



University
of Glasgow

O'Brien, B.J. and Briggs, A.H. (2002) Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Statistical Methods in Medical Research* 11(6):pp. 455-468.

<http://eprints.gla.ac.uk/4456/>

Deposited on: 24 June 2008

Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods

BJ O'Brien Department of Clinical Epidemiology and Biostatistics, McMaster University, Canada, and Centre for Evaluation of Medicines, St. Joseph's Hospital, Hamilton, Canada and **AH Briggs** Health Economics Research Centre, Department of Public Health, University of Oxford, Oxford, UK

Cost-effectiveness analysis is now an integral part of health technology assessment and addresses the question of whether a new treatment or other health care program offers good value for money. In this paper we introduce the basic framework for decision making with cost-effectiveness data and then review recent developments in statistical methods for analysis of uncertainty when cost-effectiveness estimates are based on observed data from a clinical trial. Although much research has focused on methods for calculating confidence intervals for cost-effectiveness ratios using bootstrapping or Fieller's method, these calculations can be problematic with a ratio-based statistic where numerator and/or denominator can be zero. We advocate plotting the joint density of cost and effect differences, together with cumulative density plots known as cost-effectiveness acceptability curves (CEACs) 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.

1 Introduction

Cost-effectiveness analysis is now an integral part of health technology assessment and addresses the question of whether a new treatment or other health care program offers good value for money. Economic evaluation has been most prominent and formalized in the context of public-payer reimbursement of new medicines. For example, the national Pharmaceutical Benefits Scheme in Australia and the Ontario Drug Benefit Plan in Canada both require economic evidence from pharmaceutical manufacturers in support of new submissions for formulary listing.^{1,2} In the UK, the National Institute of Clinical Excellence (NICE) uses economic evidence in setting guidance for the use of new technologies in the National Health Service.³ In the United States, the US Public Health Service has issued influential guidelines in how health care cost-effectiveness studies should be conducted.⁴

Interest in statistical issues surrounding cost-effectiveness analysis has grown rapidly in recent years, largely because more randomized trials of new therapies have begun collecting patient-level data on resource usage and costs. This development has led

Address for correspondence: Dr Bernie O'Brien, Centre for Evaluation of Medicines, 105 Main Street East, Level P1, Hamilton, Ontario, L8N 1G6, Canada. E-mail: obrienb@mcmaster.ca

Table 1 Cost-effectiveness of the implantable cardioverter defibrillator^a

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

^aSource: Abridged abstract from O'Brien *et al.*⁶

to methodological discussions on methods for estimating the parameters of cost-effectiveness and the role of statistical inference and hypothesis testing in relation to decision making with cost-effectiveness data.⁵

In this paper we introduce the basic framework for decision making with cost-effectiveness data and then review recent developments in statistical methods for analysis of uncertainty when cost-effectiveness estimates are based on observed data from a clinical trial. To illustrate and compare the methods, a common dataset is used throughout the paper based on our own work on the cost-effectiveness of the implantable cardioverter defibrillator (ICD).^{6,7} The example illustrates many of the challenging aspects of trial-based cost-effectiveness analysis and a brief summary can be found in Table 1.

As we will show, although much research has focused on methods for calculating confidence intervals for cost-effectiveness ratios using bootstrapping or Fieller's method, these calculations can be problematic with a ratio-based statistic where numerator and/or denominator can be zero. We advocate plotting the joint density of cost and effect differences, together with cumulative density plots known as cost-effectiveness acceptability curves (CEACs) 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.

2 The cost-effectiveness plane

In Figure 1 we illustrate the cost-effectiveness (CE) plane, due originally to Black.⁸ The CE plane is a two-dimensional space with the x-axis being the mean difference (treatment minus control) in effectiveness (ΔE) per patient and the y-axis being the mean difference in

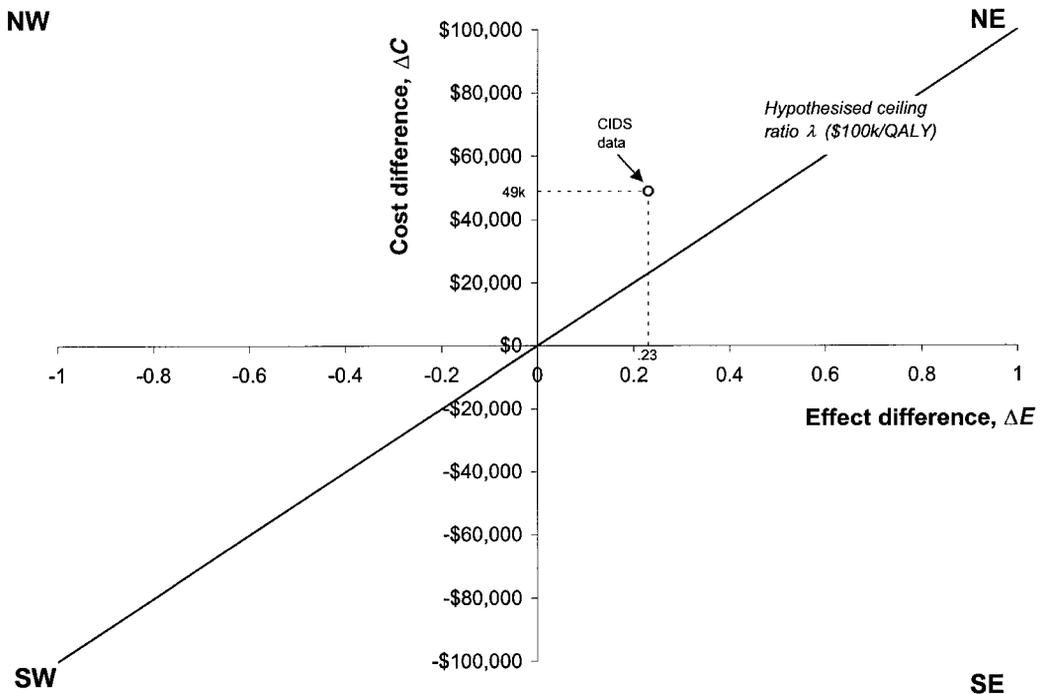


Figure 1 Incremental cost-effectiveness plane showing four quadrants, line representing the ceiling ratio for decision making and the location of the point estimate of incremental costs and effects for the CIDS data example

cost (ΔC) per patient. Costs are in money units such as dollars and effectiveness units are typically health outcomes such as survival ('life-years' gained) which may be adjusted for decrements in health-related quality of life to yield quality-adjusted life-years (QALYs).⁹ 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 aid reference, 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.

- **North-West quadrant:** treatment dominates control, being less costly and more effective, and the efficiency-based decision rule is to adopt treatment.
- **South-East quadrant:** treatment is dominated by control, being more costly and less effective, and the decision rule is not to adopt treatment.
- **North-East quadrant:** increased effectiveness with treatment is achieved at increased cost. In this situation, the decision to adopt the new therapy will depend on where the (x,y) coordinate falls in the NE quadrant and whether this point lies below the acceptable 'ceiling ratio' of the decision maker. As illustrated by the line 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 (denoted as λ). If the incremental cost-effectiveness ratio (ICER) of the new therapy ($\Delta C/\Delta E$), that is, 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 maximum willingness to pay (λ), then the treatment should be adopted.

- **South-West quadrant:** reduced effectiveness with treatment is associated with lower cost. This is the mirror image of the NE quadrant. If the ICER of the new therapy $(\Delta C/\Delta E)$ —which will be a positive ratio—is greater than λ , then the treatment should be adopted. Note that the decision rule is reversed since the ICER can equivalently be considered as the additional cost of control over treatment divided by the additional effect.

Using our example, if we assume, for the moment, that our ICD data (Table 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.

3 Quantifying the precision of ΔC and ΔE

Given that cost and effect differences are estimates it is important that uncertainty in those estimates is also presented. Based on the standard errors of the means 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 CE plane.⁵ In our example, 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 CE plane in Figure 2, which, in addition to a point estimate of the cost and effect difference of ICD therapy, also shows error bars representing the confidence intervals around those estimates.

The horizontal error bar in Figure 2 represents the confidence interval for the effect difference, and the vertical error bar represents the confidence interval for the cost difference. Both have the point estimate of the cost and effect differences at their center and together the intervals define a 'box' on the CE 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 with a two-tailed p -value of 0.10 but that the difference in cost was significant ($p < 0.05$). Clearly, one crude approach to quantifying the precision of the ICER (the ratio $\Delta C/\Delta E$), discussed by O'Brien *et al.*⁵ would be to define upper and lower bounds based on upper and lower 95% limits for ΔC and ΔE separately. For example, the lower bound for ICER would be the ratio of the upper 95% limit on ΔC and the upper 95% limit on ΔE .

There are two central problems with the so-called 'box' method. The first is that it ignores covariation between costs and effects. This is a symptom of the more general failure to model the joint density of ΔC and ΔE , which defines ellipses—contours of

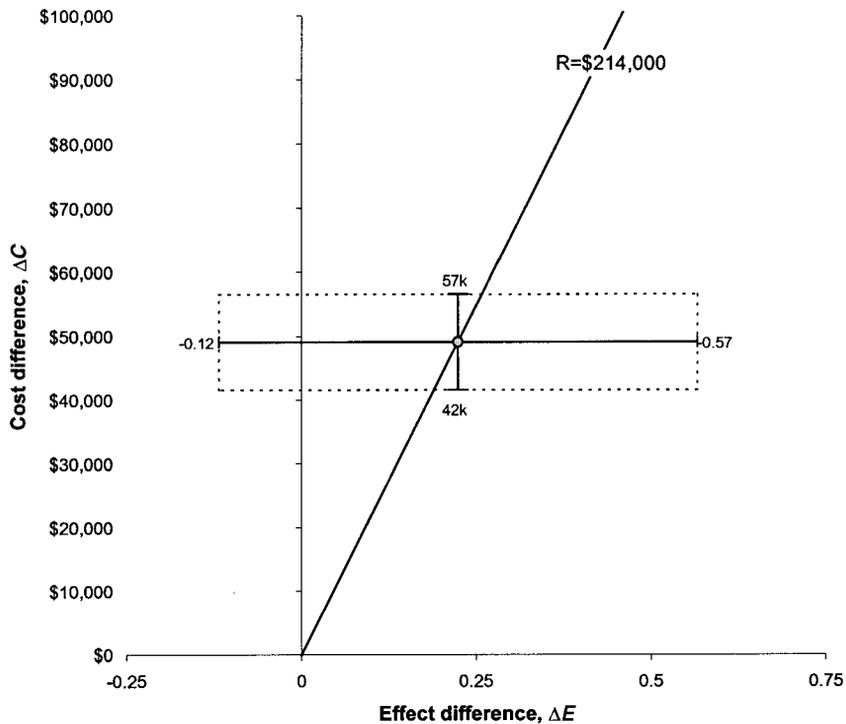


Figure 2 Confidence limits and the confidence box on the cost-effectiveness plane for the ICD data example

equiprobability—on the CE plane, and of differing shape depending on the sign and extent of covariation. The second problem is an issue of how such confidence intervals are interpreted. For the ICD data in Figure 2 it may be tempting 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 sequential inferential logic is that failure to show a significant difference is not the same as establishing equivalence. Focusing on hypothesis testing leads to an overemphasis on type I errors at the expense of type II errors. Our contention is that emphasis should be placed on the estimation of the joint density of ΔC and ΔE in the CE plane, using either parametric or non-parametric methods.

4 ICER confidence intervals based on the joint density of ΔC and ΔE

The ICER statistic is a ratio of two random variables, either of which can take the value zero, and this makes for an unstable distribution with discontinuities. For example, 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.

A general consensus has emerged in support of two main approaches: the parametric method introduced by Fieller¹¹ in 1954 and the non-parametric approach of bootstrapping,¹² both of which have been described in relation to cost-effectiveness analysis.¹³⁻¹⁶ We now illustrate each approach in turn, employing the example data from the CIDS trial (Table 1).

4.1 Fieller's theorem

Fieller's method is parametric and based on the assumption that the cost and effect differences (ΔC and ΔE) 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 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 R (the ICER) and are obtained from straightforward application of the standard formula for quadratic equations. After some messy manipulation we obtain

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

The attraction of this closed-form solution is that it requires only five pieces of information: the estimated effect difference, the estimated cost difference, their respective variances, and the covariance between them. Recently, Willan and Lin¹⁷ have also derived the covariance term for the Fieller method when costs and effects are subject to censoring.

Figure 3a shows the assumption of joint normality on the CE plane for the ICD data of Table 1 by plotting ellipses of equal probability covering 5, 50, and 95% of the

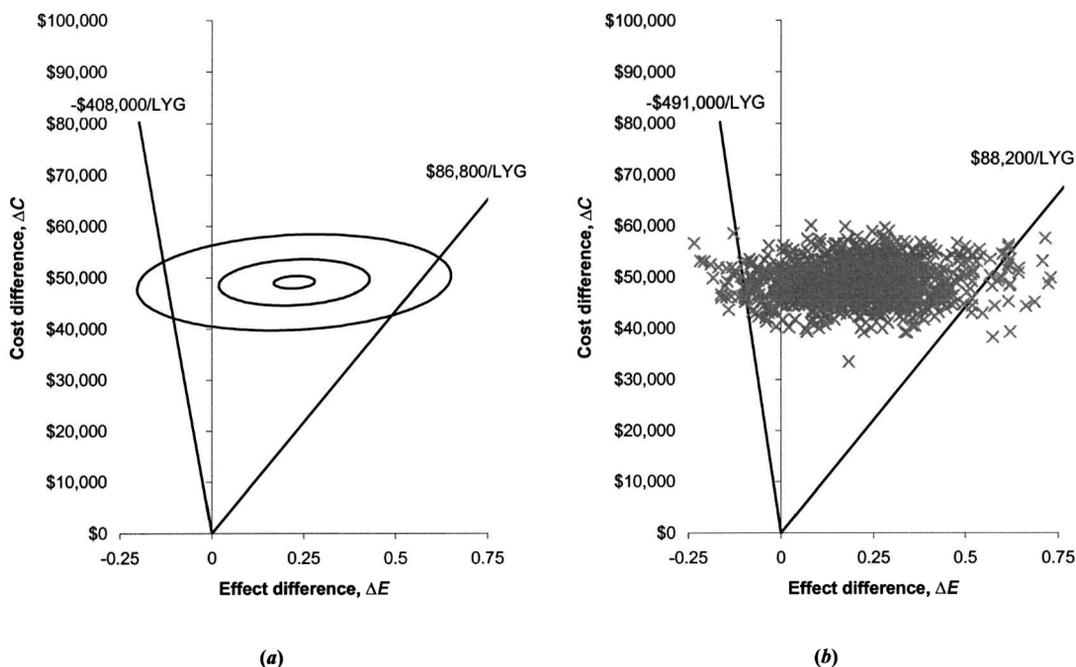


Figure 3 Fieller's theorem (a) and bootstrap (b) confidence limits on the CE plane for the ICD data example

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.

4.2 Bootstrapping

The approach of nonparametric bootstrapping has been gaining in popularity with the advent of powerful desktop computing.¹² It is a resampling procedure that estimates 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 statistic's 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 CE plane in Figure 3b 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 CE plane defined by the confidence limits. It should be immediately apparent from comparing Figures 3a and 3b that 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.

4.3 The problem of negative (and positive) ratios

The confidence interval methods outlined in the previous section are suitable for 'well-behaved' data, in the sense that the joint density of $(\Delta C, \Delta E)$ lies wholly within the North-East or South-West quadrants of the CE plane. But problems arise if parts of the joint density lie in the two 'dominance' quadrants (SE and NW) that generate negative ratios.¹⁸ First, a bootstrap interval for the ICER will order bootstrap replicates from low-to-high in a distribution. But two replicates with negative ratios can come from quadrants with totally opposite meaning—the NW quadrant (less effective, more costly) and the SE quadrant (more effective, less costly). In the CIDS example, this problem was countered by rank ordering all bootstrapped negative ratios in the North-West quadrant of the cost-effectiveness above the positive ratios before obtaining the percentile interval (see Figure 3b).

The more general problem is that data points in the negative quadrants have no meaningful ordering. 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.

Note, however, that the problem of ICER ordering also applies to positive ratios in different quadrants. As was pointed out above, the decision rule in the North-East quadrant is to implement the new treatment if the ICER is below the ceiling ratio. In the South-East quadrant the decision rule is to implement the new treatment if the ICER is greater than the ceiling ratio. Therefore, when uncertainty in the cost-effectiveness

results extend across different quadrants of the CE plane, problems arise for the interpretation of confidence intervals for the ICER.

5 Cost-effectiveness acceptability curves

Given the problems of negative cost-effectiveness ratios a new approach has been introduced based on the cost-effectiveness decision rule. Recall that 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 CE plane in Figure 3b, 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. This representation has been termed a cost-effectiveness acceptability curve (CEAC)¹⁹ as it directly summarizes the evidence in support of the intervention being cost-effective for all potential values of the decision rule. It should be noted that, although the exact value of λ is unknown, it must be resolved—explicitly or implicitly—at the time of any decision; assigning a money value to health outcomes is inevitable.

To compute a nonparametric CEAC, the bootstrap resamples from Figure 3b can be used in a plot of 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 infinity. For a parametric CEAC, the assumption of joint normality in the distribution of costs and effects is invoked, and we can integrate the function over the CE plane to determine the proportion of the parametric joint density— $f(\Delta C, \Delta E)$ —that falls on the cost-effective surface of the CE plane.²⁰ The CEAC is computed as the following double integral:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{\lambda \Delta E} f(\Delta C, \Delta E) d\Delta C d\Delta E$$

We employ this parametric approach and the resulting curve for the ICD example based on the joint normal assumption shown in Figure 3a is presented in Figure 4. 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. A particularly nice feature of

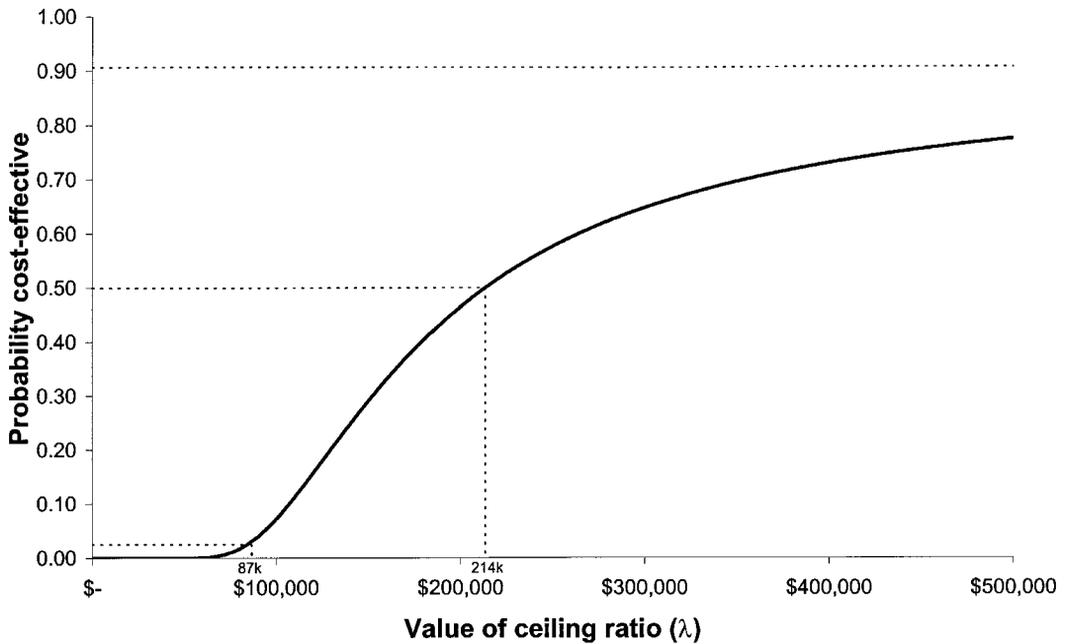


Figure 4 Cost-effectiveness acceptability curve for the ICD data example

the CEAC is that it can be used to show uncertainty and variability (e.g., variation in cost-effectiveness by age or risk factors) using multiple CEACs on the same plot. This approach to presenting stratified cost-effectiveness was used in our subsequent work on the cost-effectiveness of the ICD.²¹

6 The net-benefit framework

A recent development has been the rearrangement of the cost-effectiveness decision rule to overcome the problems associated with ICERs.^{22–25} In particular, Stinnett and Mullahy²⁴ 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 or on the effect scale. For simplicity, we focus on the cost scale of Net Monetary Benefit (NMB):

$$\text{NMB} = \lambda \cdot \Delta E - \Delta C$$

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 NMB is given by

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

Similar to the acceptability curve, the NMB is plotted conditionally as a function of λ . Figure 5 is the NMB plot for the CIDS example and includes the 95% confidence intervals on NMB based on the parametric variance formula above. The NMB curve crosses the horizontal axis at the point estimate of cost-effectiveness of the intervention, which is C\$214 000 in our ICD example. Where the confidence limits on NMB cross the axis gives the confidence interval on cost-effectiveness. We see from the figure that while the lower limit of cost-effectiveness is \$86 800, the upper 95% limit of NMB does not cross the axis, which indicates that the upper limit on cost-effectiveness is not defined. This is the same result obtained from the analysis of the acceptability curve in Figure 4.

NMB and parametric acceptability curves are closely related concepts. Each point of an acceptability curve can be calculated from the p -value on the null hypothesis of $\text{NMB}=0$. An acceptability curve calculated in this way gives the same acceptability

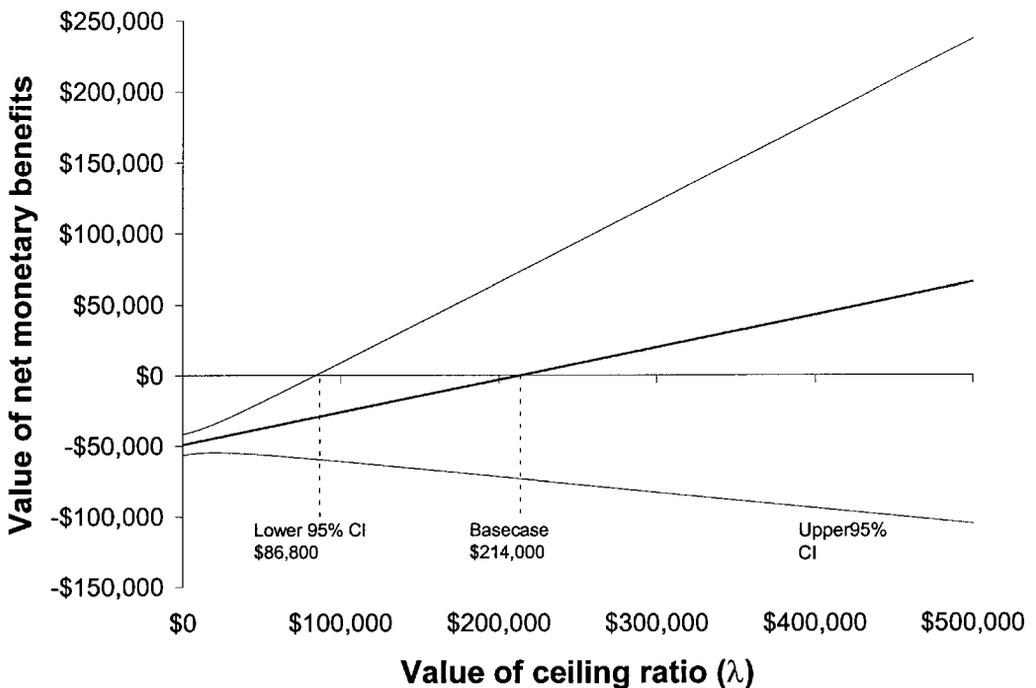


Figure 5 Net monetary benefit statistic as function of ceiling ratio for the ICD data example including 95% CI on net monetary benefit. Where the net-benefit curves intersect with the $\text{NMB} = 0$ axis defines the point estimate and 95% confidence interval on cost-effectiveness. Note that the upper 95% limit on cost-effectiveness is not defined in this example

curve as that by van Hout *et al.*,¹⁹ based on the joint normal distribution of cost and effect differences. Similarly, the net-benefit method and Fieller's theorem, being both based on the assumption of joint normality, have a formal equivalence.²⁶

6.1 Summary

Since confidence intervals for cost-effectiveness ratios are not always defined, we strongly recommend that analysts plot their results on the CE plane, using either bootstrap replications or ellipses under the assumption of joint normality (see Figure 3a, b). This gives a visual representation of the joint uncertainty that aids interpretation. Further summarization can be made with an acceptability curve or net-benefit plot. 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?

7 Statistical decision theory as a unifying framework

Although a strict frequentist interpretation of cost-effectiveness acceptability curves is possible through the consideration of the p -value on net benefits,²⁰ the natural way to interpret these curves is as the probability that the intervention is cost-effective. A number of commentators have stressed that such a view of probability in cost-effectiveness analysis is only possible in a Bayesian framework.²⁷⁻²⁹

Bayesian methods are central to the analytic framework of statistical decision theory³⁰ and differ from the frequentist in several fundamental ways. First, Bayesian methods formally incorporate previous evidence ('priors') with observed data ('likelihood') to estimate a posterior probability distribution for the hypothesis under study. Hence the Bayesian conditions on the data, not the null hypothesis being true, which is the frequentist perspective. Secondly, a full Bayesian decision analysis will parameterize a loss function associated with alternative decision consequences such that the expected utility of alternative courses of action can be computed. This second aspect is particularly attractive for economic analysis, which has a central focus on expected utility maximization. As shown by Claxton,²² in the case of cost-effectiveness analysis a composite metric such as NMB provides a logical means by which to parameterize a loss function.

But if one uses the framework of statistical decision theory a puzzle emerges. In Bayesian mode, Claxton²² argues that classical statistical inference is 'irrelevant' to decision making with cost-effectiveness data; the only defensible Type 1 error rate is 50% and the expected utility maximizer should always 'play the winner'. So what value is there to estimating uncertainty using the methods outlined above if the Bayesian perspective is to choose the decision option with the highest expected value? The answer lies in the value of collecting new data that will further reduce the expected utility loss associated with 'wrong decisions'; that is, circumstances where $NMB < 0$. Hence the relationship between the reduction in parameter uncertainty and reduced opportunity loss provides an efficiency-based framework for whether it is worthwhile to collect new data. Claxton and Posnett²³ have recently illustrated how the principles of expected information value that can be found in texts such as Pratt *et al.*,³⁰ can be applied to the economic appraisal of clinical trial design.

8 Conclusions

Data visualization is an important principle when examining uncertainty in cost-effectiveness data and we recommend the use of the CE plane as a graphical tool for arraying the joint density of cost and effect differences. We have also emphasized the estimation of uncertainty rather than hypothesis testing, arguing that the latter can lead to underpowered and misleading inference. Given that the ceiling ratio for the monetary value of health (λ) is not known, plus the distributional problems with the ratio-formed ICER, we believe that there is very limited scope for formal tests of hypotheses in cost-effectiveness studies. We advocate the use of CEACs that directly address the concern of the decision maker: how likely is it that the intervention is cost-effective if I am willing to pay $\$ \lambda$ for a unit of health outcome? The CEAC is a flexible tool and permits interpretation by the frequentist—who may look to 95% confidence limits for inference—and the Bayesian, who is interested in playing the winner and choosing the alternative that has the highest expected value.

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.³¹ NMB also provides a logical metric for a Bayesian loss function and therefore permits the analyst to assess the monetary value of further reducing parameter uncertainty by collecting new data.

Acknowledgements

This work was completed while Bernie O'Brien was a Visiting Professor at the University of Sydney and we thank the hosts at the Centre for Health Economics Research and Evaluation and the NHMRC Clinical Trials Centre. Andrew Briggs is the recipient of a Public Health Career Scientist Award from the UK Department of Health and Bernie O'Brien is supported by a Senior Investigator Award from the Canadian Institutes of Health Research. We thank Gordon Blackhouse for his research assistance. The Canadian Implantable Defibrillator Study was funded by the Medical Research Council of Canada.

References

- 1 Department of Health and Aging Commonwealth of Australia. Guideline for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: Including submissions involving economic analysis. 1990. Woden (ACT).
- 2 Ontario Ministry of Health. Ontario guidelines for economic analysis of pharmaceutical products. Toronto: Drug Programs Branch, 1994.
- 3 Sculpher MJ, Drummond M, O'Brien B. Effectiveness, efficiency, and NICE. A NICE start but evidence costs money. *British Medical Journal* 2001; **322**: 943–44.
- 4 Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine*. Oxford: Oxford University Press, 1996.
- 5 O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; **32**(2): 150–63.

- 6 O'Brien BJ, Connolly SJ, Goeree R *et al.* Cost-effectiveness of the implantable cardioverter defibrillator: results from the Canadian Implantable Defibrillator Study (CIDS). *Circulation* 2001; **103**: 1416–21.
- 7 Connolly SJ, Gent M, Roberts RS *et al.* Canadian Implantable Defibrillator Study (CIDS): A randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation* 2000; **101**: 1297–1302.
- 8 Black WC. The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making* 1990; **10**: 212–14.
- 9 Torrance GW, Feeny D. Utilities and quality-adjusted life years. *International Journal of Technical Assessment in Health Care* 1989; **5**: 559–75.
- 10 Briggs AH, O'Brien BJ. The death of cost-minimisation analysis? *Health Economics* 2001; **10**: 179–84.
- 11 Fieller EC. Some problems in interval estimation. *Journal of The Royal Statistical Society Series B* 1954; **16**: 175–83.
- 12 Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
- 13 Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* 1997; **6**: 327–40.
- 14 Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statistics in Medicine* 1996; **15**: 1447–58.
- 15 Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics* 1997; **6**: 243–52.
- 16 Willan A, O'Brien BJ. Confidence intervals for cost-effectiveness ratios in clinical trials: a new method using Fieller's Theorem. *Health Economics* 1996; **5**: 297–305.
- 17 Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Statistics in Medicine* 2001; **20**: 1563–74.
- 18 Stinnett AA, Mullahy J. The negative side of cost-effectiveness analysis. *JAMA* 1997; **277** (24): 1931–32.
- 19 van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics* 1994; **3**(5): 309–19.
- 20 Lothgren M, Zethraus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics* 2000; **9**(7): 623–30.
- 21 Sheldon R, O'Brien BJ, Blackhouse G *et al.* The effect of clinical risk stratification on the cost-effectiveness of the implantable cardioverter defibrillator. The Canadian Implantable Defibrillator Study. *Circulation* 2001; **104**(14): 1622–26.
- 22 Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 1999; **18**(3): 341–64.
- 23 Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Economics* 1996; **5**(6): 513–24.
- 24 Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998; **18**(Suppl): S65–S80.
- 25 Tambour M, Zethraus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technical Assessment in Health Care* 1998; **14**: 467–71.
- 26 Heitjan DF. Fieller's method and net health benefits. *Health Economics* 2000; **9**(4): 327–35.
- 27 Heitjan DF, Moskowitz AJ, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics* 1999; **8**(3): 191–201.
- 28 Luce BR, Claxton K. Redefining the analytical approach to pharmacoeconomics. *Health Economics* 1999; **8**(3): 187–89.
- 29 O'Hagan A, Stevens JW, Montmartin J. Inference for the cost-effectiveness acceptability curve and cost-effectiveness ratios. *Pharmacoeconomics* 2000; **17**(4): 339–49.
- 30 Pratt JW, Raiffa H, Schlaifer R. *Introduction to Statistical Decision Theory*. 2nd printing, 1996 ed. Cambridge, MA: MIT Press, 1995.
- 31 Hoch JS, Briggs AH, Willan A. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics* 2002; **11**: 415–430.