



Briggs, A.H. and O'Brien, B.J. and Blackhouse, G. (2002) Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual Review of Public Health* 23:pp. 377-401.

<http://eprints.gla.ac.uk/4146/>

Deposited on: 8 May 2008

THINKING OUTSIDE THE BOX: Recent Advances in the Analysis and Presentation of Uncertainty in Cost-Effectiveness Studies

Andrew H. Briggs,¹ Bernie J. O'Brien,^{2,3} and Gordon Blackhouse^{2,3}

¹Health Economics Research Centre, Department of Public Health, University of Oxford, Oxford OX3 7LF, United Kingdom; e-mail: andrew.briggs@ihs.ox.ac.uk; ²Centre for Evaluation of Medicines, St. Joseph's Hospital, Hamilton, Ontario L8N 4A6, Canada; and ³Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton L8N 3Z5, Ontario, Canada

INTRODUCTION

The past decade has seen a rapid increase in the use of clinical trials as a vehicle for collecting economic information and estimating the cost-effectiveness of interventions (43). The existence of patient-level information on both costs and effects from clinical trials has generated interest in statistical methods for cost-effectiveness analysis, with a key focus on the quantification and presentation of uncertainty. This paper reviews recent developments and provides an overview of the state-of-the-art of quantitative methods for cost-effectiveness analysis.

A key structural feature of the paper is the use of a common example to illustrate the various methodological issues and techniques that are discussed. We

have chosen to use our own work on the cost-effectiveness of the implantable cardioverter defibrillator (ICD) versus drug therapy for patients at high risk of sudden cardiac death. This study was chosen for two reasons: First, it illustrates many of the challenging analytical aspects of contemporary trial-based cost-effectiveness analysis, and second, we have the data! A brief summary of the published cost-effectiveness analysis (36) is presented in Box 1. A further published analysis of cost-effectiveness by risk strata (47) is summarized in Box 2.

Box 1

Example: Cost-effectiveness of the implantable cardioverter defibrillator*

Background: In the Canadian Implantable Defibrillator Study (CIDS) we assessed the cost-effectiveness of the implantable cardioverter defibrillator (ICD) in reducing the risk of death in survivors of previous ventricular tachycardia (VT) or fibrillation (VF).

Methods: Health care resource use was collected prospectively on the first 430 patients enrolled in CIDS ($n = 212$ ICD, $n = 218$ amiodarone). Mean cost per patient, adjusted for censoring, was computed for each group based on initial therapy assignment. Incremental cost-effectiveness of ICD therapy was computed as the ratio of the difference (ICD–amiodarone) in cost to the difference in life expectancy (both discounted at 3% per year). All costs are in 1999 Canadian dollars; C\$1 \approx US \$0.65.

Results: Over 6.3 years, mean cost per patient in the ICD group was C\$87,715 versus C\$38,600 in the amiodarone group (difference C\$49,115; 95% CI C\$41,597 to C\$56,593). Life expectancy for the ICD group was 4.58 years versus 4.35 years for amiodarone (difference 0.23, 95% CI -0.12 to 0.57), for incremental cost-effectiveness of ICD therapy of C\$213,543 per life-year gained.

Box 2

Example: Effect of clinical risk stratification on cost-effectiveness of the implantable cardioverter-defibrillator. The Canadian Implantable Defibrillator Study**

Background: Three randomized clinical trials showed that implantable cardioverter-defibrillators (ICDs) reduce the risk of death in survivors of ventricular tachyarrhythmias, but the cost per year of life gained is high. A substudy of the Canadian Implantable Defibrillator Study (CIDS) showed that 3 clinical factors, age ≥ 70 years, left ventricular ejection fraction $\leq 35\%$, and New York Heart Association class III, predicted the risk of death and benefit from the ICD. We estimated the extent to which selecting patients for ICD therapy based on these risk factors makes ICD therapy more economically attractive.

Methods: Patients in CIDS were grouped according to whether they had 2 or more of 3 risk factors. Incremental cost-effectiveness of ICD therapy was computed as the ratio of the difference in mean cost to the difference in life expectancy between the 2 groups.

*Source: Abridged abstract from Reference (36).

**Source: Abridged abstract from Reference (47).

Results: Over 6.3 years, the mean cost per patient in the ICD group was Canadian (C) \$87,715 versus \$38,600 in the amiodarone group (C\$1 \approx US\$0.67). Life expectancy for the ICD group was 4.58 years versus 4.35 years for amiodarone, for an incremental cost-effectiveness of ICD therapy of C\$213,543 per life-year gained. The cost per life-year gained inpatients with ≥ 2 factors was C\$65,195, compared with C\$916,659 with < 2 risk factors.

We begin by introducing the “cost-effectiveness (CE) plane” as a device for presenting and relating the two central parameters of interest in economic evaluation: the difference (treatment minus control) in effectiveness (ΔE) and the difference in cost (ΔC). We show how the CE plane is useful for presenting uncertainty in the location of these two parameters and also uncertainty in the ratio between them, $\Delta C/\Delta E$, known as the incremental cost-effectiveness ratio (ICER). Using the CE plane, we review methods for estimating and presenting the uncertainty that can arise in cost-effectiveness results. Although much research has focused on methods for calculating confidence intervals for cost-effectiveness ratios, this calculation can be problematic with ratio-based statistics. Instead, we advocate plotting the joint density of cost and effect differences on the cost-effectiveness plane, together with cumulative density plots over the cost-effectiveness surface known as cost-effectiveness acceptability curves to summarize the overall value for money of interventions. We also outline the net-benefit formulation of the cost-effectiveness problem and show that it has particular advantages over the standard incremental cost-effectiveness ratio formulation. In the final section of the paper we consider some areas of continuing development of statistical methods for cost-effectiveness analysis such as the use of a Bayesian interpretation of probability, arguing that this is most natural given the decision-making context of cost-effectiveness analysis.

THE COST-EFFECTIVENESS PLANE

In Figure 1 we illustrate the cost-effectiveness (CE) plane, due originally to Black (4). The CE plane is a two-dimensional space with the x-axis being the average difference (treatment minus control) in effectiveness (ΔE) per patient and the y-axis being the average difference in cost (ΔC) per patient. Although costs are in money units such as dollars, the effectiveness units are typically health outcomes such as life-years gained or quality-adjusted life-years (18). In principle, the axes are unbounded from positive to negative infinity, and the origin represents the control group because scales are in difference form. To minimize confusion, we label the four quadrants using the points of the compass.

If we consider the ideal circumstance of knowing our (x, y) coordinates on the CE plane for sure, with no uncertainty, then a number of eventualities can arise. For example, treatment is said to “dominate” control, being less costly and more effective if the (x, y) coordinates are located in the SE quadrant and the

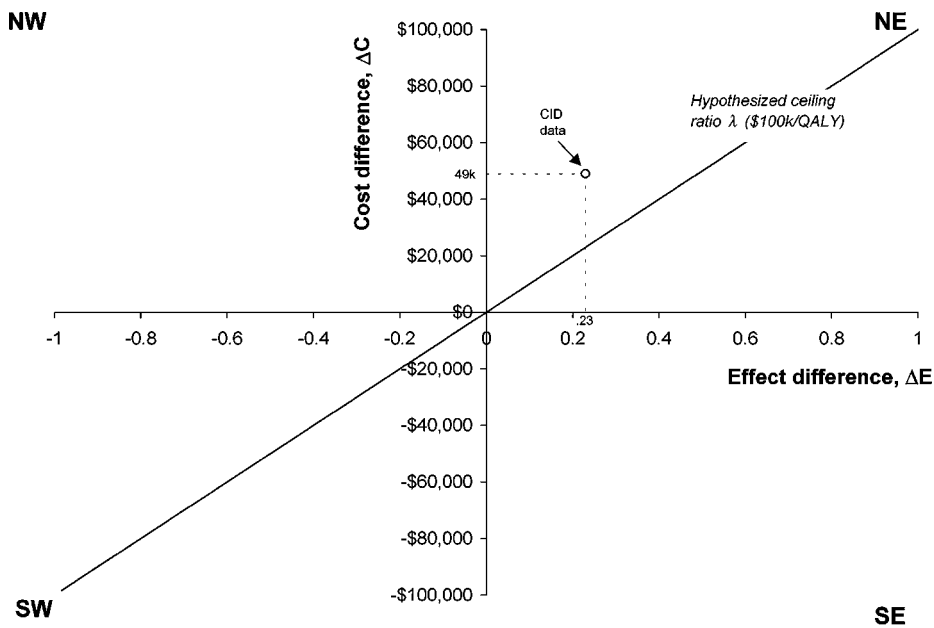


Figure 1 Incremental cost-effectiveness plane showing four quadrants, line representing the ceiling ratio for decision making and the location of the point estimate of incremental costs and effects for the CIDS data example.

mirror image—control dominates treatment—in the NW quadrant. In these two circumstances, the efficiency-based decision to adopt the new therapy or not is self-evident. But most new therapies locate in the NE quadrant where increased effectiveness is achieved at increased cost. In this situation, the decision to adopt the new therapy will depend on where the (x, y) coordinates fall in the NE quadrant and whether this point lies below the acceptable “ceiling ratio” of the decision-maker. As illustrated by the ray extending from the origin, the assumption is that the dollar amount that the decision-maker is willing to pay for a unit of effectiveness is known (call this λ). If the incremental cost-effectiveness ratio (ICER) of the new therapy ($\Delta C/\Delta E$), i.e., the slope of a straight line from the origin that passes through the ($\Delta E, \Delta C$) coordinate, is less than the decision-maker’s willingness to pay (λ), then the treatment should be adopted.

Using our example, if we assume, for the moment, that our ICD data (Box 1) had no uncertainty, then the true cost difference per patient would be C\$49,100 and the true increase in survival would be 0.23 years for an ICER of C\$214,000 per life-year gained. If we assume that the maximum that society is willing to pay for a year of life is C\$100,000, then ICD therapy should not be adopted. This is shown graphically in Figure 1 by the point estimate of cost-effectiveness falling above and to the left of the line with slope $\lambda = \text{C}\$100,000$. Of course, the problem

is that all the parameters are uncertain, including the amount society is willing to pay for a unit of effect.

ESTIMATION OR HYPOTHESIS TESTING?

In practice, we have only estimates of the cost and effect differences, and it is important that uncertainty in those estimates is also presented. It is straightforward to calculate confidence intervals for each of the cost and effect differences, ΔC and ΔE , using standard methods, and these intervals can also be plotted on the cost-effectiveness plane (38). For example, in the CIDS trial, the 95% confidence intervals for ΔC are (C\$41,600 to C\$56,600) and for ΔE are (-0.12 to 0.57). These results are represented on the cost-effectiveness plane in Figure 2, which, in addition to a point estimate of the cost and effect difference of ICD therapy, also shows I-bars representing the confidence intervals around those estimates. The horizontal I-bar represents the confidence interval for the effect difference, and the vertical I-bar represents the confidence interval for the cost difference.

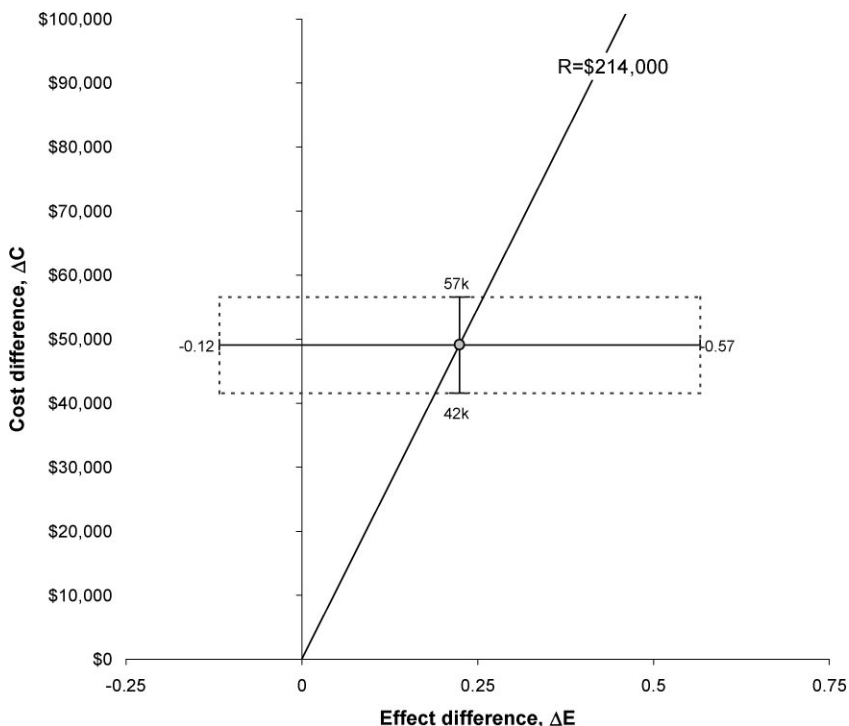


Figure 2 Confidence limits and the confidence box on the cost-effectiveness plane for the ICD data example.

Both have the point estimate of the cost and effect differences at their center and together the intervals define a box on the cost-effectiveness plane. Of note with our ICD example is that the box “straddles” the y-axis but lies completely above the x-axis, reflecting the fact that the difference in survival in the CIDS trial was not significant ($p > 0.05$) but that the difference in cost was significant ($p < 0.05$).

Our example in Figure 2 is just one situation that can arise when analyzing the results of an economic analysis conducted alongside a clinical trial with respect to the significance or otherwise of the cost and effect differences. In fact, there are nine possible situations that could arise, and these are illustrated on the cost-effectiveness plane in Figure 3 with multiple “confidence boxes.”

In situations 1 and 2, one intervention has been shown to be significantly more effective and significantly cheaper than the other and is therefore clearly the treatment of choice—the new treatment is preferred in the SE quadrant (situation 1) and the control treatment in the NW quadrant (situation 2). In situations 7 and 8, we have one treatment shown to be significantly more costly, but also significantly more effective. It is in these situations that it is clearly appropriate to estimate an ICER and where much research effort has been employed to ascertain the most appropriate method for estimating the ICER confidence interval.

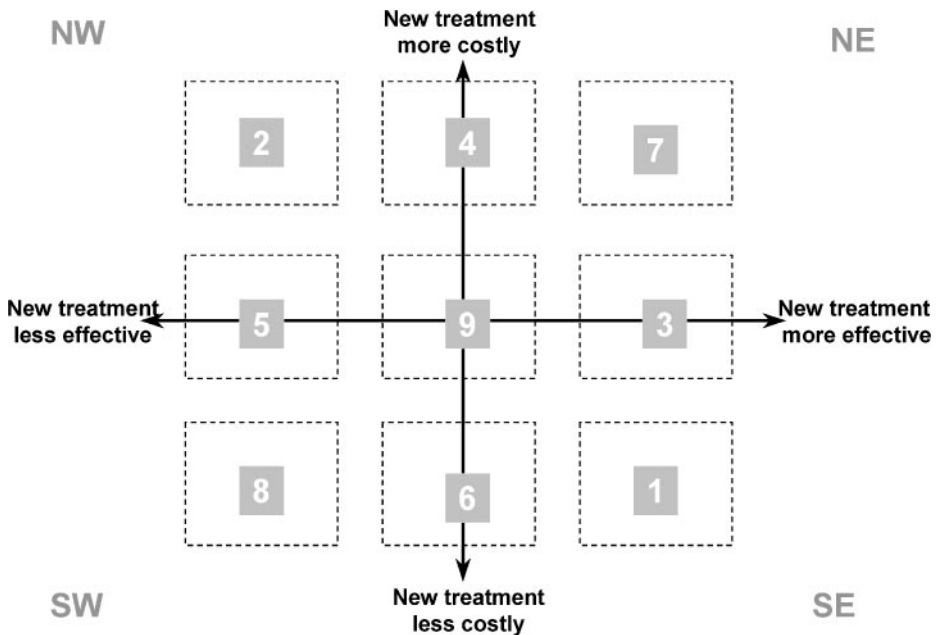


Figure 3 Nine possible situations that can arise concerning the significance (or otherwise) of cost and effect differences illustrated on the cost-effectiveness plane. Boxes indicate the area bounded by the individual confidence limits on cost and effect: statistically significant differences are indicated where the box does not straddle the relevant axis.

A potential problem arises in the situations where either the cost difference (situations 3 and 5) or the effect difference (situations 4 and 6) is not statistically significant. (Note that our ICD example falls into situation 4.) It is common to find analysts in these situations adapting the decision rule to focus only on the dimension where a difference has been shown. For example, it might be tempting in situation 4, our ICD example, to assume that ICD and amiodarone have the same life expectancy and only compare them in terms of cost. This form of analysis, known as cost-minimization analysis, uses the logic that among outcome-equivalent options one should choose the less costly option.

As we have argued elsewhere (10), the problem with this simple approach to decision making in situations where either cost or effect is not statistically significant is that it is based on simple and sequential tests of hypotheses. But the deficiencies of hypothesis testing (in contrast to estimation) are well known and gave rise to the memorable adage, “absence of evidence is not evidence of absence” (2). The concern is that a focus on hypothesis testing leads to an overemphasis on type I errors (the rejection of the null hypothesis of no difference when there is in fact no difference) at the expense of type II errors (the failure to reject the null hypothesis of no difference when in fact a difference does exist). In a review of clinical evaluations, Freiman and colleagues (24) showed how a substantial proportion of studies reporting “negative” results had insufficient power to detect quite important differences in treatment effect. Consistent with these recent debates in the clinical evaluation literature, the goal of economic evaluation should be the estimation of a parameter—incremental cost-effectiveness—with appropriate representation of uncertainty, rather than hypothesis testing.

ESTIMATING UNCERTAINTY: THINKING OUTSIDE THE BOX

The point estimates (means) from the effect and cost distributions provide the best estimate of the treatment and cost effects and should be used in the primary analysis. While confidence intervals for cost-effectiveness ratios are a valid approach to addressing uncertainty in cost-effectiveness analysis for situations 7 and 8, problems arise when uncertainty is such that the ICER could be negative (48). However, these problems can be overcome through either the appropriate representation of uncertainty on the cost-effectiveness plane (6, 51), or the use of the net-benefit statistic that represents a new framework for handling uncertainty in CEA and which does not suffer from the problems associated with the ICER in situations where negative ratios arise (49). In this section we review each of these issues in turn to emphasize how analysts should be estimating and presenting uncertainty in the results of their analyses in the potential situations outlined above.

Confidence Limits for Cost-Effectiveness Ratios

With patient-level information on the costs and effects of treatment interventions, it is natural to consider representing uncertainty in the ICER using confidence

intervals. However, as a ratio statistic, the solution to confidence-interval estimation is not straightforward.

The intuition behind this problem is that where there is nonnegligible probability that the denominator of the ratio could take a zero value, the ICER becomes unstable since for a zero denominator the ICER would be infinite. For a positive cost difference (the numerator of the ICER) as the effect difference approaches zero from the positive direction, the ICER tends to positive infinity. As the effect difference approaches zero from the negative direction, the ICER tends to negative infinity. For negative cost differences the ICER signs are reversed. This discontinuity about the zero effect difference causes statistical problems for estimating confidence limits; for example, there is no mathematically tractable formula for the variance of the statistic. Even where the effect difference is significantly different from zero, it would be inappropriate to assume that the ICER's sampling distribution followed a normal distribution.

There have been many proposed solutions to the problem of estimating confidence limits for the ICER, many of which were simply approximations that could perform rather poorly in some situations. However, a general consensus has emerged in support of two main approaches: the parametric method introduced by Fieller (23) half a century ago and the nonparametric approach of bootstrapping (19), both of which have been described in relation to cost-effectiveness analysis (9, 11, 14, 40, 42, 46, 53). We now illustrate each approach in turn, employing the example data from the CIDS trial (Box 1).

FIELLER'S THEOREM CONFIDENCE INTERVALS In Fieller's approach, it is assumed that the cost and effect differences (represented by ΔC and ΔE , respectively) follow a joint normal distribution. The standard cost-effectiveness ratio calculation of $R = \Delta C / \Delta E$ can be expressed as $R\Delta E - \Delta C = 0$, with known variance $R^2 \text{var}(\Delta E) + \text{var}(\Delta C) - 2R \text{cov}(\Delta E, \Delta C)$. Therefore, we can generate a standard normally distributed variable by dividing the reformulated expression through by its standard error:

$$\frac{R\Delta E - \Delta C}{\sqrt{R^2 \text{var}(\Delta E) + \text{var}(\Delta C) - 2R \text{cov}(\Delta E, \Delta C)}} \sim N(0, 1).$$

Setting this expression equal to the critical point from the standard normal distribution, $z_{\alpha/2}$ for a $(1 - \alpha)$ 100% confidence interval, yields the following quadratic equation in R :

$$R^2[\Delta E^2 - z_{\alpha/2}^2 \text{var}(\Delta E)] - 2R[\Delta E \cdot \Delta C - z_{\alpha/2}^2 \text{cov}(\Delta E, \Delta C)] + [\Delta C^2 - z_{\alpha/2}^2 \text{var}(\Delta C)] = 0.$$

The roots of this equation give the Fieller confidence limits for the ICER, R . These roots are reproduced in the appendix; while apparently complicated, recall that in order to calculate the roots, only five pieces of information are required: the estimated effect difference, the estimated cost difference, their respective variances

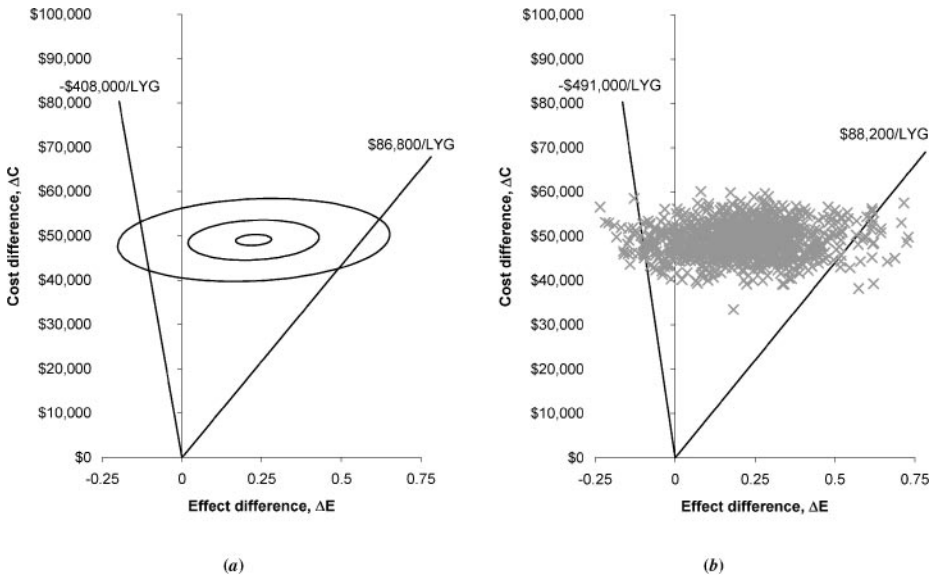


Figure 4 Fieller's theorem (a) and bootstrap (b) confidence limits on the CE plane for the ICD data example.

and the covariance between them. Figure 4a shows the assumption of joint normality on the cost-effectiveness plane for the ICD data of Box 1 by plotting ellipses of equal probability covering 5%, 50%, and 95% of the integrated joint density. Also plotted are the estimated confidence limits using Fieller's theorem (C\$86,800 to C\$-408,000), represented by the slopes of the lines on the plane passing through the origin. Note that the "wedge" defined by the Fieller confidence limits falls inside the 95% ellipse—taking tangents to the 95% ellipse, as was suggested in an early paper as a possible method for approximating the interval (51), would overestimate the width of the interval since the wedge area covers not only the 95% of the joint density covered by the ellipse but also areas above and below the 95% ellipse. By contrast, Fieller's approach automatically adjusts to ensure that 95% of the integrated joint density falls within the wedge, which makes Fieller's approach an exact method (subject to the parametric assumption of joint normality of costs and effects holding).

BOOTSTRAP CONFIDENCE INTERVALS The approach of nonparametric bootstrapping has been gaining in popularity with the advent of powerful desktop computing. It is a resampling procedure that employs raw computing power to estimate an empirical sampling distribution for the statistic of interest rather than relying on parametric assumptions. Bootstrap samples of the same size as the original data are drawn with replacement from the original sample and the statistic of interest is calculated. Repeating this process a large number of times generates a vector of

bootstrap replicates of the statistic of interest, which is the empirical estimate of that statistics' sampling distribution.

In terms of the cost-effectiveness application, the approach involves a three-step procedure:

1. Sample with replacement n_C cost/effect pairs from the patients in the control group (where n_C is the number of observed patients in the control group) and calculate the mean cost and effect in this bootstrap resample.
2. Sample with replacement n_T cost/effect pairs from the patients in the treatment group (where n_T is the number of observed patients in the treatment group) and calculate the mean cost and effect in this bootstrap resample.
3. Using the bootstrapped means from the steps above, calculate the difference in effect between the groups, the difference in cost between the two groups, and an estimate of the incremental cost-effectiveness.

This three-step procedure provides one bootstrap replication of the statistic of interest; repeating this process a large number of times (at least 1000 times is recommended for confidence interval calculation) generates the empirical distribution of cost-effectiveness.

Each of 1000 bootstrapped effect and cost differences from step 3 above are plotted on the cost-effectiveness plane in Figure 4b for the ICD data example. Confidence limits can be obtained by selecting the 26th and 975th of the 1000 replicates [which excludes 25 (or 2.5%) of observations from either end of the empirical distribution]; this effectively ensures that 95% of the estimated joint density falls within the wedge on the cost-effectiveness plane defined by the confidence limits. As is clearly apparent from Figure 4b, the bootstrap estimate of the joint density and the bootstrap confidence limits (C\$88,200 to C\$-491,000) are very similar to those generated by Fieller's theorem. This suggests that for this particular example, the assumption of joint normality for the cost and effect differences is reasonable. The Fieller limits are therefore preferred in this case for two main reasons: (a) Parametric methods are commonly more powerful than their nonparametric counterparts when the parametric assumptions hold; and (b) Fieller's approach always generates the same result; two analysts both employing the bootstrap method with the same data will generate slightly different results due to the play of chance.

Beyond the Confidence Interval: Acceptability Curves

Although commentators are now largely agreed on the most appropriate methods for ICER confidence interval estimation, such intervals are not appropriate in all the nine situations outlined above. One important problem concerns negative ratios. In the NW and SE quadrants, the ICER is negative and its magnitude conveys no useful meaning. The problem is that in the positive quadrants low ICERs are preferred to high ICERs (from the point of view of the more costly more effective treatment). However, no such simple arrangement exists in the negative

quadrants. Consider the three following points in the SE quadrant: A (1LY, -\$2000); B (2LYs, -\$2000); C (2LYs, -\$1000); giving negative ICERs of -\$2000/LY, -\$1000/LY and -\$500/LY, respectively. Therefore, in terms of magnitude, A has the lowest ICER, with C the highest and B between the two. However, it should be clear that B is preferred to both A and C as it has the highest number of life years saved and the greatest cost-saving. Furthermore, negative ICERs in the NW quadrant of the plane (favoring the existing treatment) are qualitatively different from negative ICERs in the SE quadrant (favoring the new treatment) yet will be grouped together in any naïve rank-ordering exercise (note the treatment of negative ratios in the bootstrapping of the ICD data above; since the negative ratios were in the NE quadrant they were ranked above the highest positive ratios to give a negative upper limit to the ratio).

A solution to this problem can be found by returning to the original decision rule introduced above. If the estimated ICER lies below some ceiling ratio, λ , reflecting the maximum that decision-makers are willing to invest to achieve a unit of effectiveness, then it should be implemented. Therefore, in terms of the bootstrap replications on the cost-effectiveness plane in Figure 4b, we could summarize uncertainty by considering what proportion of the bootstrap replications fall below and to the right of a line with slope equal to λ , lending support to the cost-effectiveness of the intervention. Of course, the appropriate value of λ is itself unknown. However, it can be varied to show how the evidence in favor of cost-effectiveness of the intervention varies with λ . In terms of the bootstrap method, we would simply plot the proportion of bootstrap replications falling on the cost-effective side of the line as λ is varied across its full range from 0 through to ∞ . Alternatively, if we are happy with an assumption of joint normality in the distribution of costs and effects, we can consider the proportion of the parametric joint density that falls on the cost-effective surface of the CE plane. We employ this parametric approach and the resulting curve for the ICD example based on the joint normal assumption shown in Figure 4a is presented in Figure 5 and has been termed a cost-effectiveness acceptability curve (51), as it directly summarizes the evidence in support of the intervention being cost-effective for all potential values of the decision rule.

This acceptability curve presents much more information on uncertainty than do confidence intervals. The curve cuts the vertical axis at the p-value (one-sided) for the cost difference (which is $p < 0.0001$ in our ICD example) since a value of zero for λ implies that only the cost is important in the cost-effectiveness calculation. The curve is tending toward 1 minus the p-value for the effect difference (which in the ICD example is $p = 0.10$), since an infinite value for λ implies that effect only is important in the cost-effectiveness calculation. The median value ($p = 0.5$) corresponds to the base-case ICER, which is C\$214,000 in our example.

As well as summarizing, for every value of λ , the evidence in favor of the intervention being cost-effective, acceptability curves can also be employed to obtain a confidence interval on cost-effectiveness. For the ICD example, the 95% upper bound is not defined and the 95% lower bound is equal to C\$86,800.

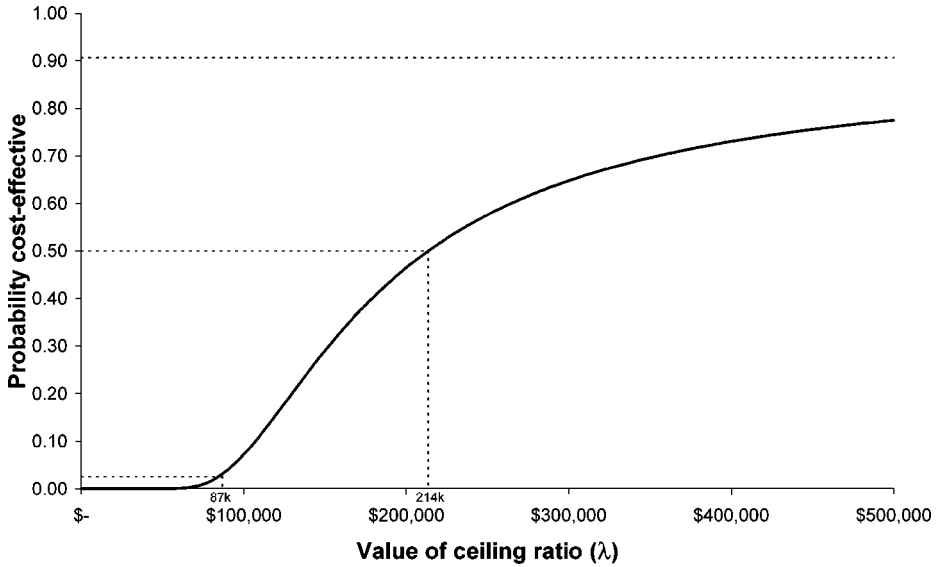


Figure 5 Parametric cost-effectiveness acceptability curve for the ICD data example (assuming joint normality of cost and effect differences).

Acceptability Curves and Stratified Cost-Effectiveness

In addition to the presentation of precision around parameter estimates such as cost-effectiveness, it is important to understand heterogeneity in data. For most (if not all) medical technologies there is variability in response to therapy, and this can often be systematic, identifying subgroups of patients where the treatment effect is larger or smaller. Although the standard cautions regarding the “trawling” for subgroups apply (17, 41), such information is important for presenting cost-effectiveness data to decision-makers. Selective use of therapies in patients where it is more effective and cost-effective requires the analyst to present the decision-maker with data showing both precision and heterogeneity.

The cost-effectiveness acceptability curve is a convenient method for presenting stratified analyses. Consider the ICD example again, based on the summary presented in Box 2 where clinical risk stratification by age (≥ 70 years), left ventricular ejection fraction ($\leq 35\%$), and New York Heart Association class (III) indicated patients who were likely to have a higher mortality benefit. In Figure 6a, we show how the presence of 0 through to 3 risk factors impacts on the point estimates of cost-effectiveness, with the cost-effectiveness of treatment being more favorable in persons with more risk factors (i.e., higher prior probability of death). In Figure 6b, the acceptability curves for the same groups are presented so the decision-maker can determine the probability of ICD therapy being cost-effective among subgroups and conditional upon the value of a life-year (λ).

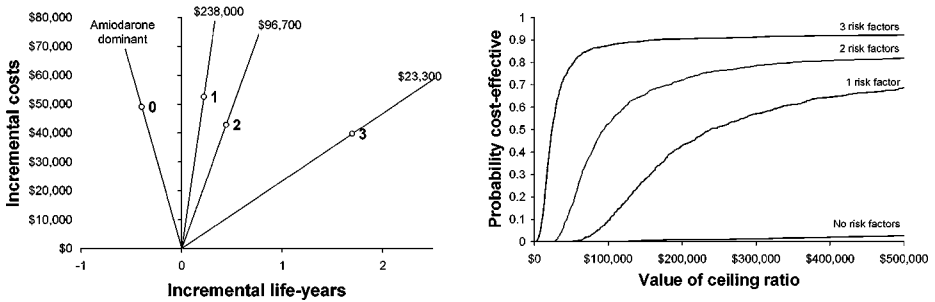


Figure 6 Risk stratified CEA for the ICD data example: (a) basecase results on the CE plane, (b) risk stratified acceptability curves.

The Net-Benefit Framework

Relatively recently, a number of researchers have employed a simple rearrangement of the cost-effectiveness decision rule to overcome the problems associated with ICERs (15, 16, 49, 50). In particular, Stinnett & Mullahy (49) offer a comprehensive account of the net-benefit framework and make a convincing case for employing the net-benefit statistic to handle uncertainty in stochastic cost-effectiveness analysis. The standard cost-effectiveness decision rule, to implement a new treatment only if $\Delta C/\Delta E < \lambda$, can be rearranged to give two alternative inequalities on either the cost scale (15, 16, 50) or on the effect scale (49). For simplicity, we focus on the cost scale of Net Monetary Benefit (NMB):

$$NMB = \lambda \cdot \Delta E - \Delta C.$$

The advantage of formulating the cost-effectiveness decision rule in this way is that, by using the value of λ to turn the decision rule into a linear expression, the variance for the net-benefit statistics is tractable and the sampling distribution is much better behaved (in that with sufficient sample size net-benefits are normally distributed). The variance expression for net-benefit on the cost scale is given by

$$\text{var}(NMB) = \lambda^2 \cdot \text{var}(\Delta E) + \text{var}(\Delta C) - 2\lambda \cdot \text{cov}(\Delta E, \Delta C).$$

Since the net-benefit statistic relies on the value of the ceiling ratio λ to avoid the problems of ratios statistics when in fact the value of the ceiling ratio is unknown, the net-benefit can be plotted as a function of λ . Figure 7 shows this for the net monetary benefit formulation of net-benefits and includes the 95% confidence intervals on net-benefits using the formula for the variance given above and assuming a normal distribution. The net-benefit curve crosses the horizontal axis at the point estimate of cost-effectiveness of the intervention, which is CS\$214,000 in our ICD example. Where the confidence limits on net-benefits cross the axis gives the confidence interval on cost-effectiveness. We see from the figure that while

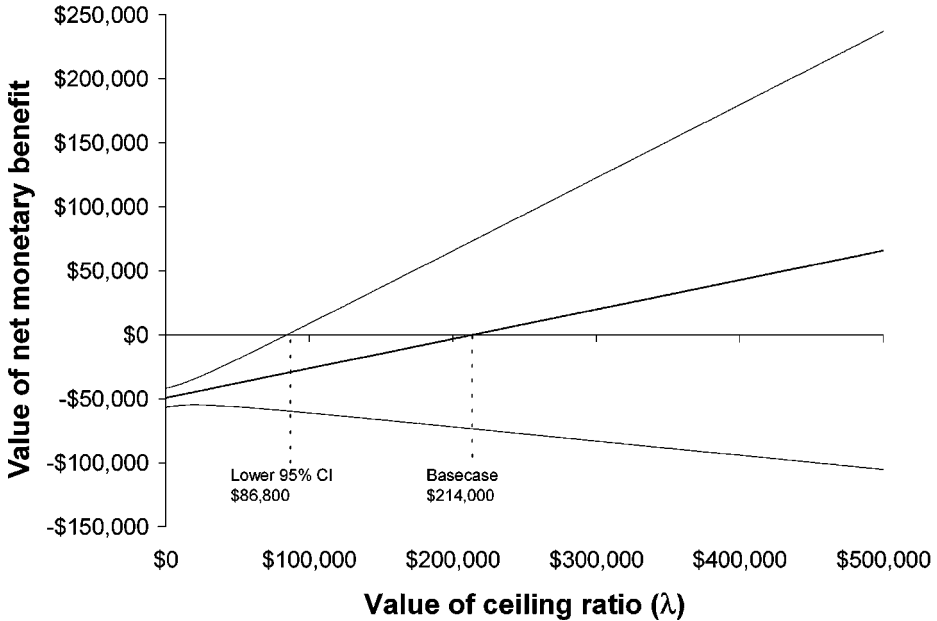


Figure 7 Net monetary benefit statistic as function of ceiling ratio for the ICD data example including 95% CI on net monetary benefit. Where the net benefit curves intersect with the $NMB = 0$ axis defines the point estimate and 95% confidence interval on cost-effectiveness. Note that the upper 95% limit on cost-effectiveness is not defined in this example.

the lower limit of cost-effectiveness is \$86,800, the upper 95% limit of net-benefit does not cross the axis, which indicates that the upper limit on cost-effectiveness is not defined. This is precisely the same result obtained from the analysis of the acceptability curve in Figure 5. Indeed, the net-benefit statistic provides a straightforward method to estimate the acceptability curve. Each point of the acceptability curve can be calculated from the p-value on the net-benefit being positive. Note that an acceptability curve calculated in this way gives the exact same acceptability curve as the analysis on the CE plane suggested by van Hout and colleagues (51), based on the joint normal distribution of cost and effect differences.

There is much common ground between the net-benefit method and Fieller's theorem. Indeed, the formal equivalence of the confidence limits described from the net-benefit method and from Fieller's theorem (and by extension the limits obtained from the acceptability curve above) have recently been demonstrated (26). Although Fieller's method fails to produce confidence limits for the ICER in some situations at a specified level of alpha, the type I error rate, this reflects a problem not of the method itself, but of the level of uncertainty. While such an interval can be defined for net benefit, that interval, by definition, will include zero at the specified level of confidence.

Since confidence intervals for cost-effectiveness ratios are not always defined, we strongly recommend that analysts plot their results on the cost-effectiveness plane, using either bootstrap replications or ellipses under the assumption of joint normality (see Figure 4*a,b*). This gives a visual representation of the joint uncertainty that is straightforward to interpret. Further summary can be obtained through the acceptability or net-benefit frameworks. Our own preference is the use of acceptability curves since these curves directly address the question of the study: How likely is it that the new intervention is cost-effective?

Power and Sample Size Calculations for Cost-Effectiveness

Up to this point we have been considering the analysis of cost and effect information generated alongside clinical trials, and we have recommended the reporting of estimated uncertainty in cost-effectiveness results rather than tests of hypothesis due to a concern of low power. These concerns are exacerbated by the fact that cost data are generally considered to have higher variance than effect data and that health economists are rarely invited to contribute to the power calculations at the design stage of a clinical trial. On the rare occasions that economists have been involved, it has tended to be the case that calculations are undertaken on costs and effects separately. However, if the purpose of economic evaluation is to make inference about cost-effectiveness then sample size and power calculations should be directly related to this cost-effectiveness result.

A number of authors have suggested the idea of basing power calculations on the methods used for approximating confidence intervals for cost-effectiveness ratios (7,45), including the use of simulation techniques (1). However, the introduction of the net-benefit statistic has simplified matters considerably. Sample size calculations can now be derived for cost-effectiveness following exactly the same procedure used for mean effectiveness.

Note that an observed net benefit is significantly positive providing

$$NMB - z_{\alpha/2} \sqrt{\text{var}(NMB)} > 0,$$

where $\text{var}(NMB)$ is as given above. Although it is tempting to base the sample size calculation on the numbers of patients required to show an observed difference as significant, in fact sample size calculations should be based on the hypothesized cost and effect differences (denoted $\Delta \tilde{E}$, $\Delta \tilde{C}$, generating a hypothesized net monetary benefit $N\tilde{M}B$) such that the study has the appropriate power to detect the hypothesized net-benefit as different from zero. In algebraic terms:

$$N\tilde{M}B - z_{\beta} \sqrt{\text{var}(N\tilde{M}B)} > z_{\alpha/2} \sqrt{\text{var}(N\tilde{M}B)},$$

where z_{β} is the critical value from the standard normal distribution corresponding to a required power of $1 - \beta$, and the variance expressions for net-benefit are as given above, but based on the hypothesized variance in cost, effect, and their covariance.

Substituting the sample variance calculations into the inequality above allows a straightforward (if rather extensive) rearrangement of the expression to give the sample size requirement (see the derivation given in the appendix). Note that as well as the hypothesized cost and effect differences, their associated variances, and covariance, the sample size also depends on the power and significance levels as well as the ceiling ratio λ . Assuming power and significance are fixed by convention, the sample size calculation can be presented as a function of the remaining unknown value λ ; however, at the design stage a single value must be chosen to give the final number of patients to be recruited.

As an example, consider the risk stratification analysis of the CIDS data example as a hypothesis-generating exercise that leads us to suppose that although implantable defibrillators do not seem good value overall, they may be cost-effective for patients with all three of the risk factors specified above. Further suppose that we now wish to design a cost-effectiveness trial to test this hypothesis and we are prepared to use the observed data from the CIDS study as the basis for the sample size calculations for the new study. Figure 8 shows the sample size requirements for such a study for different levels of power to detect a cost-effectiveness ratio significantly below the ceiling ratio at the 5% level as a function of the ceiling ratio. At conventional levels of power and significance (90% and 5%, respectively), we

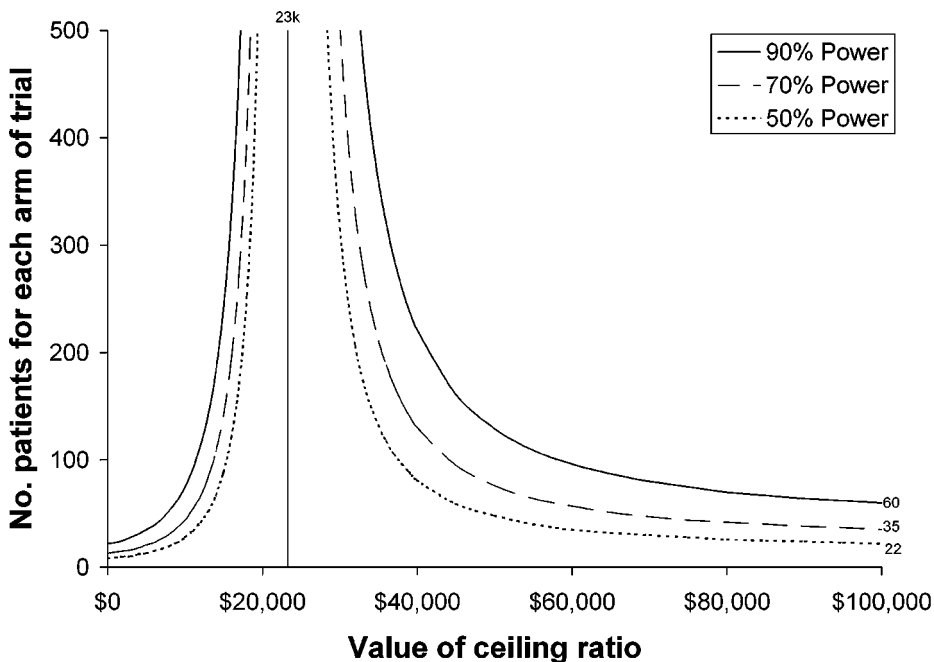


Figure 8 Sample size requirements for a hypothetical cost-effectiveness study to look at the cost-effectiveness of ICDs in patients with three risk factors.

would have to recruit 60 patients with all three risk factors to each arm of the trial, assuming a ceiling ratio of C\$100,000 per LYG.

Alternatively, the information in Figure 8 can be used to determine the power of a cost-effectiveness of known study size to show cost-effectiveness significantly below a given ceiling ratio. Note the discontinuity of the figure around the hypothesized point estimate of C\$23,300; this occurs when the ceiling ratio used by decision-makers corresponds to the true cost-effectiveness result. In this case then, no study, however large, will be able to show a significant difference. The implications of this are that where interventions are only marginally cost-effective, it is likely to prove very costly to run trials to demonstrate conclusive proof of cost-effectiveness.

FURTHER ISSUES AND FUTURE DIRECTIONS

Statistical methods for economic evaluations running alongside clinical trials is in a state of evolution, and we are likely to see many developments and refinements of the methods in the coming years. We begin by considering the use of Bayesian methods given the decision-making basis of economic evaluation research. We then go on to consider the nature and distribution of cost data and issues relating to their completeness that present particular statistical challenges.

On Being Bayesian with Probability

Although a strict frequentist interpretation of cost-effectiveness acceptability curves is possible through the consideration of the p-value on net benefits (32), the natural way to interpret these curves is as the probability that the intervention is cost-effective. Indeed, this is the way cost-effectiveness acceptability curves have been presented in the literature to date (44, 51). It has also been argued that the widespread mistaken interpretation of traditional p-values by researchers as a probability that the null hypothesis is false may be due to the fact that researchers want to make probability statements about the null hypothesis in this way (5). A number of commentators have stressed that such a view of probability in cost-effectiveness analysis is only possible in a Bayesian framework (27, 33, 39).

Fundamentally, the Bayesian approach includes a learning process whereby beliefs concerning the distributions of parameters (prior distributions) are updated (to posterior distributions), as information becomes available, through the use of Bayes' Theorem. Historically, advocates of the Bayesian approach were seen to inhabit a different scientific paradigm that was at odds with the frequentist paradigm: Frequentists considered Bayes methods as subjective and highly dependent on the prior beliefs employed, whereas frequentist methods were objective and robust. However, the adoption of such an extreme position would be to reject a set of very powerful methods that may be of import, even for frequentists (13). The empirical Bayes methods and Bayesian analysis based on uninformative prior distributions

are not subjective and have much to offer the frequentist analyst. Acceptability curves based on observed data, such as those presented in Figures 5 & 6, can be given the Bayesian interpretation assuming an uninformative prior distribution (8). Of course, if there is good prior information available on the cost-effectiveness of an intervention, then analysts may want to use this to formulate the prior in a Bayesian analysis.

At present, and most likely in the immediate future, health economists conducting economic analyses alongside clinical trials will have to work within the sample size constraints imposed by clinical investigators. This is likely to generate the situation where important economic differences cannot be detected at conventional levels of power and significance. A number of commentators have suggested that it may be appropriate for economic analysts to work with “error rates” (in the frequentist sense) that are higher than those employed in clinical evaluation (18, 37). This suggestion indicates the desire of economic analysts to consider the weight of evidence relating to the cost-effectiveness of the intervention under evaluation rather than relying on showing significance at conventional levels. This is most easily achieved through the use of cost-effectiveness acceptability curves, which show the weight of evidence for the intervention being cost-effective for all possible values of the ceiling ratio, λ . Furthermore, a Bayesian view of probability allows analysts to directly address the study question: How likely is it that the intervention is cost-effective?

Work is currently ongoing to reanalyze the cost-effectiveness analysis of the ICD data in a Bayesian framework.

Costing Challenges in Clinical Trials

In the discussion of the previous section on design and analysis issues in cost-effectiveness, we treated the cost data as if it were complete and followed a standard normal distribution. In practice, cost data present particular statistical challenges both in terms of the construct of the cost information and in the expected level of completeness.

The interest of decision-makers is in the mean total cost for a patient group. Patient costs are calculated by observing counts of resources used (e.g., visits to a general practitioner, prescribed medication, outpatient appointments, inpatient procedures, days spent in hospital), weighting these counts by a unit cost related to each resource item and summing across items. When considering this cost stochastically in a clinical trial, it is almost always the case that it is the resource use events that are truly stochastic, but that the unit costs applied are deterministic, with a single fixed value. Hence, total cost is really a weighted mixture of other distributions. Typically, this distribution of cost will be highly skewed with a few patients incurring rare but highly expensive costs (such as inpatient hospital procedures with all the associated costs) and many patients having few or no costs. Where cost data are highly skewed in this way, very large numbers of patients will be required before the assumption of normality (through the central limit theorem) can be applied.

Due to the mixture nature of the cost distribution and the inappropriateness of ordinal methods such as Wilcoxon and ordinal logistic regression for cost data (33a), much recent research has focused on the use of sophisticated statistical models to explain cost distributions. In particular, two-stage (or hurdle) models can be employed to distinguish between groups of patients incurring high and low (often zero) costs (31, 34, 35). However, to prove useful for decision-making these models need to be able to distinguish defining characteristics of patients that make them candidates for high- or low-cost pathways. All too often, it is impossible to predict a priori which patients will turn out to be high cost. Fortunately, however, and as Lumley and colleagues show (33a), standard t-tests and linear regression are robust to nonnormality of the data. Furthermore, that the skew coefficient of the population will be reduced by a factor of \sqrt{n} in the sampling distribution of the mean of that population (where n is the sample size) may guide the analyst as to whether that skew will have an important effect on the sampling distribution of the mean (7a).

Another problem relates to the completeness of the data, both in terms of administrative censoring and missing observations. Decision-makers are interested in the mean cost per patient for the lifetime of the patient. However, clinical trials rarely follow every patient to death. Instead, a cut-off point is specified at which time data collection stops and the analysis of the data begins. Where patients were recruited to the trial over a substantial recruitment period, there can be an administrative censoring problem such that the follow-up time for patients in the trial is different, with some having reached the endpoint of interest and some having been censored.

Of course, censoring is a problem for standard clinical results, not just costs, and the first attempts to handle censoring in cost data employed standard statistical approaches to survival analysis with cost as the survival metric rather than time (21, 22). Unfortunately, this approach is invalid since it can be shown that censoring (which occurs on the time scale) is no longer independent of cost: Patients accruing cost at a slow rate will more likely be censored in a naïve cost censoring analysis leading to a bias upwards in the censor-adjusted cost estimate (25). Instead, a technique has been advocated (known as the Kaplan-Meier sample average estimator) whereby costs are partitioned over time and uncensored costs are aggregated at each time interval and weighted by the probability of survival: Summing across these weighted estimates gives the censor-adjusted total cost estimate (20, 30).

However, this technique too has disadvantages. It is only unbiased in the limit as the partition size tends to zero; it cannot handle covariate adjustment and cannot be used to predict beyond the follow-up of the trial. New techniques are beginning to emerge that address these problems: The inverse probability weighting method is unbiased (3); an extension to the KMSA estimator has been developed that allows for covariate adjustment (29); a two-stage estimator has been developed that when implemented parametrically can be used to predict beyond the study period (12); and the survival analysis problem for costs has been extended to

cost-effectiveness through the use of the net benefit statistic (52). Further refinements of these methods are expected to provide a complete solution for analysts wanting to simultaneously handle censored cost and effect data, while adjusting for covariates and predicting beyond the follow-up of the trial.

SUMMARY AND CONCLUSIONS

In this paper, we have been concerned with the emerging quantitative techniques for analyzing the results of cost-effectiveness analyses undertaken alongside clinical trials. In particular, we have emphasized the use of the cost-effectiveness plane as a device to present and explore the implications of uncertainty. As a general rule, we would encourage analysts to make more use of the cost-effectiveness plane because we believe that it gives the clearest intuitive understanding of the implications of uncertainty for the analysis.

We stress the importance of estimation in cost-effectiveness studies rather than hypothesis testing: demonstrating that the application of separate and sequential tests of hypothesis could result in poor inference due to lack of power. Furthermore, any direct test of a cost-effectiveness hypothesis must involve the ceiling ratio for decision-making, λ , which is itself unknown. Therefore, formal tests of hypothesis are unlikely to be useful in economic evaluation studies; however, the use of confidence intervals for representing uncertainty in the ICER is limited. Rather, we advocate the use of acceptability curves that directly address the concern of the decision-maker: How likely is it that the intervention is cost-effective? This interpretation requires a Bayesian view of probability, but a Bayesian approach is the most natural approach for decision-making.

The net-benefit framework provides a very important contribution to the analysis of uncertainty for incremental cost-effectiveness by removing the reliance on ratio statistics, which are inherently problematic from a statistical point of view. In particular, net-benefit methods allow straightforward calculation of acceptability curves, a simple solution to the problem of power calculation, and have recently been employed to directly estimate cost-effectiveness within a regression framework (28). The use of regression for cost-effectiveness is important because it provides both a framework to handle censoring of the data and a mechanism for exploring subgroup analysis. Both these issues are likely to receive increasing attention, and we look forward to continued refinement of the methods in this area.

ACKNOWLEDGMENTS

AB is the recipient of a Public Health Career Scientist Award from the U.K. Department of Health. The Canadian Implantable Defibrillator Study was funded by the Medical Research Council of Canada. However, views expressed in this paper are those of the authors and should not be attributed to any funding bodies.

TECHNICAL APPENDIX

Fieller's theorem for ICER confidence limits

We start from the quadratic equation in R, the limits of which are the Fieller confidence limits:

$$R^2[\Delta E^2 - z_{\alpha/2}^2 \text{var}(\Delta E)] - 2R[\Delta E \cdot \Delta C - z_{\alpha/2}^2 \text{cov}(\Delta E, \Delta C)] + [\Delta C^2 - z_{\alpha/2}^2 \text{var}(\Delta C)] = 0.$$

This equation is solved using the standard quadratic formula

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where:

$$a = \Delta E^2 - z_{\alpha/2}^2 \text{var}(\Delta E)$$

$$b = -2[\Delta E \cdot \Delta C - z_{\alpha/2}^2 \text{cov}(\Delta E, \Delta C)]$$

$$c = \Delta C^2 - z_{\alpha/2}^2 \text{var}(\Delta C).$$

Substituting these values into the expression above simplifies only slightly with the 2s cancelling to give

$$\frac{[\Delta E \cdot \Delta C - z_{\alpha/2}^2 \text{cov}(\Delta E, \Delta C)] \pm \sqrt{[\Delta E \cdot \Delta C - z_{\alpha/2}^2 \text{cov}(\Delta E, \Delta C)]^2 - [\Delta E^2 - z_{\alpha/2}^2 \text{var}(\Delta E)] \cdot [\Delta C^2 - z_{\alpha/2}^2 \text{var}(\Delta C)]}}{\Delta E^2 - z_{\alpha/2}^2 \text{var}(\Delta E)}$$

In order to estimate these limits, only five simple sample statistics require estimation. For a comparison of control and treatment interventions indicated by the subscripts C and T respectively we have:

$$1. \Delta E = \bar{E}_T - \bar{E}_C = \frac{1}{n_T} \sum_{i=1}^{n_T} E_{Ti} - \frac{1}{n_C} \sum_{j=1}^{n_C} E_{Cj}$$

$$2. \Delta C = \bar{C}_T - \bar{C}_C = \frac{1}{n_T} \sum_{i=1}^{n_T} C_{Ti} - \frac{1}{n_C} \sum_{j=1}^{n_C} C_{Cj}$$

$$3. \text{var}(\Delta E) = \text{var}(\bar{E}_T) + \text{var}(\bar{E}_C) = \frac{s_{ET}^2}{n_T} + \frac{s_{EC}^2}{n_C}$$

$$4. \text{var}(\Delta C) = \text{var}(\bar{C}_T) + \text{var}(\bar{C}_C) = \frac{s_{CT}^2}{n_T} + \frac{s_{CC}^2}{n_C}$$

$$5. \text{cov}(\Delta E, \Delta C) = \text{cov}(\bar{E}_T, \bar{C}_T) + \text{cov}(\bar{E}_C, \bar{C}_C) \\ = \rho_T \sqrt{\text{var}(\bar{E}_T) \text{var}(\bar{C}_T)} + \rho_C \sqrt{\text{var}(\bar{E}_C) \text{var}(\bar{C}_C)}$$

where E and C represent effect and cost respectively, s_{ij}^2 is the sample variance for $i = \text{cost or effect}$ in the $j = \text{control or treatment groups}$ and ρ_j is the correlation coefficient between costs and effects in each group. Both the sample variance and correlation can be estimated using standard methods and are output by all standard statistical packages.

Sample size calculations

We start from the desire to have the power to show a hypothesized NMB as significant:

$$N\tilde{M}B - z_{\beta}\sqrt{\text{var}(N\tilde{M}B)} > z_{\alpha/2}\sqrt{\text{var}(N\tilde{M}B)}$$

and rearrange to get

$$N\tilde{M}B > (z_{\alpha/2} + z_{\beta})\sqrt{\text{var}(N\tilde{M}B)}.$$

Substituting into the above expression the standard expressions for NMB and its variance gives:

$$\lambda \cdot \Delta\tilde{E} - \Delta\tilde{C} > (z_{\alpha/2} + z_{\beta})\sqrt{\lambda^2 \cdot \text{var}(\Delta\tilde{E}) + \text{var}(\Delta\tilde{C}) - 2 \cdot \lambda \cdot \text{cov}(\Delta\tilde{E}, \Delta\tilde{C})}.$$

This gives an expression in the same five statistics as given above. Substituting in the expressions for these five sample statistics and rearranging on n (assuming equal sample sizes in each trial arm) gives

$$n > \frac{(z_{\alpha/2} + z_{\beta})^2 \cdot \{\lambda^2 \cdot [\tilde{s}_{ET}^2 + \tilde{s}_{EC}^2] + [\tilde{s}_{CT}^2 + \tilde{s}_{CC}^2] - 2 \cdot \lambda \cdot [\rho_T \tilde{s}_{ET} \tilde{s}_{CT} + \rho_C \tilde{s}_{EC} \tilde{s}_{CC}]\}}{[\lambda \cdot (\tilde{E}_T - \tilde{E}_C) - (\tilde{C}_T - \tilde{C}_C)]^2}$$

remembering that the variance expressions above relate to the hypothesized variances in the population, not the variances of the estimators.

Visit the Annual Reviews home page at www.annualreviews.org

LITERATURE CITED

1. Al MJ, van Hout BA, Michel BC, Rutten FF. 1998. Sample size calculation in economic evaluations. *Health Econ.* 7(4):327–35
2. Altman DG, Bland JM. 1995. Absence of evidence is not evidence of absence. *Br. Med. J.* 311(7003):485
3. Bang H, Tsiatis AA. 2000. Estimating medical costs with censored data. *Biometrika* 87(2):329–43
4. Black WC. 1990. The CE plane: a graphic representation of cost-effectiveness. *Med. Decis. Mak.* 10:212–14
5. Bland JM, Altman DG. 1998. Bayesians and frequentists. *Br. Med. J.* 317(7166):1151–60
6. Briggs A, Fenn P. 1998. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ.* 7(8):723–40
7. Briggs A, Gray A. 1998. Power and sample size calculations for stochastic cost-effectiveness analysis. *Med. Decis. Mak.* 18(8):S81–93
- 7a. Briggs A, Gray A. 1998. The distribution of health care costs and their statistical analysis for economic evaluation. *J. Health Serv. Res. & Policy* 3(4):233–45

8. Briggs AH. 1999. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ.* 8(3):257–61
9. Briggs AH, Mooney CZ, Wonderling DE. 1999. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and nonparametric techniques using Monte Carlo simulation. *Stat. Med.* 18(23):3245–62
10. Briggs AH, O'Brien BJ. 2001. The death of cost-minimisation analysis? *Health Econ.* 10:179–84
11. Briggs AH, Wonderling DE, Mooney CZ. 1997. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ.* 6(4):327–40
12. Carides GW, Heyse JF, Iglewicz B. 2000. A regression-based method for estimating mean treatment cost in the presence of right-censoring. *Biostatistics* 1(3):299–313
13. Carlin RP, Louis AT. 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall
14. Chaudhary MA, Stearns SC. 1996. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat. Med.* 15:1447–58
15. Claxton K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J. Health Econ.* 18(3):341–64
16. Claxton K, Posnett J. 1996. An economic approach to clinical trial design and research priority-setting. *Health Econ.* 5(6):513–24
17. Collins R, Gray R, Godwin J, Peto R. 1987. Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews. *Stat. Med.* 6(3):245–54
18. Drummond MF, O'Brien B, Stoddart GL, Torrance G. 1997. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford Univ. Press. 2nd ed.
19. Efron B, Tibshirani R. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall
20. Etzioni RD, Feuer EJ, Sullivan SD, Lin D, Hu C, Ramsey SD. 1999. On the use of survival analysis techniques to estimate medical care costs. *J. Health Econ.* 18(3):365–80
21. Fenn P, McGuire A, Backhouse M, Jones D. 1996. Modelling programme costs in economic evaluation. *J. Health Econ.* 15(1):115–25
22. Fenn P, McGuire A, Phillips V, Backhouse M, Jones D. 1995. The analysis of censored treatment cost data in economic evaluation. *Med. Care* 33(8):851–63
23. Fieller EC. 1954. Some problems in interval estimation. *J. R. Stat. Soc., Ser. B* 16:175–83
24. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. 1978. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. *N. Engl. J. Med.* 299(13):690–94
25. Hallstrom AP, Sullivan SD. 1998. On estimating costs for economic evaluation in failure time studies. *Med. Care* 36(3):433–36
26. Heitjan DF. 2000. Fieller's method and net health benefits. *Health Econ.* 9(4):327–35
27. Heitjan DF, Moskowitz AJ, Whang W. 1999. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Econ.* 8(3):191–201
28. Hoch JS, Briggs AH, Willan A. 2002. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ.* In press
29. Lin DY. 2000. Linear regression analysis of censored medical costs. *Biostatistics* 1(1):35–47

30. Lin DY, Feuer EJ, Etzioni R, Wax Y. 1997. Estimating medical costs from incomplete follow-up data. *Biometrics* 53(2):419–34
31. Lipscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar DB. 1998. Predicting the cost of illness: a comparison of alternative models applied to stroke. *Med. Decis. Mak.* 18(2 Suppl.):S39–56
32. Lothgren M, Zethraeus N. 2000. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Econ.* 9(7):623–30
33. Luce BR, Claxton K. 1999. Redefining the analytical approach to pharmacoeconomics. *Health Econ.* 8(3):187–89
- 33a. Lumley T, Diehr P, Emerson S, Chen L. 2002. The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* 23:151–69
34. Manning WG, Mullahy J. 2001. Estimating log models: to transform or not to transform? *J. Health Econ.* 20(4):461–94
35. Mullahy J. 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *J. Health Econ.* 17(3):247–81
36. O'Brien BJ, Connolly SJ, Goeree R, Blackhouse G, Willan A, et al. 2001. Cost-effectiveness of the implantable cardioverter-defibrillator: results from the Canadian Implantable Defibrillator Study (CIDS). *Circulation* 103(10):1416–21
37. O'Brien BJ, Drummond MF. 1994. Statistical versus quantitative significance in the socioeconomic evaluation of medicines. *PharmacoEconomics* 5(5):389–98
38. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. 1994. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med. Care* 32(2):150–63
39. O'Hagan A, Stevens JW, Montmartin J. 2000. Inference for the cost-effectiveness acceptability curve and cost-effectiveness ratio. *PharmacoEconomics* 17(4):339–49
40. Obenchain RL, Melfi CA, Croghan TW, Buesching DP. 1997. Bootstrap analyses of cost effectiveness in antidepressant pharmacotherapy. *PharmacoEconomics* 11:464–72
41. Oxman AD, Guyatt GH. 1992. A consumer's guide to subgroup analyses. *Ann. Intern. Med.* 116(1):78–84
42. Polsky D, Glick HA, Willke R, Schulman K. 1997. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ.* 6:243–52
43. Pritchard C. 1999. *Trends in Economic Evaluation*. London: Off. Health Econ. OHE Brief. Pap. No. 36
44. Raikou M, Gray A, Briggs A, Stevens R, Cull C, et al. 1998. Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetic patients (HDS7): UKPDS 40. *Br. Med. J.* 317:720–26
45. Ramsey SD, McIntosh M, Sullivan SD. 2001. Design issues for conducting cost-effectiveness analyses alongside clinical trials. *Annu. Rev. Public Health* 22:129–41
46. Severens JL, De Boo TM, Konst EM. 1999. Uncertainty of incremental cost-effectiveness ratios. A comparison of Fieller and bootstrap confidence intervals. *Int. J. Technol. Assess. Health Care* 15(3):608–14
47. Sheldon R, O'Brien BJ, Blackhouse G, Goeree R, Mitchell B, et al. 2001. Effect of clinical risk stratification on cost-effectiveness of the implantable cardioverter-defibrillator: the Canadian implantable defibrillator study. *Circulation* 104(14):1622–26
48. Stinnett AA, Mullahy J. 1997. The negative side of cost-effectiveness analysis. *JAMA* 277(24):1931–32
49. Stinnett AA, Mullahy J. 1998. Net

- health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med. Decis. Mak.* 18(2 Suppl.):S68–80
50. Tambour M, Zethraeus N, Johannesson M. 1998. A note on confidence intervals in cost-effectiveness analysis. *Int. J. Technol. Assess. Health Care* 14(3):467–71
51. van Hout BA, Al MJ, Gordon GS, Rutten FF. 1994. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ.* 3(5):309–19
52. Willan AR, Lin DY. 2001. Incremental net benefit in randomized clinical trials. *Stat. Med.* 20(11):1563–74
53. Willan AR, O’Brien BJ. 1996. Confidence intervals for cost-effectiveness ratios: an application of Fieller’s theorem. *Health Econ.* 5:297–305