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.

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 \( \lambda \)), or at the very least, the analysis must be presented as a function of \( \lambda \) so that readers can apply the WTP most appropriate for them.

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 dyspeptic symptoms at 1 year. Let \( \Delta_e \) 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 \( \Delta_c \) 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

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
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 \( (1/\Delta_e) \times \Delta_e \), which is equal to \( \Delta_e/\Delta_e \), 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 \( \Delta_e/\Delta_e \) 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 \( \Delta_e \) 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 \( \Delta_e \) by \( \Delta_e \), as shown in Figure 1. The slope of the line connecting the origin to the point \( (\Delta_e,\Delta_e) \) is the ICER. For illustration, the line through the origin with slope equal to the WTP (\( \lambda \)) 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 \( \lambda \). 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 \( \lambda \).

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 \( \Delta_e \) and \( \Delta_e \) 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 \( b(\lambda) = \Delta_e \lambda - \Delta_e \). 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 \( \Delta_e \). On the cost-effectiveness plane, \( b(\lambda) \) is the vertical distance between the point \( (\Delta_e,\Delta_e) \) 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

![Fig. 1](image-url)
of success: for example, avoiding an ulcer and avoiding an ulcer bleeding incident. Let $\Delta_{c,j}$ and $\Delta_{c,z}$ be the difference in the probabilities of avoiding an ulcer and an ulcer bleed, respectively. Then the INB can be defined as $b(\lambda) = \Delta_{c,j}h_j + \Delta_{c,z}h_z - \Delta_c$, where $h_j$ and $h_z$ are the WTP to avoid an ulcer and an ulcer bleed, respectively. Also INB is linear in $\lambda$ and thus a plot of $b(\lambda)$ by $\lambda$ is a straight line in which the slope is $\Delta_c$, 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 $\lambda$, yielding $\Delta_c/\Delta_c$. The characteristics of the plot of $b(\lambda)$ by $\lambda$ depend on which quadrant of the cost-effectiveness plane the point $(\Delta_c,\Delta_c)$ lies. This is illustrated in Figure 3. In the NE quadrant (panel b of Fig. 3) $\Delta_c$, $\Delta_c$, and the ICER are all positive. Therefore, the plot of $b(\lambda)$ has positive slope ($\Delta_c$), positive horizontal intercept ($\Delta_c/\Delta_c$), and negative vertical intercept ($-\Delta_c$). In the SE quadrant (panel d) $\Delta_c$ is positive, and $\Delta_c$ and the ICER are negative. Therefore, the plot of $\lambda$ has positive slope, negative horizontal intercept, and positive vertical intercept. In the SW quadrant (panel c) $\Delta_c$ and $\Delta_c$ 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) $\Delta_c$ is negative, $\Delta_c$ 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 $\lambda$ is particularly useful when a cost-effectiveness analysis is performed before the value of $\lambda$ 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 $T_1$, $T_2$ and $T_3$, 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 $(\Delta_c,\Delta_c)$ and the threshold. Therefore, the vertical intercept of the line passing through the point $(\Delta_c,\Delta_c)$ with slope $\lambda$ 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 $\lambda$, 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 $\lambda$, $b(\lambda)$ is the $\lambda$-th percentile of the standard normal random variable $Z$. The variance of $b(\lambda)$ is $\lambda^2$.

Thus the null hypothesis $H_0: b(\lambda) = 0$, versus the alternative hypothesis $H_1: b(\lambda) > 0$, can be rejected at the level $\alpha$ if $|b(\lambda)| = \frac{\lambda}{\sqrt{V[b(\lambda)]}} > z_{1-\alpha}$, where $z_{1-\alpha}$ is the 100(1 $-\alpha$)% 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 $b(\lambda) \pm z_{1-\alpha}\sqrt{V[b(\lambda)]}$. Most often an INB analysis is performed as a function of $\lambda$, and illustrated with a plot of $b(\lambda)$ and the corresponding confidence limits versus $\lambda$.

Let $n_T$ and $n_S$ be the number of patients randomized to $T$ and $S$, respectively. Let $e_j$ be 1 if the $i$th patient randomized to $j$ is a success, 0 otherwise, where $j = T$ or $S$, and $i = 1, 2, \ldots, n_j$. Let $e_j$ be the total cost for the $i$th patient randomized to $j$. Let $\bar{e}_T$ and $\bar{e}_S$ be the sample means (averages) of the values for $e_j$, $e_j$, 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 $\Delta_c$ is estimated by $\hat{\Delta}_c = \bar{e}_T - \bar{e}_S$. The
estimators of the variance of $\tilde{\Delta}_e$, the variance of $\tilde{\Delta}_c$, and the covariance between $\tilde{\Delta}_e$ and $\tilde{\Delta}_c$ 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 $(\Delta_e, \Delta_c)$ lies in the SE (win–win) quadrant and $b(\lambda)$ is positive for all positive values of $\lambda$. 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 $H_0$: $b(1000) \leq 0$ 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

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

Fig. 3

(a) NW: $\Delta_e < 0; \Delta_c > 0$
(b) NE: $\Delta_e > 0; \Delta_c > 0$
(c) SW: $\Delta_e < 0; \Delta_c < 0$
(d) SE: $\Delta_e > 0; \Delta_c < 0$
\[ \hat{b}(1000) = 1000 \times 0.1371 - (-53.01) = 190.11, \]

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 at \( \hat{b}(1000)/[\hat{V}[\hat{b}(1000)]^{1/2}] = 190.11/97.85 = 1.943. \) 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 \( \hat{b}(1000) \) are given by

\[ \hat{b}(1000) \pm z_{0.95} \{ \hat{V}[\hat{b}(1000)] \}^{1/2} = 190.11 \pm 1.645 \times 97.85 = 29.15, 351.1. \]

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 \( \lambda, \hat{b}(\lambda) \) 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 \( \lambda \) and plotted as shown in Figure 5. The plot of \( \hat{b}(\lambda) \) has slope 0.1371(\( \Delta_c \)), crosses the vertical axis at 53.01(\( -\Delta_c \)) 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 \( \lambda \) one can perform a cost-

---

**Table 2** Parameter estimates for the CADET-Hp Trial

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Treatment, T (( n_T = 142 ))</th>
<th>Standard therapy, S (( n_S = 146 ))</th>
<th>Difference or sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta )</td>
<td>0.507</td>
<td>0.3699</td>
<td>difference = ( \Delta_e = 0.1371 )</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>476.97</td>
<td>529.98</td>
<td>difference = ( \Delta_e = -53.01 )</td>
</tr>
<tr>
<td>( V(\pi) )</td>
<td>0.00176</td>
<td>0.001596</td>
<td>sum = ( V(\Delta_e) = 0.003356 )</td>
</tr>
<tr>
<td>( V(\sigma) )</td>
<td>2167</td>
<td>2625</td>
<td>sum = ( V(\Delta_e) = 4792 )</td>
</tr>
<tr>
<td>( C(\pi, \sigma) )</td>
<td>-0.2963</td>
<td>-0.4166</td>
<td>sum = ( C(\Delta_e, \Delta_e) = -0.7129 )</td>
</tr>
</tbody>
</table>
effectiveness analysis using an INB approach. In particular, for \( \lambda = 1000 \), the confidence interval contains only positive values, which means the null hypothesis that \( b(1000) \leq 0 \) can be rejected and that there is evidence to support the adoption of T. This is true for any value of \( \lambda \) that is greater than 611.1 (upper limit of the ICER). For values of \( \lambda \) less than 611.1 the confidence interval includes negative values and the null hypothesis that \( b(1000) \leq 0 \) 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 \( V(\Delta e) \) and \( C(\Delta e \Delta e) \) 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 analyzing 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


Appendix 1. Estimators of variance and covariance

\[
\hat{V}(\Delta e) = \frac{e_T(1 - \hat{e}_T)}{n_T + \hat{e}_S}/n_S.
\]

\[
\text{var} \hat{\Delta e} = \frac{\sum_{i=1}^{n_T} (e_{Ti} - \hat{e}_T)^2}{n_T(n_T - 1)} + \frac{\sum_{i=1}^{n_S} (e_{Si} - \hat{e}_S)^2}{n_S(n_S - 1)}.
\]

\[
C(\Delta e, \hat{\Delta e}) = \frac{(e_{Ti} - \hat{e}_T)(e_{Si} - \hat{e}_S)}{n_T(n_T - 1)} + \frac{(e_{S[i]} - \hat{e}_S)(e_{Ti} - \hat{e}_T)}{n_S(n_S - 1)},
\]

where \(e_{j[1]}\) is the sum of the costs for all the patients on the \(j\)th treatment who had a success, i.e.

\[
\hat{e}_{j[1]} = \sum_{i=1}^{n_j} e_{ji}c_{ji}.
\]

Appendix 2. The Fieller confidence interval for the ICER

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

\[
(\Delta c/\hat{\Delta e})\{1 - z_{1-\alpha}^2(a_c - a_{\hat{\alpha}})^2/[1 - z_{1-\alpha}^2\hat{a}_e]\},
\]

where \(a_c = \hat{\sigma}_\Delta^2/\hat{\Delta}_c^2\), \(a_{\hat{\alpha}} = \hat{\sigma}_\Delta^2/\hat{\Delta}_\alpha^2\), \(a_{\hat{\alpha}} = \hat{\sigma}_\Delta^2/(\hat{\Delta}_\alpha\hat{\Delta}_\alpha)\), and \(z_{1-\alpha}\) is the 100(1 - \(\alpha\))% percentile of a standard normal random variable.