Incremental net benefit in the analysis of economic data from clinical trials, with application to the CADET-Hp Trial

Andrew R. Willan

In clinical trials it has become increasingly common for cost data to be collected in addition to data on clinical effectiveness. In response to this, new statistical methodology has been developed. Initially, efforts were focused on the incremental cost-effectiveness ratio, but as the problems with ratio statistics became apparent, attention was switched to incremental net benefit. In this paper statistical methods based on an incremental net benefit approach are given and illustrated with an example from the gastrointestinal literature. The relationship between incremental net benefit and incremental cost-effectiveness is emphasized. ***Eur J Gastroenterol Hepatol*** **16:543–549 © 2004 Lippincott Williams & Wilkins**

European Journal of Gastroenterology & Hepatology 2004, **16**:543–549

Keywords: randomized controlled trials, cost-effectiveness, incremental net benefit, incremental cost-effectiveness ratio

Program in Population Health Sciences, Hospital for Sick Children, Toronto, and Department of Public Health Sciences, University of Toronto, Canada.

This work was funded through the Discovery Grant Program of the Natural Science and Engineering Research Council of Canada (grant number 44868-03).

Correspondence to Dr A.R. Willan, P.H.S., 555 University Avenue, Toronto, ON M5G 1X8, Canada.
Tel: +1 416 813 2166; fax: +1 416 813 5979; e-mail: willana@mcmaster.ca

Introduction

It is becoming increasingly common in randomized controlled trials (RCTs) to collect patient-level cost data along with the measures of effectiveness. This added dimension to RCTs has motivated the development of new statistical methodology designed to answer questions of economic policy in addition to those of purely clinical importance. Initially, attention was focused on estimating the incremental cost-effectiveness ratio (ICER) and calculating the appropriate confidence intervals [1–10]. Most recently, due to the acknowledged problems associated with ratio statistics, attention has shifted to making inference about incremental net benefit (INB) using tests of hypothesis and confidence intervals [11–23]. Typically, the INB is defined in units of cost, but it can also be defined in units of effectiveness and referred to as the incremental net health benefit (INHB). Either definition requires the specification of the willingness to pay (WTP) for a unit of effectiveness (denoted as λ), or at the very least, the analysis must be presented as a function of λ so that readers can apply the WTP most appropriate for them. By the WTP we mean the value that policy makers are willing to pay for a unit of effectiveness. In some situations the unit of effectiveness might be a year of life. In the example given later in this paper, the unit of effectiveness is a success, defined as the presence of no or minimal dyspeptic symptoms at 1 year after the beginning of treatment.

The outline of this paper is as follows. In the next section (Methods) the terms used in a cost-effectiveness analysis are defined and the relationship between the ICER and INB is given. Also contained in this section are the methods for performing a statistical analysis of the data from a clinical trial in which both effectiveness and costs have been measured. An example is given in the third section (The CADET-Hp Trial), which is followed by the final section, further discussions.

Methods

The model

Consider a two-arm RCT in which patients are randomized between treatment (T) and standard therapy (S). Let the effectiveness be defined by some binary measure of success, i.e. each patient is either a success or a failure. A success could be defined as surviving a specified period of time or, as in the case of the example in the third section, as the presence of no or minimal dyspepsia symptoms at 1 year. Let Δ_e be the probability of success for a patient on T minus the probability of success for a patient on S. Assume that total cost has been collected for each patient, and let Δ_c be the mean cost of a patient on T minus the mean cost of a patient on S. Which resources are included in total costs depends on the perspective taken by the analysis. If the analysis takes the perspective of the health care system, only resources covered under the system would be included. However, if a broader societal perspective is adopted then aspects such as time lost from work and care by a family member could also be included.

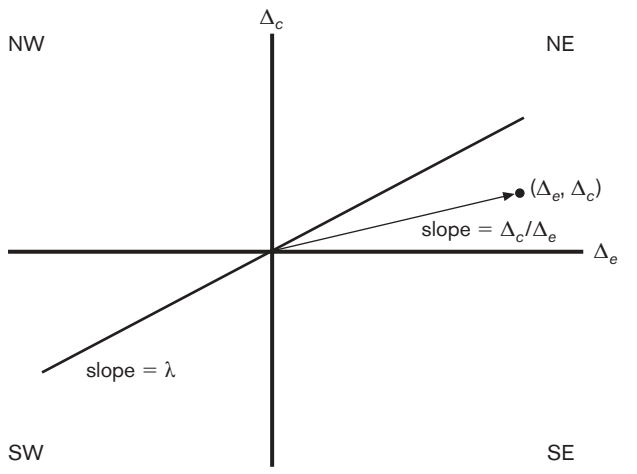
The expression $1/\Delta_e$ is known as the number needed to treat because it is the number of patients you would need to treat with T rather than S to expect one additional success. For example, if the probability of

0954-691X © 2004 Lippincott Williams & Wilkins

DOI: 10.1097/01.meg.0000108356.41221.ac

success was 0.8 with T and 0.6 with S, then $\Delta_e = 0.2$ and $1/\Delta_e = 5$. Therefore, for every five patients given T rather than S, one additional success would be expected. Therefore the expression $(1/\Delta_e) \times \Delta_c$, which is equal to Δ_c/Δ_e, is the total cost of treating the additional patients with T and, therefore, is the additional cost of achieving one extra success from using T rather than S. The expression Δ_c/Δ_e is known as the incremental cost-effectiveness ratio (ICER), and is the additional cost of achieving an extra unit of effectiveness. In our example a unit of effectiveness is a success, but in other RCTs it might be a year of life or quality adjusted year of life. Policy makers can use the ICER to judge the cost-effectiveness of adopting T rather than S and, subject to budget constraints and assuming Δ_e is positive, adoption of T would be recommended if the ICER was less than the WTP for a unit of effectiveness. The ICER is often displayed on the cost-effectiveness plane, which is a plot of Δ_c by Δ_e, as shown in Figure 1. The slope of the line connecting the origin to the point (Δ_e,Δ_c) is the ICER. For illustration, the line through the origin with slope equal to the WTP (λ) is shown. We shall refer to this line as the *threshold* in the remainder of this paper. For any point falling below the line, T is considered cost-effective. In the north-east (NE) quadrant, to be cost-effective a point must fall below the line so that the additional cost of T is off-set by the increase in effectiveness, i.e. ICER is less than the WTP. In the south-east (SE) quadrant (known as the win–win quadrant), since T costs less and is more effective, it is cost-effective regardless of the value of λ. In the south-west (SW) quadrant, for a point to be cost-effective it must fall below the line so that the reduction in effectiveness is off-set by the reduction in cost. In the north-west (NW) quadrant (known as the lose–lose quadrant), since T costs more and is less effective, it is not cost-effective regardless of the value of λ.

Fig. 1



The cost-effectiveness plane showing the incremental cost-effectiveness ratio (ICER).

An argument in favour of INB is that the ICER is not properly ordered in the SE and NW quadrants. Consider comparing S to three different treatments (T1, T2 and T3), for which the values of Δ_e and Δ_c are shown in Table 1. T1 increases the probability of success by 0.1 and reduces the cost by 2000, for an ICER of −20 000. T2 increases the probability of success by 0.2 and reduces the cost by 4000, and is clearly superior to T1, yet it has the same ICER. Therefore, two treatments compared to the same standard can have the same ICER, yet one may be clearly superior to the other. It gets worse. T3 increases the probability of success by 0.2 and reduces the cost by 2000, and is superior to T1 because, although it reduces cost by the same amount, its increase in the probability of success is double. However, the ICER for T3 is larger than for T1. It is shown below how INB orders these treatments properly. Other problems with the ICER involve constructing its confidence intervals. If the observed difference in effectiveness is small, the confidence interval may include undefined values, i.e. either the positive or negative vertical axis. Further, if the observed differences in effectiveness and cost are both small, the confidence interval may be undefined, meaning that no better inference regarding the ICER can be made after conducting the trial than before conducting it: clearly, a disappointing outcome.

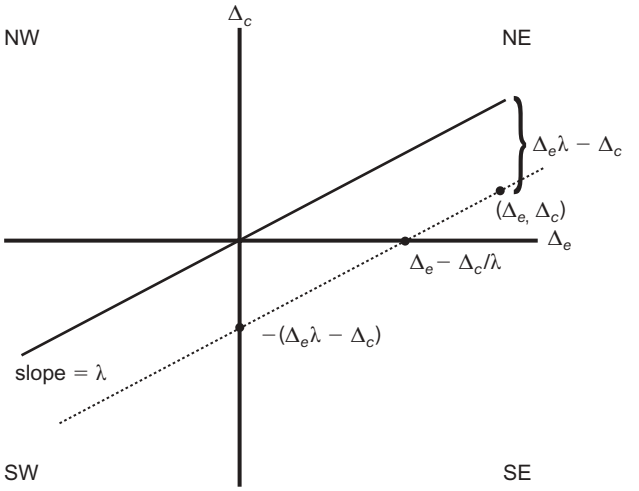
To address these issues researchers have turned to INB. INB in units of cost is defined as $b(\lambda) = \Delta_e\lambda - \Delta_c$. The first term is the increase in the number of units of effectiveness multiplied by what we are willing to pay for a unit of effectiveness, yielding the benefit of the increase in effectiveness expressed in monetary terms. The presence of the second term subtracts the increase in cost, leaving the incremental *net* benefit. T is cost-effective if, and only if, $b(\lambda)$ is positive, regardless of the sign of Δ_e. On the cost-effectiveness plane, $b(\lambda)$ is the vertical distance between the point (Δ_e,Δ_c) and the threshold, being positive if the point is below the line and negative if above (Fig. 2).

INB can be modified to handle more than one measure of effectiveness. Consider an RCT with two measures

Table 1 Three treatments (T1 to T3) in the win–win quadrant

Treatment	Δ_e	Δ_c	ICER
T1	0.1	−2000	−20 000
T2	0.2	−4000	−20 000
T3	0.2	−2000	−10 000

Fig. 2



The cost-effectiveness plane showing the incremental net benefit (INB).

of success: for example, avoiding an ulcer and avoiding an ulcer bleeding incident. Let $\Delta_{e,1}$ and $\Delta_{e,2}$ be the difference in the probabilities of avoiding an ulcer and an ulcer bleed, respectively. Then the INB can be defined as $b(\lambda_1,\lambda_2) = \Delta_{e,1}\lambda_1 + \Delta_{e,2}\lambda_2 - \Delta_c$, where λ_1 and λ_2 are the WTP to avoid an ulcer and an ulcer bleed, respectively. Also INB is linear in λ and thus a plot of $b(\lambda)$ by λ is a straight line in which the slope is Δ_e, the vertical intercept is $-\Delta_c$, and the horizontal intercept is the ICER (Fig. 3). This last point is seen by setting $b(\lambda)$ to 0 and solving for λ, yielding Δ_c/Δ_e. The characteristics of the plot of $b(\lambda)$ by λ depend on which quadrant of the cost-effectiveness plane the point (Δ_e,Δ_c) lies. This is illustrated in Figure 3. In the NE quadrant (panel b of Fig. 3) Δ_e, Δ_c, and the ICER are all positive. Therefore, the plot of $b(\lambda)$ has positive slope (Δ_e), positive horizontal intercept (Δ_c/Δ_e), and negative vertical intercept ($-\Delta_c$). In the SE quadrant (panel d) Δ_e is positive, and Δ_c and the ICER are negative. Therefore, the plot of λ has positive slope, negative horizontal intercept, and positive vertical intercept. In the SW quadrant (panel c) Δ_e and Δ_c are negative and the ICER is positive. Therefore, the plot of $b(\lambda)$ has negative slope, positive horizontal intercept, and positive vertical intercept. In the NW quadrant (panel a) Δ_e is negative, Δ_c is positive and the ICER is negative. Therefore, the plot of $b(\lambda)$ has negative slope, negative horizontal intercept, and negative vertical intercept.

A plot of $b(\lambda)$ by λ is particularly useful when a cost-effectiveness analysis is performed before the value of λ is identified, or if the researchers want to leave it to the readers to apply their own WTP. An advantage of INB is that, being expressed in monetary terms, it allows for comparisons of treatments applied to different patient populations where different measures of effectiveness are used. In fact, it allows for comparisons between any two public policy interventions, say, comparing highway safety improvements with breast cancer screening.

Returning to our comparison of T1, T2 and T3, it is easy to see how INB orders them properly. Recall that $b(\lambda)$ is the vertical distance on the cost-effectiveness plane between the point (Δ_e,Δ_c) and the threshold. Therefore, the vertical intercept of the line passing through the point (Δ_e,Δ_c) with slope λ is equal to $-b(\lambda)$ (Fig. 4). The more negative this intercept is, the larger the value of $b(\lambda)$. Essentially, INB projects each point on the cost-effectiveness plane onto the vertical axis via straight lines with slope λ, with negative values of $b(\lambda)$ above the horizontal axis and positive below. The projections for our example are shown in Figure 4. For any value of λ, $b_2(\lambda) > b_3(\lambda) > b_1(\lambda)$, which is the ordering consistent with any rational evaluation of the treatments.

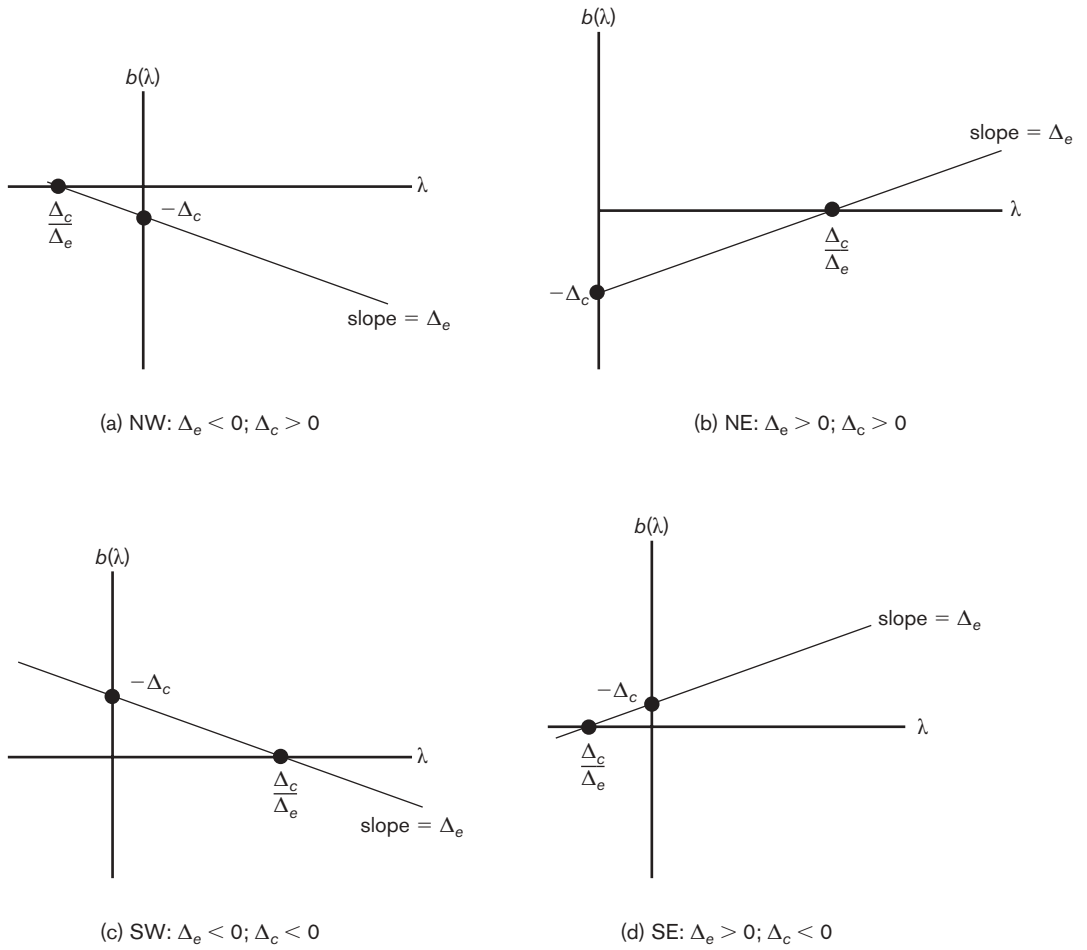
Statistical analysis

To conduct a statistical analysis of cost-effectiveness in this context the following five parameters must be estimated: Δ_e, Δ_c, $V(\hat{\Delta}e)$, $V(\hat{\Delta}c)$ and $C(\hat{\Delta}e, \hat{\Delta}c)$, where V and C are the variance and covariance functions, respectively, and $\hat{}$ means 'estimator of'. The parameter $b(\lambda)$ is then estimated by $\hat{b}(\lambda) = \lambda\hat{\Delta}e - \hat{\Delta}c$. The variance of $\hat{b}(\lambda)$ is estimated by $\hat{V}[\hat{b}(\lambda)] = \lambda^2\hat{V}(\hat{\Delta}e) + \hat{V}(\hat{\Delta}c) - 2\lambda\hat{C}(\hat{\Delta}e, \hat{\Delta}c)$.

Thus the null hypothesis H_0: $b(\lambda) \leqslant 0$, versus the alternative hypothesis H_1: $b(\lambda) > 0$, can be rejected at the level α if $\hat{b}(\lambda)/\{\hat{V}[\hat{b}(\lambda)]\}^{\frac{1}{2}}$ exceeds $Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)^{\text{th}}$ percentile of the standard normal random variable. If H_0 is rejected, there is evidence that T is cost-effective. In addition, the $100(1-2\alpha)\%$ confidence limits for $b(\lambda)$ are given by $\hat{b}(\lambda) \pm Z_{1-\alpha}\{\hat{V}[\hat{b}(\lambda)]\}^{\frac{1}{2}}$. Most often an INB analysis is performed as a function of λ and illustrated with a plot of $\hat{b}(\lambda)$ and the corresponding confidence limits versus λ.

Let n_T and n_S be the number of patients randomized to T and S, respectively. Let e_{ji} be 1 if the ith patient randomized to j is a success, 0 otherwise, where $j = $ T or S, and $i = 1, 2, \ldots n_j$. Let c_{ji} be the total cost for the ith patient randomized to j. Let $\bar{e}j$ and $\bar{c}j$ be the sample means (averages) of the values for e_{ji} and c_{ji}, respectively. Therefore, $\bar{e}_T$ and $\bar{e}_S$ are the proportion of successes for patients randomized to T and S, respectively, and provide the estimator for the probability of success in each arm. The parameter Δ_e is estimated by $\bar{e}_T - \bar{e}_S$. Likewise, Δ_c is estimated by $\bar{c}_T - \bar{c}_S$. The

Fig. 3



INB by lambda for each quadrant of the cost-effectiveness plane.

estimators of the variance of $\hat{\Delta}_e$, the variance of $\hat{\Delta}_c$, and the covariance between ar $\hat{\Delta}_e$ and $\hat{\Delta}_c$ are given in Appendix 1.

The CADET-Hp Trial

The CADET-Hp Trial was a double blind, placebo controlled, parallel group, multi-centre, randomized controlled trial, performed in 36 family practitioner centres across Canada [24]. Patients 18 years and over with uninvestigated dyspepsia of at least moderate severity presenting to their family physicians were eligible, provided they did not have any alarm symptoms and were eligible for empirical drug therapy. Patients were randomized between treatment (T) with omeprazole 20 mg, metronidazole 500 mg and clarithromycin 250 mg; and standard therapy (S) omeprazole 20 mg, placebo metronidazole and placebo clarithromycin.

Both regimens were given twice daily for 7 days. Treatment success was defined as the presence of no or minimal dyspepsia symptoms at 1 year. Total costs were determined from the societal perspective and are given in Canadian dollars.

A summary of the parameter estimates is given in Table 2. T is observed to increase success and decrease costs, and consequently, the point ($\hat{\Delta}e$, $\hat{\Delta}c$) lies in the SE (win–win) quadrant and $\hat{b}(\lambda)$ is positive for all positive values of λ. Suppose the WTP for a success was \$1000, then if the INB evaluated at 1000 is less than or equal to zero, T should not be adopted. This, then, becomes the null hypothesis, expressed below as H_0. On the other hand, if the INB evaluated at 1000 is greater than zero, T should be adopted. This becomes the alternative hypothesis, expressed below as H_1. Thus the investigators wish to test the hypothesis H_0: $b(1000) \leqslant 0$ versus H_1: $b(1000) > 0$ at the 5% level. That is, if the WTP for success is \$1000, is there evidence that T is cost-effective? First, the investigators must determine

$$\hat{b}(1000) = 1000 \times 0.1371 - (-53.01) = 190.11,$$

and

$$\hat{V}[\hat{b}(1000)] = 1000^2 \times 0.003356 + 4792 - 2 \times 1000 \times (-0.7129) = 9574.$$

Then the Z-statistic for testing H_0 is ar $\hat{b}(1000)/\{\hat{V}[\hat{b}(1000)]\}^{\frac{1}{2}} = 190.11/97.85 = 1.943$. Since $1.943 > z_{0.95}$ $(= 1.645)$, H_0 can be rejected in favour of H_1, and the investigators can conclude that there is evidence that T is cost-effective compared to S if the WTP is at least \$1000. The 90% confidence limits for $b(1000)$ are given by

$$\hat{b}(1000) \pm z_{0.95}\{\hat{V}[\hat{b}(1000)]\}^{\frac{1}{2}}$$

$$= 190.11 \pm 1.645 \times 97.85$$

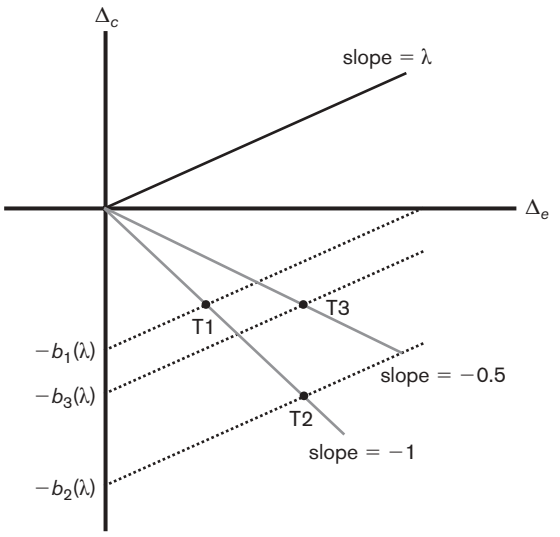
$$= 29.15, 351.1.$$

Since the lower limit is greater than 0, we can reject H_0 in favour of H_1, at the 5% level. A two-sided 90% confidence interval is constructed because we wish to test a one-sided null hypothesis at the 5% level.

For any given value of λ, $\hat{b}(\lambda)$ and $\hat{V}[\hat{b}(\lambda)]$ are given by $0.1371\lambda + 53.01$ and $0.003356\lambda^2 + 4792 + 1.426\lambda$, respectively. The quantity $\hat{b}(\lambda)$ and corresponding confidence limits can be calculated for a large range of λ and plotted as shown in Figure 5. The plot of $\hat{b}(\lambda)$ has slope $0.1371(\Delta_e)$, crosses the vertical axis at $53.01(-\hat{\Delta}c)$ and the horizontal axis at −386.7 (= ICER). By observing where the confidence limits cross the vertical axis, one can make inference regarding the difference between arms with respect to cost, in what is essentially a cost-minimization analysis. Since the confidence interval includes 0, the null hypothesis of no difference in mean cost cannot be rejected at the 10% level (i.e. this is a two-sided test). The horizontal intercepts of the confidence limits define the Fieller confidence limits for the ICER [5,6]. Therefore, by focusing attention on the horizontal axis one can perform a cost-effectiveness analysis using an ICER approach. A direct method for calculating the Fieller confidence limits for ICER is given in Appendix 2. The 90% confidence limits for the ICER are −1710 and 611.1. The hypothesis that the ICER is greater than any value above 611.1 can be rejected at the 5% level since 611.1 is the upper limit of the 90% confidence interval.

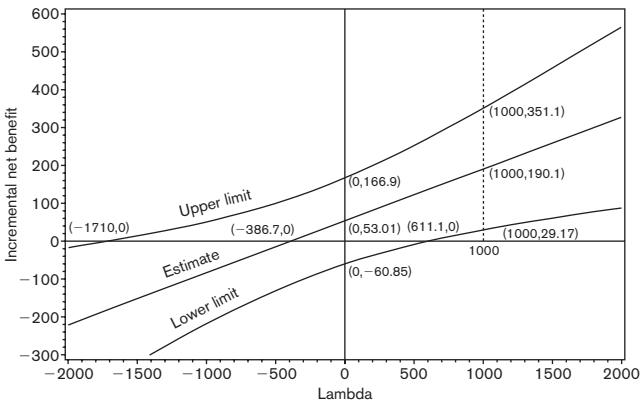
For any given value of λ one can perform a cost-

Fig. 4



Ordering INB on the cost-effectiveness plane.

Fig. 5



INB by lambda for the CADET-Hp Trial.

Table 2 Parameter estimates for the CADET-Hp Trial

Parameter	Treatment, T ($n_T = 142$)	Standard therapy, S ($n_S = 146$)	Difference or sum
$\bar{e}_j$	0.507	0.3699	difference $= \hat{\Delta}_e = 0.1371$
$\bar{c}_j$	476.97	529.98	difference $= \hat{\Delta}_c = -53.01$
$\hat{V}(\bar{e}_j)$	0.00176	0.001596	sum $= \hat{V}(\hat{\Delta}_e) = 0.003356$
$\hat{V}(\bar{c}_j)$	2167	2625	sum $= \hat{V}(\hat{\Delta}_c) = 4792$
$\hat{C}(\bar{e}_j, \bar{c}_j)$	−0.2963	−0.4166	sum $= \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) = -0.7129$

effectiveness analysis using an INB approach. In particular, for $\lambda = 1000$, the confidence interval contains only positive values, which means the null hypothesis that $b(1000) \leqslant 0$ can be rejected and that there is evidence to support the adoption of T. This is true for any value of λ that is greater than 611.1 (upper limit of the ICER). For values of λ less than 611.1 the confidence interval includes negative values and the null hypothesis that $b(1000) \leqslant 0$ cannot be rejected.

Discussion

In this paper I have provided an illustrated overview of the use of the incremental net benefit approach in assessing the cost-effectiveness of a new health care intervention. Detailed methods for performing a statistical analysis of cost-effectiveness data from a clinical trial in which the measure of effectiveness is binary are given in section 2 and illustrated in section 3. The estimators of $\hat{V}(\hat{\Delta}e)$ and $\hat{C}(\hat{\Delta}e\hat{\Delta}c)$ appropriate for when the measure of effectiveness is continuous, such as survival time (i.e. time from randomization to death), are given in Appendix 3.

Presenting the results of an INB analysis as a function of the WTP allows a reader to perform analysis for a value of the WTP most appropriate to them. Furthermore, a reader can concentrate on the horizontal axis to perform an analysis of the ICER, or the vertical axis for an assessment of the between-treatment difference in costs.

In the methods provided above there is an assumption that all patients are followed for the duration of interest. In the example from the CADET-Hp Trial the durations of interest were 1 year because the primary outcome was defined as no or minimal dyspepsia symptoms at 1 year. In many trials, especially for those with longer duration of interest, some patients will not be followed for the entire duration. These patients are said to be censored. There are two main reasons for censoring. The first, called administrative censoring, occurs because the final analysis is performed at a time after the last patient was randomized that is shorter than the duration of interest. If the duration of interest is 5 years, as it might be for a cancer trial, then administrative censoring will occur if the analysis is performed, as it often is, within 5 years of the last patient being randomized. The second reason for censoring, referred to as lost to follow-up occurs simply because some patients will fail to show up for follow-up clinic visits, perhaps because they have moved away or because they have lost their motivation. Statistical analysis for censored data, while still only requiring that the five parameters of interest be estimated, is decidedly more complex. There are two main approaches for analysing censored data in the context of a cost-effectiveness analysis: the direct method [21,22,25] and inverse probability weighting [23,26–31].

The methods given in the statistical analysis section for estimating the mean and variance of cost are based on the least-squares methodology. The use of least-squares for cost data is often criticized [32] because of the presence of right-skewing. Nonetheless, the robustness of least-squares methodology for skewed data has been repeatedly demonstrated [5,33–35].

Acknowledgements

The author is grateful to Dr N. Chiba and AstraZeneca Canada, Inc. for supplying the data from the CADET-Hp Trial, to Dr Gary Foster for producing the figures, and to Ms Susan Tomlinson for her meticulous proofreading. The opinions expressed are those of the author and should not be attributed to any funding agency.

References

1 O'Brien BJ, Drummond MF, Labelle RJ, Willan AR. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in heath care. *Medical Care* 1994; **32**:150–163.
2 Mullahy J, Manning W. Statistical issues of cost-effectiveness analysis. In: Sloan F (editor). *Valuing Health Care*. Cambridge: Cambridge University Press; 1994, pp. 149–184.
3 van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E ratios alongside a clinical trial. *Health Economics* 1994; **3**:309–319.
4 Wakker P, Klaassen MP. Confidence intervals for cost/effectiveness ratios. *Health Economics* 1995; **4**:373–381.
5 Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Economics* 1996; **5**:297–305.
6 Chaudhary MA, Stearns SC. Confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statistics in Medicine* 1996; **15**:1447–1458.
7 Mullahy J. What you don't know can't hurt you? Statistical issues and standards for medical technology evaluation. *Medical Care* 1996; **34 (12 suppl)**:DS124–DS135.
8 Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC (editors). *Cost Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
9 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**:327–340.
10 Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics* 1997; **6**:243–252.
11 Phelps CE, Mushlin AI. On the (near) equivalence of cost-effectiveness and cost-benefit analysis. *Intern J Technol Assessment in Health Care* 1991; **7**:12–21.
12 Ament A, Baltussen R. The interpretation of results of economic evaluation: explicating the value of health. *Health Economics* 1997; **6**: 625–635.
13 Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998; **18 (suppl)**:S68–S80.
14 Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *Intern J Technol Assessment* 1998; **14**:467–471.
15 van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E ratios alongside a clinical trial. *Health Economics* 1994; **3**:309–319.
16 Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics* 1998; **7**:723–740.
17 Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999; **8**:257–261.
18 Lothgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics* 2000; **9**: 623–630.
19 Heitjan DF. Fieller's method and net health benefit. *Health Economics* 2000; **9**:327–335.

20 Willan AR. Analysis, sample size and power for estimating incremental net health benefit from clinical trial data. *Controlled Clinical Trials* 2001; **22**:228–237.
21 Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Statistics in Medicine* 2001; **20**:1563–1574.
22 Willan AR, Chen EB, Cook RJ, Lin DY. Incremental net benefit in clinical trials with quality-adjusted survival. *Statistics in Medicine* 2003; **22**: 353–362.
23 Willan AR, Lin DY, Cook RJ, Chen EB. Using inverse-weighting in cost-effectiveness analysis with censored data. *Statistical Methods in Medical Research* 2002; **11**:539–551.
24 Chiba N, van Zanten SJ, Sinclair P, Ferguson RA, Escobedo S, Grace E. Treating *Helicobacter pylori* infection in primary care patients with uninvestigated dyspepsia: the Canadian adult dyspepsia empiric treatment – *Helicobacter pylori* positive (CADET-Hp) randomised controlled trial. *BMJ* 2002; **324**:1012–1016.
25 Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**:419–434.
26 Lin DY. Linear regression analysis of censored medical costs. *Biostatistics* 2000; **1**:35–47.
27 Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika* 2000; **87**:329–343.
28 Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics* 2001; **57**: 1002–1008.
29 Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* 1997: **84**:339–348.
30 Zhao H, Tsiatis AA. Efficient estimation of the distribution of quality adjusted survival time. *Biometrics* 1999; **55**:1101–1107.
31 Zhao H, Tsiatis AA. Testing equality of survival functions of quality adjusted lifetime. *Biometrics* 2001; **57**:861–867.
32 O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Economics* (in press).
33 Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health* 2002; **23**:151–169.
34 Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *BMJ* 2000; **320**:1197–1200.
35 Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Economics* (submitted).

Appendix 1. Estimators of variance and covariance

$$\hat{\mathrm{V}}(\hat{\Delta}e) = \bar{e}_T(1-\bar{e}_T)/n_T + \bar{e}_S)/n_S.$$

$$\operatorname{var}\hat{\Delta}C = \sum_{i=1}^{nT}\frac{(c_{Ti}-\bar{c}_T)^2}{n_T(n_T-1)} + \sum_{i=1}^{nS}\frac{(c_{Si}-\bar{c}_S)^2}{n_S(n_S-1)}.$$

$$C(\hat{\Delta}e,\hat{\Delta}c) = \frac{(c_{T\{1\}}-n_T\bar{e}_T\bar{c}_T)}{n_T(n_T-1)} + \frac{(c_{S\{1\}}-n_S\bar{e}_S\bar{c}_S)}{n_S(n_S-1)},$$

where $c_{j\{1\}}$ is the sum of the costs for all the patients on the jth treatment who had a success, i.e.

$$c_{j\{1\}} = \sum_{i=1}^{n_j} e_{ji}c_{ji}.$$

Appendix 2. The Fieller confidence interval for the ICER

The ICER is estimated by $\hat{\Delta}c/\hat{\Delta}e$, with the Fieller [5,6] solution for the $100(1-2\alpha)\%$ confidence limits given by

$$(\hat{\Delta}c/\hat{\Delta}e)\{[1-z_{1-\alpha}^2 a_{ec} \pm z_{1-\alpha}$$

$$[a_e + a_c - z_{1-\alpha}^2(a_e a_c - a_{ec}^2]^{\frac{1}{2}}]/(1-z_{1-\alpha}^2 a_e)\},$$

where $a_e = \hat{\sigma}^2_{\Delta_e}/\hat{\Delta}_e^2$, $a_c = \hat{\sigma}^2_{\Delta_c}/\hat{\Delta}_c^2$, $a_{ec} = \hat{\sigma}_{\Delta_c\Delta_e}/(\hat{\Delta}_e\hat{\Delta}_c)$, and $z_{1-\alpha}$ is the $100(1-\alpha)^{\text{th}}$ percentile of a standard normal random variable.

Appendix 3. Parameter estimates for continuous measure of effectiveness

$$\hat{V}(\hat{\Delta}e) = \sum_{i=1}^{nT}\frac{(e_{Ti}-\bar{e}_T)^2}{n_T(n_T-1)} + \sum_{i=1}^{nS}\frac{(e_{Si}-\bar{e}_S)^2}{n_S(n_S-1)}$$

and

$$\hat{C}(\hat{\Delta}e,\,\hat{\Delta}c) = \sum_{i=1}^{nT}\frac{(e_{Ti}-\bar{e}_T)(c_{Ti}-\bar{c}_T)}{n_T(n_T-1)} + \sum_{i=1}^{nS}\frac{(e_{si}-\bar{e}_S)(c_{si}-\bar{c}_S)}{n_s(n_S-1)},$$

where e_{ji} is the measure of effectiveness on the ith patient of the jth treatment, and $\bar{e}_j$ is the sample mean (average) of the values of e_{ji}, i.e.

$$\bar{e}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} e_{ji}.$$