 'visualise' uncertainty due to sampling variation more clearly. The bootstrap approach has been advocated by a number of commentators for estimating confidence limits for ICERs (O'Brien et al. 1994; Chaudhary and Stearns, 1996; Obenchain et al. 1997; Briggs et al. 1997) and has also been suggested by advocates of the net-benefit approach.

The bootstrap approach can be characterised as a simple four step process:

1. Sample with replacement n_T cost/effect pairs from the sample of patients who received the new treatment and calculate the bootstrap estimates $\overline{C}_T^*$ and $\overline{E}_T^*$ for the bootstrap sample;
2. Sample with replacement n_C cost/effect pairs from the sample of patients receiving the control treatment and calculate the bootstrap estimates $\overline{C}_C^*$ and $\overline{E}_C^*$ for the bootstrap sample;
3. On the basis of these four means, estimate the cost and effect differences, $\Delta\overline{C}^*$ and $\Delta\overline{E}^*$.
4. On the basis of these cost and effect differences calculate either the bootstrap ICER or the bootstrap net-benefit.

This process is then repeated a large number of times to build up the empirical estimate of the sampling distribution of the relevant statistic. Straightforward confidence limits can then be estimated using the 2.5^{th} and 97.5^{th} percentiles of the ordered list of bootstrap replications[4].

The confidence limits based on the bootstrap percentile approach are reproduced in Table 1 adjacent to the appropriate statistic. Since both the net-benefit intervals include zero and the upper limit of the ICER confidence interval is greater than £10,000, it cannot be concluded that the new treatment is cost-effective at the conventional 5% significant level. However, suppose a decision has to be made concerning which of the two treatments should be provided for the patients in question. Given that the baseline results from Table 1 indicate that the new treatment is cost-effective, the decision maker may be interested in the overall

[4] Alternatively, more efficient methods of calculating confidence intervals, such as the bias-corrected (*BC*) and bias-corrected-accelerated (BC_a) methods (Efron and Tibshirani, 1993) can be applied. However, these are not discussed here in order to keep the example simple and because the need for these corrections comes from the problems associated with confidence intervals for ratio estimators.

confidence that can be attached to a decision to implement the new therapy, even if that confidence does not reach conventional levels.

The cost and effect differences, from 1,000 repetitions of the first three steps of the bootstrap process based on the raw data summarised in Table 1, are presented in Figure 4. Note that in this example, the bootstrap replications fall into three of the four quadrants of the cost-effectiveness plane. In quadrant II and quadrant IV of the cost-effectiveness plane, the bootstrap replications of cost and effect will generate negative ratios. However, the negative ratios in quadrant II are associated with the new treatment dominating the old, whereas the negative ratios in quadrant IV are associated with the old treatment dominating the new treatment. However, in a rank ordering of cost-effectiveness ratios - all negative ratios are grouped together and are not distinguished by which quadrant of the cost-effectiveness plane they fall into. This illustrates the problems associated with negative ratios and gives us good reason to be wary of bootstrap based confidence limits when some replications fall into the IV quadrant of the plane (i.e., when effect differences are negative).

This problem does not arise with the net-benefit approach. Positive net benefits are favourable to the new treatment and negative net benefits are prejudicial to the new treatment. Hence, simply calculating the proportion of positive net-benefit observations in the bootstrap replications gives an estimate of the probability that can be given to the new treatment being cost-effective. In the example illustrated above, this probability is 82.4%. Hence, even though the confidence limits for net-benefits include zero at the 5% significance level, a decision-maker may decide that an 82.4% probability that the new treatment is cost-effective gives sufficient confidence in the new treatment to recommend implementation.

However, if we recognise the need to base decision-making on the cost-effectiveness plane, rather than purely on confidence intervals for cost-effectiveness ratios, we get the same results as the net-benefit approach. Figure 4 shows 19 of the bootstrap replications in quadrant IV, favouring the control treatment, and 20 in quadrant II favouring the new treatment. The remaining 961 bootstrap replications are in quadrant I and are associated with a positive bootstrap ICER. Of these positive ICERs, 804 are below the ceiling cost-effectiveness ratio of £10,000. Therefore, (804+20)/1000 = 82.4% of the bootstrap replications lie to the right of the line with slope equal to the ceiling ratio.

So far, we have assumed that the maximum willingness to pay for additional health effect is known. However, as the advocates of the net-benefit approach recognise, in fact considerable uncertainty remains concerning the ceiling ratio. One approach is to plot the net-benefit, together with confidence limits, as a function of the ceiling ratio (Stinnett and

Mullahy, 1998). An alternative, and we believe more informative approach, is to present the cost-effectiveness acceptability curve, first advocated by Van Hout and colleagues (1994). This shows the probability that the treatment under consideration is cost-effective as a function of the ceiling ratio. Figure 5 shows the cost-effectiveness acceptability curve for the above example based on the bootstrap replications of the cost and effect differences. Note that the cost-effectiveness acceptability curve is exactly the same, no matter whether it is calculated using the net-benefit or the cost-effectiveness plane approach. This is because the decision rules for both approaches are the same. Hence, if we move away from the debate about confidence intervals (either for net-benefits or ICERs) and accept that cost-effectiveness acceptability curves generate the appropriate information for decision-makers, it does not matter which approach we adopt.

Conclusions

The net-benefit approach to cost-effectiveness analysis is a powerful approach. The properties of the net-benefit statistic mean that it has many practical advantages over the ICER statistic when it comes to representing uncertainty in the form of confidence intervals. We show how, at a design stage of a trial intending to collect both health outcome effect and resource cost data, the net-benefits approach can be employed to generate a more efficient sample-size formula for cost-effectiveness analysis, and that the previously suggested formula based on cost-effectiveness ratios is simply a special case of our more general formula. At the analysis stage of an economic analysis, it is natural to report point estimates of cost-effectiveness using the ICER rather than net-benefits as we believe this statistic remains the most straightforward to interpret. Rather than presenting confidence intervals for either cost-effectiveness ratios or net-benefits, we believe that cost-effectiveness acceptability curves offer more information to the decision-maker, since they make no assumptions concerning the appropriate level of significance or the maximum-willingness for an additional unit of health effect appropriate for decision-making purposes. Using the non-parametric approach of bootstrapping, we show that cost-effectiveness acceptability curves can equivalently be calculated using either a net-benefit approach or a cost-effectiveness approach to decision-making. However, due to the desirable properties of the net-benefit statistic, the calculation of cost-effectiveness acceptability curves is much more straightforward using a net-benefits approach.

Appendix

Proposition 1: *The Briggs and Gray sample size formula is a special case of our more general formula*
Setting the correlation coefficient in Equation 9 to -1 gives:

$$n > \frac{(z_{\alpha/2} + z_\beta)^2 \left[R_c^2(\sigma_{ET}^2 + \sigma_{EC}^2) + (\sigma_{CT}^2 + \sigma_{CC}^2) + 2R_c\sqrt{(\sigma_{ET}^2 + \sigma_{EC}^2)(\sigma_{CT}^2 + \sigma_{CC}^2)}\right]}{(R_c\Delta\tilde{E} - \Delta\tilde{C})^2}$$

The term in the square brackets is simply the square of $R_c\sqrt{\sigma_{ET}^2 + \sigma_{EC}^2} + \sqrt{\sigma_{CT}^2 + \sigma_{CC}^2}$ simplifying the expression to:

$$n > \frac{(z_{\alpha/2} + z_\beta)^2 \left[R_c\sqrt{\sigma_{ET}^2 + \sigma_{EC}^2} + \sqrt{\sigma_{CT}^2 + \sigma_{CC}^2}\right]^2}{(R_c\Delta\tilde{E} - \Delta\tilde{C})^2}$$

which is exactly the same as Equation 9 in the paper by Briggs and Gray (1998).

Proposition 2: *The required sample size is decreasing in the correlation*
Treating the sample size formula of Equation 9 as an identity and differentiating with respect to the correlation coefficient gives:

$$\frac{d}{d\rho}(n) = -\frac{2(z_{\alpha/2} + z_\beta)^2 R_c\sqrt{(\sigma_{ET}^2 + \sigma_{EC}^2)(\sigma_{CT}^2 + \sigma_{CC}^2)}}{(R_c\Delta\tilde{E} - \Delta\tilde{C})^2}.$$

Since R_c, the power and significance levels, the variances and the denominator term must all be positive, the partial derivative above is always negative and the sample size is therefore a decreasing function of the correlation coefficient.

Proposition 3: *The required sample size may be increasing in R_c*

Treating the sample size formula of Equation 9 as an identity and differentiating with respect to R_c gives:

$$\frac{d}{dR_c}(n) = 2(z_{\alpha/2} + z_\beta)^2 \left[\frac{R_c(\sigma_{ET}^2 + \sigma_{EC}^2) - \rho\sqrt{(\sigma_{ET}^2 + \sigma_{EC}^2)}\sqrt{(\sigma_{CT}^2 + \sigma_{CC}^2)}}{(R_c\Delta\tilde{E} - \Delta\tilde{C})^2} - \frac{R_c^2(\sigma_{ET}^2 + \sigma_{EC}^2) + (\sigma_{CT}^2 + \sigma_{CC}^2) - 2R_c\rho\sqrt{(\sigma_{ET}^2 + \sigma_{EC}^2)}\sqrt{(\sigma_{CT}^2 + \sigma_{CC}^2)}}{(R_c\Delta\tilde{E} - \Delta\tilde{C})^3} \cdot \Delta E \right]$$

Setting this partial derivative to zero and rearranging on the correlation coefficient gives:

$$\rho = \frac{R_c \dfrac{\Delta\tilde{C}}{\Delta\tilde{E}}(\sigma_{ET}^2 + \sigma_{EC}^2) + (\sigma_{CT}^2 + \sigma_{CC}^2)}{\left(R_c + \dfrac{\Delta\tilde{C}}{\Delta\tilde{E}}\right)\sqrt{(\sigma_{ET}^2 + \sigma_{EC}^2)}\sqrt{(\sigma_{CT}^2 + \sigma_{CC}^2)}}$$

Hence, this expression gives the value for the correlation coefficient at which the sample sizes generated by a formula based just on effectiveness and a formula based on cost-effectiveness are equal (for the same effectiveness measure). Where the correlation coefficient is greater than that given by the identity above, the sample size will be an increasing function of the ceiling ratio. Of course, since $-1 < \rho < 1$, it will not always be the case that there exists a correlation coefficient such that the sample sizes required for cost-effectiveness analysis and clinical evaluation are equal.

References

Anderson, J.P., Bush, J.W., Chen, M. and Dolenc, D. (1986) Policy space areas and properties of benefit-cost/utility analysis. *JAMA* **255**, 794-795.

Black, W.C. (1990) The CE plane: A graphic representation of cost-effectiveness. *Medical Decision Making* **10**, 212-214.

Briggs, A.H. and Gray, A.M. (1998) Sample size and power calculations for stochastic cost-effectiveness analysis. *Medical Decision Making* **18**; S81-S92.

Briggs, A.H., Wonderling, D.E. and Mooney, C.Z. (1997) Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* **6**, 327-340.

Chaudhary, M.A. and Stearns, S.C. (1996) Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Statistics in Medicine* **15**, 1447-1458.

Coyle, D. (1996) Statistical analysis in pharmacoeconomic studies. *PharmacoEconomics* **9**, 506-516.

Drummond, M. and O'Brien, B. (1993) Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. *Health Economics* **2**, 205-212.

Efron, B. and Tibshirani, R. (1993) *An introduction to the bootstrap*, New York: Chapman & Hall.

Karlsson, G. and Johannesson, M. (1996) The decision rules of cost-effectiveness analysis. *PharmacoEconomics* **9**, 113-120.

O'Brien, B.J., Drummond, M.F., Labelle, R.J. and Willan, A. (1994) In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* **32**, 150-163.

Obenchain, R.L., Melfi, C.A., Croghan, T.W. and Buesching, D.P. (1997) Bootstrap analyses of cost effectiveness in antidepressant pharmacotherapy. *PharmacoEconomics* **11**, 464-472.

Sacristan, J.A., Day, S.J., Navarro, O., Ramos, J. and Hernandez, J.M. (1995) Use of confidence intervals and sample size calculations in health economic studies. *Annals of Pharmacotherapy* **29**, 719-725.

Stinnett, A.A. and Mullahy, J. (1998) Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* **18**, S65-S80

Tambour, M., Zethraeus, N. and Johannesson, M. (1998) A note of confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care* (forthcoming).

van Hout, B.A., Al, M.J., Gordon, G.S. and Rutten, F.F. (1994) Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics* **3**, 309-319.

Wakker, P. and Klaassen, M. (1995) Confidence intervals for cost-effectiveness ratios. *Health Economics* **4**, 373-382.

Willan, A.R. and O'Brien, B.J. (1996) Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Economics* **5**, 297-305.

Table 1
(Hypothetical) Results from a trial based economic analysis

	Control Group		Treatment Group		Difference	
Statistics	Effect	Cost	Effect	Cost	Effect	Cost
n	20	20	20	20	n/a	n/a
Mean	8.89	£ 25,080	9.62	£28,645	0.74	£ 3,564
St.dev.	1.14	£ 4,734	1.02	£ 6,355	n/a	n/a
S.E.	0.26	£ 1,058	0.23	£ 1,421	0.34	£ 1,772
ICER:	£ 4,836 (95%CI: -£772 to £24,646)**					
NB_H*	0.38 (95%CI: -0.36 to 1.06)**					
NB_C*	£3,806 (95%CI: -£3,644 to £10,563)**					

*Assuming maximum willingness to pay of £10,000 per unit of health effect
**Based on the 2.5th and 97.5th centiles of 1,000 bootstrap replications

Figure 1
Decision rules and the cost-effectiveness plane

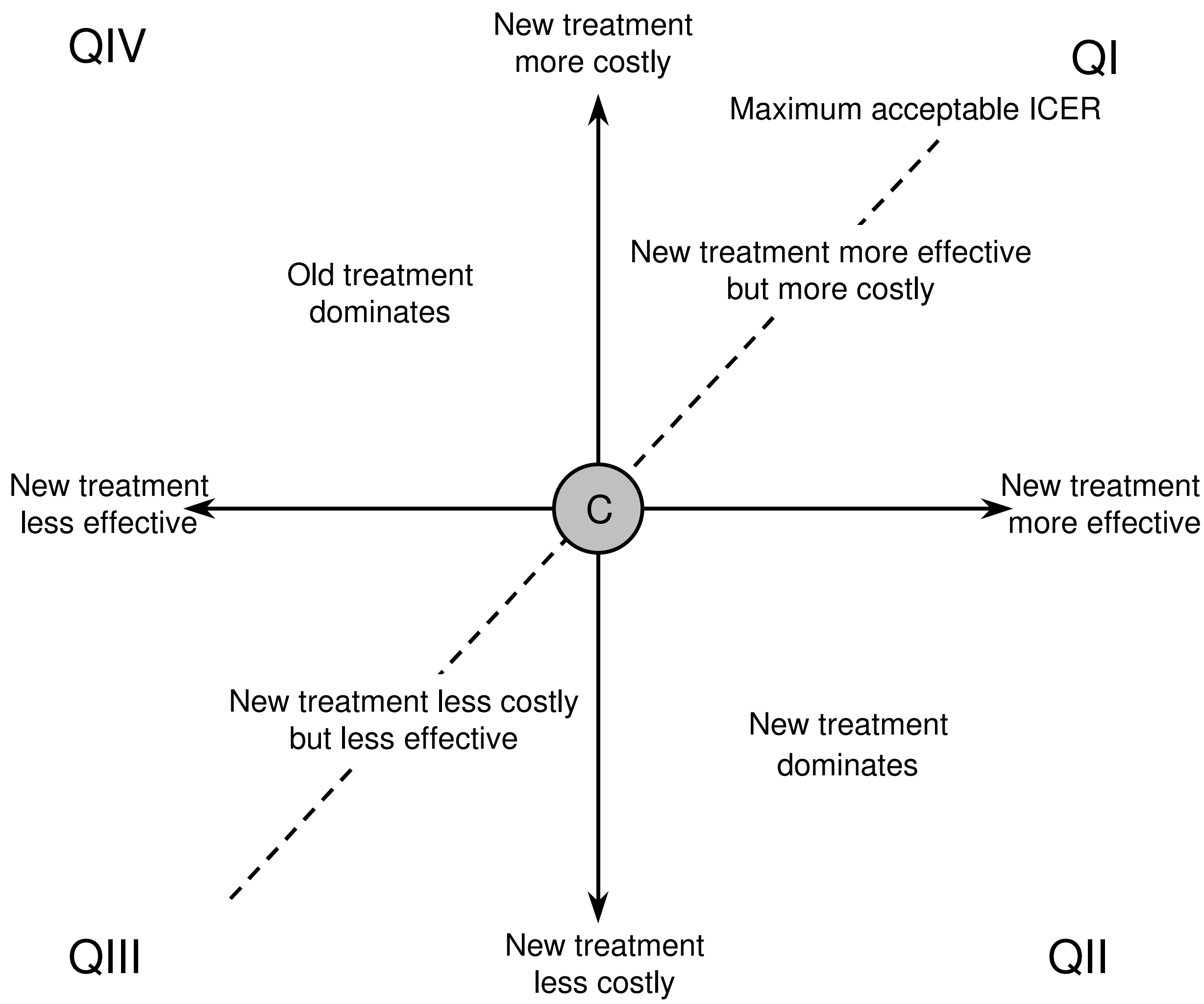

Figure 2
Power and sample size by the willingness to pay for additional health effect (at 5% significance)

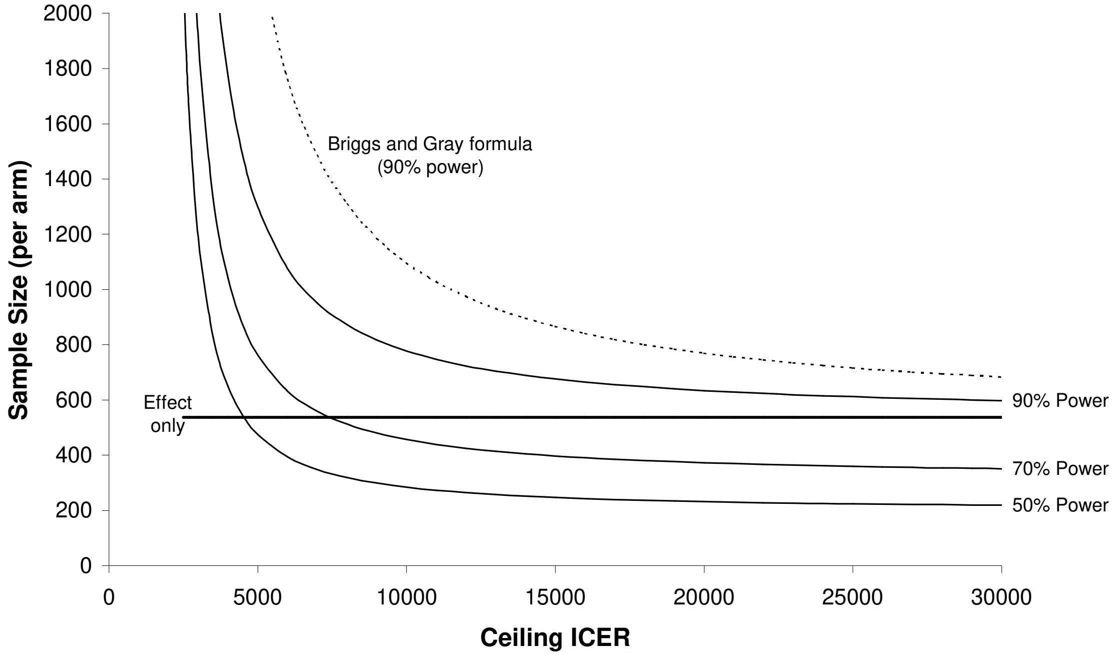

Figure 3
Sample size by willingness to pay for additional health effect for different values of the correlation between cost and effect

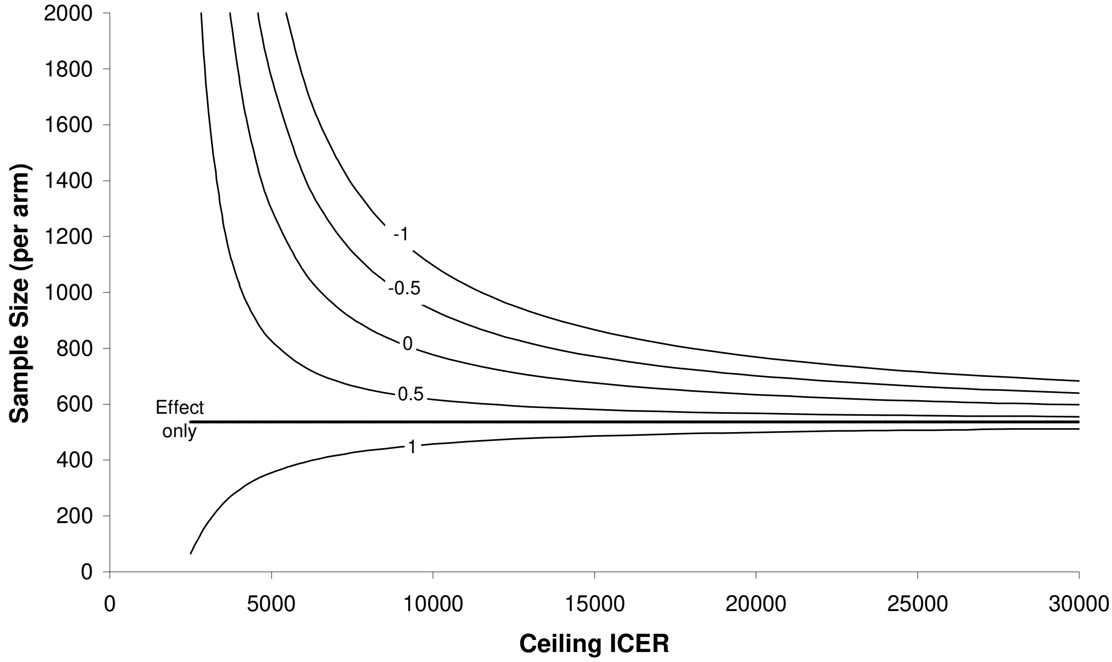

Figure 4

1,000 bootstrap replications of cost and effect differences on the cost-effectiveness plane

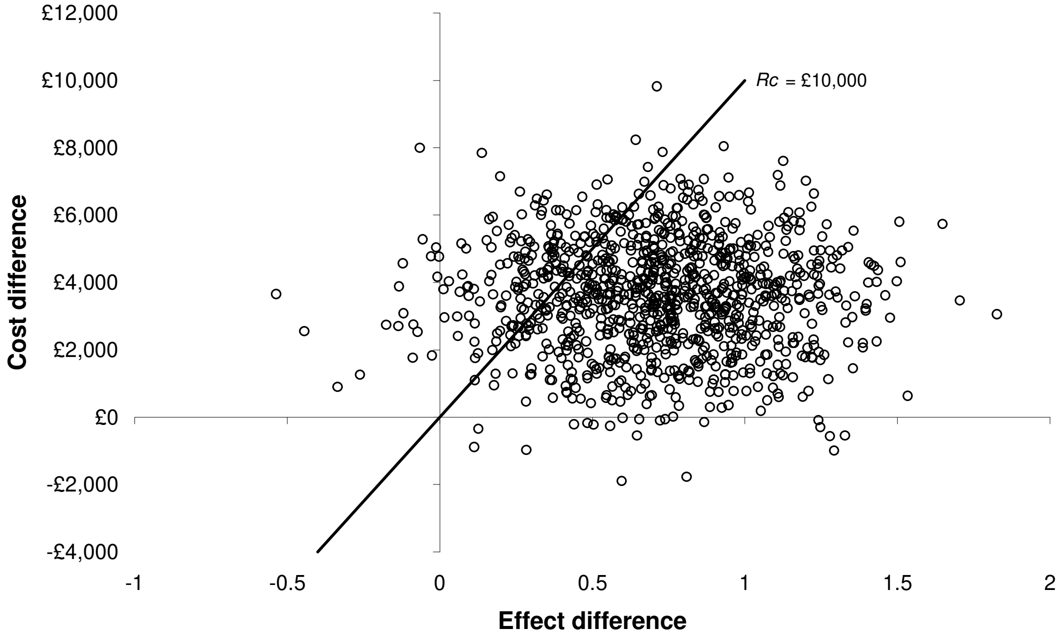

Figure 5
Cost-effectiveness acceptability curve showing the probability that the new intervention is cost-effective as a function of the ceiling ratio