Incremental net benefit in randomized clinical trials

Andrew R. Willan[1,*,†] and D. Y. Lin[2]

[1]*Department of Clinical Epidemiology and Biostatistics, McMaster University and Centre for Evaluation of Medicines, St Joseph's Hospital, 600-143 James Street South, Hamilton, ON, L8P 3A1, Canada*
[2] *Department of Biostatistics, Box 357232, University of Washington, Seattle, WA 98195, U.S.A.*

SUMMARY

There are three approaches to health economic evaluation for comparing two therapies. These are (i) cost minimization, in which one assumes or observes no difference in effectiveness, (ii) incremental cost-effectiveness, and (iii) incremental net benefit. The latter can be expressed either in units of effectiveness or costs. When analysing data from a clinical trial, expressing incremental net benefit in units of cost allows the investigator to examine all three approaches in a single graph, complete with the corresponding statistical inferences. Furthermore, if costs and effectiveness are not censored, this can be achieved using common two-sample statistical procedures. The above will be illustrated using two examples, one with censoring and one without. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

It is becoming quite common in randomized controlled trials to collect patient-level cost data along with measures of effectiveness. This allows conventional principles of statistical inference to be used to quantify the uncertainty in cost-effectiveness analyses. Initial efforts were concentrated on providing confidence intervals for incremental cost-effectiveness ratios (ICER) [1–10], but more recently the concept of incremental net benefit (INB) [11–14] has been proposed as an alternative. Using a net benefit approach solves a number of problems associated with the ICER. The confidence intervals for the ICER can include undefined values or may even be completely undefined. It fails to account for opportunity costs, and even more perplexing, two totally opposite results can have the same ICER. Also, negative ICERs are difficult to interpret and are not properly ordered.

In this paper we show that by varying the value given to a unit of effectiveness, the net benefit approach can include cost minimization and ICER analyses. The statistical model is outlined in the rest of this section. The methods for uncensored and censored data are given in Section 2, followed by examples of both in Section 3. Power and sample size issues are discussed in Section 4.

---

*Correspondence to: Andrew R. Willan, Department of Clinical Epidemiology and Biostatistics, McMaster University and Centre for Evaluation of Medicines, St Joseph's Hospital, 600-143 James Street South, Hamilton, ON, L8P 3A1, Canada
†E-mail: willana@mcmaster.ca

2. METHODS

2.1. The model

In a two-arm randomized control trial let e_{ji} and c_{ji} be the respective measures of effectiveness and cost for patient i on therapy j, where $j = $ T (treatment), S (standard), $i = 1, 2, \ldots, n_j$ and n_j is the respective sample size. Let

$$E \begin{pmatrix} e_{ji} \\ c_{ji} \end{pmatrix} = \begin{pmatrix} \mu_j \\ v_j \end{pmatrix}; \quad V \begin{pmatrix} e_{ji} \\ c_{ji} \end{pmatrix} = \begin{pmatrix} \sigma_j^2 & \rho_j \sigma_j \omega_j \\ \rho_j \sigma_j \omega_j & \omega_j^2 \end{pmatrix}; \quad \mu = \mu_T - \mu_S \quad \text{and} \quad v = v_T = v_S$$

where V is the variance-covariance function. Typically, e_{ji} is the patient's survival time (perhaps quality-adjusted) from randomization to death. Unless all patients are followed until death or for the entire duration of interest, censoring is said to occur.

In cost minimization, where μ is assumed to be zero, the parameter of interest is v, which is the added cost of giving a patient treatment rather than standard. A negative difference favours treatment and a positive difference favours standard. The ICER is defined as v/μ, and measures the increase in cost of achieving each extra unit of effectiveness (for example, year of life) using treatment rather than standard. The incremental net benefit, expressed in costs, is defined as $b = \mu\lambda - v$, where λ is the value given to a unit of effectiveness. The quantity b is the net benefit, expressed in money, of giving a patient treatment rather than standard, with positive difference favouring treatment and a negative difference favouring standard. Typically λ is varied in a sensitivity analysis, and incremental net benefit is expressed as $b(\lambda)$.

Noting that $b(0) = -v$ illustrates that cost minimization is a special case of net benefit. Similarly, observing that $b(v/\mu) = 0$ demonstrates the connection between the ICER and incremental net benefit. Therefore, in a cost-effectiveness analysis, one need only estimate $b(\lambda)$, and its confidence limits, and graph them as a function of λ. These graphs cross the vertical axis at minus the cost difference and the horizontal axis at the ICER, defining the respective estimates and the corresponding confidence intervals. The confidence interval for the ICER given by the horizontal intercepts are identical to those provided by Fieller's theorem [2, 3].

2.2. Uncensored data

If all patients are followed until death or for the entire duration of interest, then the observed measures of effectiveness and cost are not censored. In this case the sample means, variance and covariances can be used to estimate the model parameters, and the estimator of $b(\lambda)$ and its estimated variance are given by

$$\hat{b}(\lambda) = (\hat{\mu}_T - \hat{\mu}_S)\lambda - (\hat{v}_T - \hat{v}_S) = \hat{\mu}\lambda - \hat{v}$$

and

$$\hat{V}[\hat{b}(\lambda)] = \sum_{j=S,T} \frac{1}{n_j}(\hat{\sigma}_j^2 \lambda^2 + \hat{\omega}_j^2 - 2\hat{\sigma}_j \hat{\omega}_j \hat{\rho}_j \lambda)$$

respectively. The $100(1-\alpha)$ per cent confidence limits are given by $\hat{b}(\lambda) \pm z_{(1-\alpha/2)}\{\hat{V}[\hat{b}(\lambda)]\}^{1/2}$, where $z_{(1-\alpha/2)}$ is the $(1 - \frac{\alpha}{2})$ 100th percentile of the standard normal distribution.

Alternatively, identical results can be achieved by defining net benefit for each patient as $b_{ji}(\lambda) = e_{ji}\lambda + c_{ji}$, and defining $\hat{b}_j(\lambda)$ and $\hat{v}_j(\lambda)$ as the sample means and variances, respectively. The estimator of $b(\lambda)$ and its estimated variance are then given by

$$\hat{b}(\lambda) = \hat{b}_{\mathrm{T}}(\lambda) - \hat{b}_{\mathrm{S}}(\lambda)$$

and

$$\hat{V}[\hat{b}(\lambda)] = \sum_{j=\mathrm{S,T}} \frac{\hat{v}_j(\lambda)}{n_j}$$

$$= \sum_{j=\mathrm{S,T}} \frac{1}{n_j(n_j-1)} \left[\lambda^2 \sum_{i=1}^{n_j}(e_{ji}-\hat{\mu}_j)^2 + \sum_{i=1}^{n_j}(c_{ji}-\hat{v}_j)^2 - 2\lambda\sum_{i=1}^{n_j}(e_{ji}-\hat{\mu}_j)(c_{ji}-\hat{v}_j) \right]$$

Thus, net benefit analysis for uncensored data can be accomplished using common two-sample procedures.

By calculating $\hat{b}(\lambda)$ and the confidence limits for a wide range of λ, and graphing them as a function of λ, inference regarding incremental net benefit can be made as a function of λ. The estimate and confidence limits for the difference in cost are given by the vertical intercepts and for the ICER by the horizontal intercepts.

2.3. Censored data

If patients are not followed until death or for the entire duration of interest, the data will be censored and some of the c_{ji}'s and e_{ji}'s will not be observed. In this case, assuming that duration of survival is the primary measure of effectiveness, life-table methods must be employed to estimate the means, variances and covariances. Let the duration of interest be τ, and let t_{jk} ($k = 1, 2, \ldots, L_j$) be all the times to death for patients on therapy j. (The t_{jk}'s are the unique L_j values of the observed e_{ji}'s.) The product-limit estimates [15] of the survival curves are

$$\hat{S}_j(t) = \prod_{l=0}^{k-1}\left(1 - \frac{d_{jl}}{n_{jl}}\right)$$

for $t_{j,k-1} \leqslant t < t_{jk}$ where d_{jl} are the number of deaths on therapy j at time t_{jl}, n_{jl} are the number of patients still at risk at time $t_{jl}, t_{j0} = d_{j0} = 0$ and $n_{j0} = n_j$. The estimates of mean survival time [16] is the area under the curves to time τ and is given by

$$\hat{\mu}_j = \int_0^\tau \hat{S}_j(u)\,\mathrm{d}u = \sum_{k:0\leqslant t_{jk} < \tau}\{\hat{S}_j(t_{jk})[\min(t_{j,k+1},\tau) - t_{j,k}]\}$$

The formulation of $\hat{V}(\hat{\mu}_j)$, the variance estimator for $\hat{\mu}_j$, is given in the Appendix. These methods assume that the censoring in non-informative.

A similar approach, in which the costs accumulated until death replace the t_{jk}'s, cannot be used for estimating the mean cost because the censoring is informative on the cost scale. Patients accumulate costs at different rates. Therefore, there is a positive correlation between the costs a patient accumulates until death and the costs he or she accumulates until censoring.

As a result, the life-table estimate of mean cost is positively biased. A valid method is proposed by Lin *et al.* [17]. In this method the duration of interest is divided into the K intervals $[\alpha_k, \alpha_{k+1})$, where $0 = \alpha_1 < \alpha_2 < \cdots < \alpha_{K+1} = \tau$. The quantity m_{jk} is defined as the number of patients on therapy j who are alive at α_k and are not censored (although may die) before α_{k+1}, and c_{jk} is the average cost incurred during the interval $[\alpha_k, \alpha_{k+1})$ by these m_{jk} patients. Then $\hat{v}_j = \sum_{k=1}^{K} \hat{S}_j(\alpha_k) c_{jk}$. The formulations of $\hat{V}(\hat{v}_j)$, the variance estimator for $\hat{v}_j$, and $\hat{C}(\hat{\mu}_j, \hat{v}_j)$, the estimator of the covariance between $\hat{\mu}_j$ and $\hat{v}_j$, are given in the Appendix.

3. EXAMPLES

3.1. Uncensored data – prostate

In a trial of symptomatic hormone resistant prostate cancer [18, 19], 161 patients were randomized between prednisone alone (S) and prednisone plus mitoxantrone (T). Although there was no statistically significant difference in survival, there was better palliation with T. Cost data, including hospital admissions, outpatient visits, investigations, therapies and palliative care, were collected retrospectively on the 114 patients from the three largest centres. Survival was quality-adjusted using the EORTC quality of life questionnaire QLQ-C30. All patients were followed until death.

The sample means and the sample variance-covariance information can be found in Table I. Cost is given in Canadian dollars (CAD$) and effectiveness in quality-adjusted life-weeks (QALW). Using the values given in Table I, $\hat{b}(\lambda)$ and the corresponding 90 per cent confidence limits were calculated for various values of λ and plotted in Figure 1. For $\lambda = 1000$ (approximately \$50 000/quality-adjusted life-year) the estimated of net benefit is \$14 517 with confidence interval \$3662 to \$25 372. Ninety per cent confidence intervals were used to be consistent with a required 5 per cent test of the hypothesis $b(\lambda) = 0$ versus $b(\lambda) > 0$. The slope of $\hat{b}(\lambda)$ is positive (that is, $\hat{\mu} = 12.8 > 0$), illustrating that treatment was observed to increase effectiveness. The vertical intercept is also positive (that is, $\hat{v} = -1717 < 0$), illustrating that treatment was observed to decrease cost. This 'win–win' observation is illustrated by a negative value for the estimate of ICER, given by the horizontal intercept, of -134 CAD\$/QALW. The graphs provide confidence intervals for v of -7946 to 4512 (the negative of the vertical intercepts) and for the ICER of -1764 to 378 (the horizontal intercepts). The confidence intervals for the ICER are identical to those provided by Fieller's theorem [2, 3]. The lower 90 per cent confidence limit for the incremental net benefit is positive for values of λ greater than 378 CAD\$/QALW, that is, greater than approximately $20\,000$ CAD\$/QALY.

If there are concerns regarding the sample estimates in the presence of right skewing, as is often the case with cost data, the bootstrap estimates provide an alternative. The bootstrap estimates, using 1500 re-samples, provide a 90 per cent confidence interval for the ICER of -1802 to 362, which is very close to the Fieller interval.

Table I. Sample sizes and parameter estimates for the prostate example.

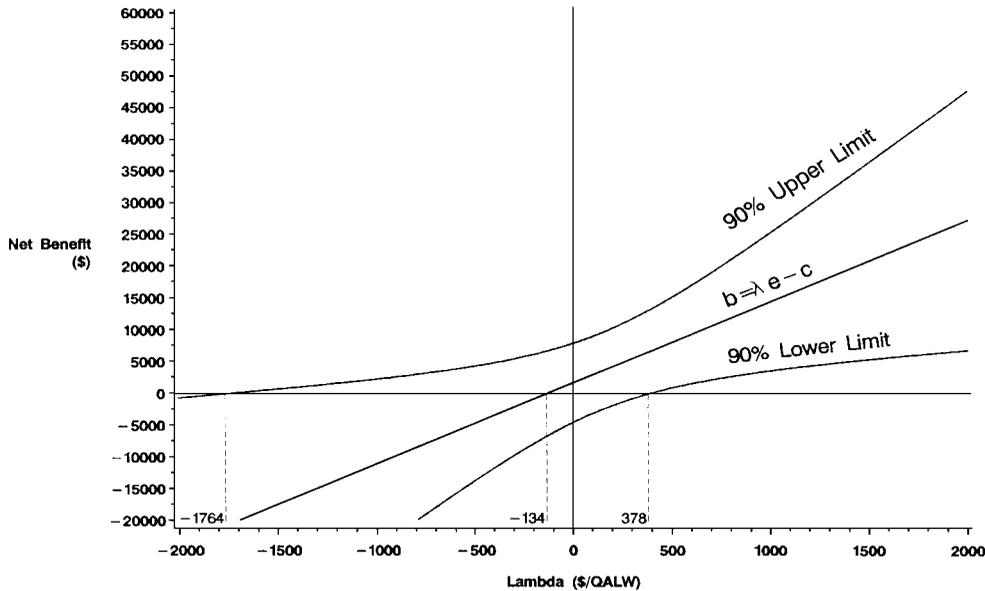|	n_j	$\hat{\mu}_j$	$\hat{v}_j$	$\hat{\sigma}_j^2/n_j$	$\hat{\omega}_j^2/n_j$	$\hat{\rho}_j \hat{\sigma}_j \hat{\omega}_j/n_j$
Standard	53	28.1	29 039	16.4	7 872 681	2876
Treatment	61	40.9	27 322	24.1	6 466 351	2771

Figure 1. Net benefit as a function of lambda for the prostate example.

Table II. Sample sizes and parameter estimates for the CIDS example.

	n_j	$\hat{\mu}_j$	$\hat{v}_j$	$\hat{V}(\hat{\mu}_j)$	$\hat{V}(\hat{v}_j)$	$\hat{C}(\hat{\mu}_j, \hat{v}_j)$
Standard	218	4.651	38 864	0.01913	6 519 142	14.20
Treatment	212	4.907	87 103	0.01754	8 461 538	124.9

3.2. Censored data – the Canadian Implantable Defibrillator Study (CIDS)

In a trial of patients at risk of cardiac arrest, 659 patients were randomized between amio-darone (S) and implantable cardioverter defibrillator (T). The measure of effectiveness was unadjusted survival. Costs were collected prospectively on 430 patients from the centres participating in the economic analysis. Not all patients were followed for the duration of interest of 77 months. By declaring a duration of interest of 77 months, there is an implicit assumption that beyond 77 months the survival curves for the two treatment groups are negligibly different and the two treatment groups are incurring costs at the same rate. Practical considerations, however, require that the duration of interest not exceed the longest survival time, since mean survival is not estimable beyond that point.

The means, variances and covariances were estimated using procedures given in Section 2.2 and the Appendix and can be found in Table II. Cost is given in CAD\$ and effectiveness in years. $\hat{b}(\lambda)$ and the corresponding 95 per cent confidence limits are plotted in Figure 2. For $\lambda = \$50\,000$ per life-year, the estimated of net benefit is $-\$35\,246$ with confidence interval $-\$54\,322$ to $-\$16\,569$. The slope of $\hat{b}(\lambda)$ is positive (that is, $\hat{\mu} = 0.26 > 0$), illustrating that treatment was observed to increase survival. The vertical intercept is negative (that
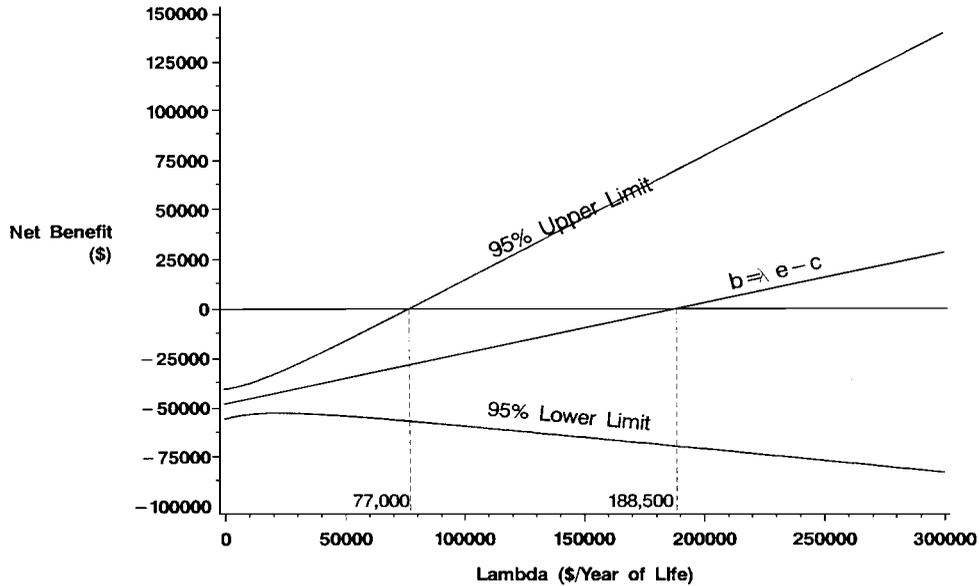
Figure 2. Net benefit as a function of lambda for the CIDS example.

is, $\hat{v} = 49\,071 > 0$), illustrating that treatment was observed to increase cost. This 'win–lose' observation is illustrated by a positive value for the estimate of the ICER, given by the horizontal intercept, of 188 500 CAD\$/life-year. The graphs provide a confidence interval for v of 40 653 to 55 825 (the negative of the vertical intercepts). The lower confidence limit for the ICER, given by the horizontal intercept of the upper confidence limit for $b(\lambda)$, is 77 000. The lower confidence limit for $b(\lambda)$ does not cross the horizontal axis, indicating that there is no upper limit for the ICER. One interprets this to mean that arbitrarily large values for the ICER are not inconsistent with the data, which is to be expected since the difference in effectiveness is not statistically significant. The confidence interval for the ICER, calculated using Fieller's theorem, is displayed on the cost-effectiveness plane in Figure 3. The positive vertical axis is contained in the confidence interval, again illustrating that arbitrarily large values for the ICER are not inconsistent with the data.

4. POWER AND SAMPLE SIZE

The formula for total sample size for net benefit is given by

$$N = \frac{2(z_{1-\alpha} + z_{1-\beta})^2 \sigma_b^2}{\delta^2}$$

$$\sigma_b^2 = \sum_{j=S,T} (\sigma_j^2 \lambda^2 + \omega_j^2 - 2\sigma_j \omega_j \rho_j \lambda)$$

and δ is the smallest clinically important value of $b(\lambda)$. If we assume that the covariance matrices are equal between treatment groups, the above sample size formula is equivalent to
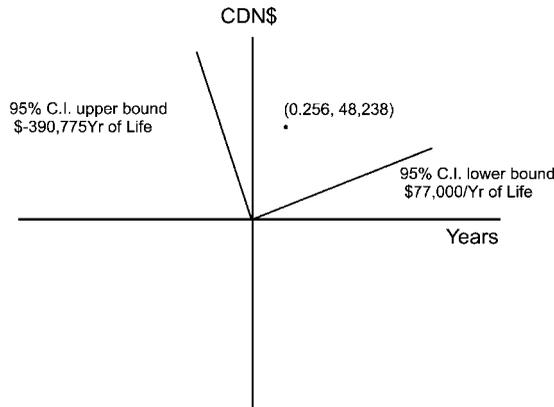
Figure 3. Cost-effectiveness plane for the CIDS example.

equation (3) in Laska *et al.* [20]. The corresponding power function is given by

$$P(\delta) = \Phi\left(\frac{\delta}{(2\frac{\sigma_b^2}{N})^{1/2}} - z_{1-\alpha}\right) = \Phi\left(\frac{\delta}{\{V[\hat{b}(\lambda)]\}^{1/2}} - z_{1-\alpha}\right)$$

where $\Phi(z_{1-\gamma}) = 1 - \gamma$. The power curve gives the probability of rejecting the hypothesis $H : b(\lambda) \leqslant 0$ in favour of the hypothesis $A : b(\lambda) > 0$, at the level α, for a given δ.

Providing an estimate of σ_b^2 prior to the study can be problematic. An estimate of σ_j^2, at least for the standard arm, should be available from previous studies. However, since collecting patient-level cost data in clinical trials has not been common in the past, estimates of ω_j^2 and ρ_j may not be readily available. As an alternative, investigators could perform a chart review of some patients with the same condition as proposed for the planned trial (most likely having received standard), collecting those healthcare utilization items that are most responsible for costs, such as days in hospital and procedures. Applying a set of prices to the healthcare utilization items provides a cost per patient from which an estimate of ω_S^2 can be obtained. Values of ρ_j could be varied in a sensitivity analysis.

Power curves for the prostate and CIDS examples are found in Figures 4 and 5, respectively. The quantity λ is equal to 1000 CAD$ per QALW in Figure 4 and 50 000 CAD$ per life-year in Figure 5. There are two graphs in each figure. One corresponds to $\alpha = 0.05$, a fairly standard approach, and the other to $\alpha = 0.5$. Willan [21] suggests that the power curve should pass through 0.5 at the point of indifference. Assuming that we are indifferent between therapies if the net benefit is zero, this is achieved by setting $\alpha = 0.5$, and H is rejected if the net benefit is observed to be positive. This approach has intuitive appeal since there is at least a 50 per cent probability of rejecting H whenever net benefit is positive. Examining Figure 4 for $\alpha = 0.5$ reveals that if this approach was used, there would be only a 5 per cent probability of rejecting H if standard is superior by 10 500 CAD$ (that is, 10.5 QALW). By symmetry, there would be a 95 per cent probability of rejecting H if treatment were superior by 10 500 CAD$. The corresponding value in Figure 5 is 470 000 CAD$ (that is, 9.4 years of life). This is an indication that the prostate trial is more powerful than the CIDS trial.
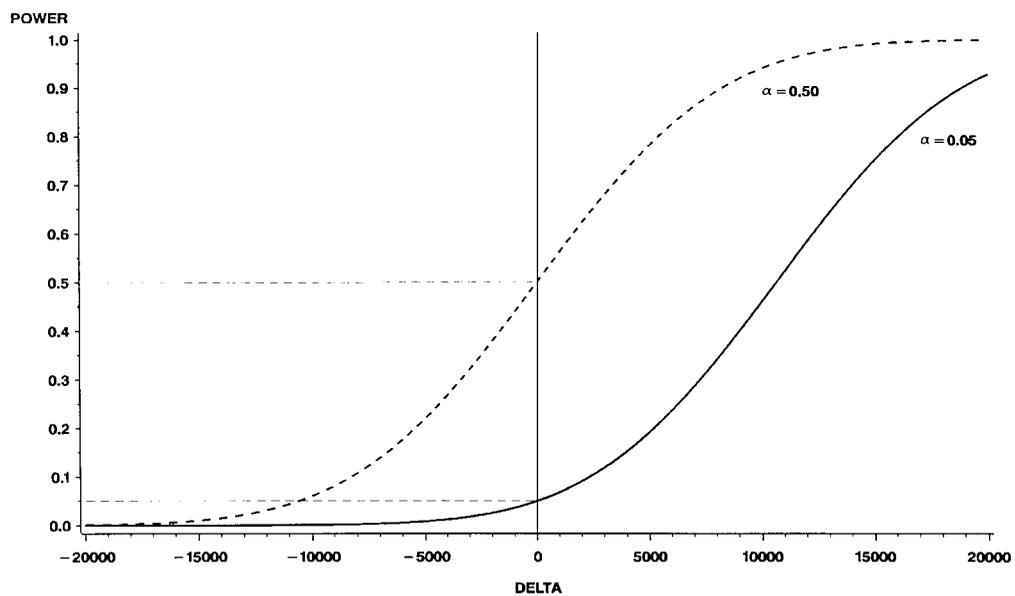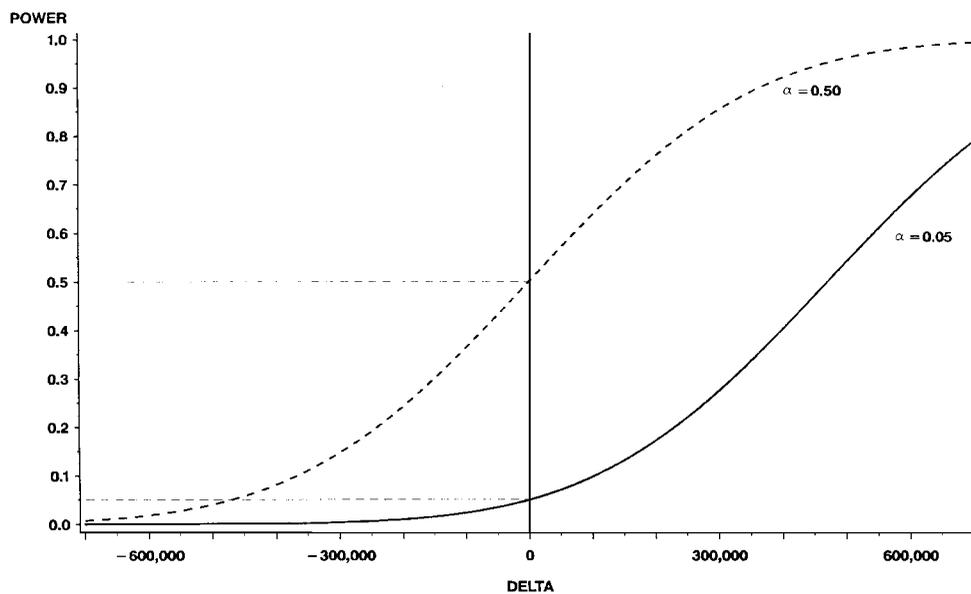
Figure 4. Power curves for the prostate example.

Figure 5. Power curves for the CIDS example.

5. DISCUSSION

In this paper we argue in favour of using INB methodology for analysing cost-effectiveness data where both cost and effectiveness are measured at the patient level. The main advantages are that the INB analysis (i) provides the cost minimization and ICER analyses as special cases and (ii) considers value as well as cost. The ICER is the cost of an extra unit of effectiveness, whereas INB is the difference between the value and the cost. There are other advantages as well.

The ICER analysis derived from INB methodology provides better insight than the Fieller's theorem approach. In the CIDS example the upper limit of the ICER is less than the lower limit. This appears confusing and may cause investigators mistakenly to reverse the limits and reach totally incorrect conclusions. The situation is seen more clearly using INB methodology. The lower limit of $b(\lambda)$ is always negative, implying that no value ascribed to a unit of effectiveness would lead to rejection of the null hypothesis that $b(\lambda) = 0$ in favour of $b(\lambda) > 0$. This means that no matter how much one values a year of life, say, there is no evidence that treatment is cost-effective compared to standard. Although the fact the ICER limits include the positive vertical axis amounts to the same thing, it is less obvious. For some examples the Fieller's theorem limits may not exist, and the investigators may not know what to conclude. For the INB analysis this situation is characterized by neither INB limit crossing the horizontal axis. Thus zero is always contained in the confidence interval, and no matter how much or how little one values a unit of effectiveness, there is no evidence that either standard or treatment is more cost-effective than the other.

Another advantage is the INB's ability to generalize to more than one measure of effectiveness. (In a trial of anticoagulants one might be interested in deaths, strokes and major bleeds.) If there are p measures of effectiveness, and letting $\mu^{(k)}$ be the difference of the expected value for the kth measure, then $b(\lambda) = \sum_{k=1}^{p} \mu^{(k)} \lambda_k - v$, where λ, is now the vector $(\lambda_1, \lambda_2, \ldots, \lambda_p)'$ of values ascribed to the various units of effectiveness. This results in a single measure of cost-effectiveness, whereas an ICER approach would have p measures – one for each measure of effectiveness. Patient preferences can be incorporated into INB. Suppose treatment was more invasive or less convenient than standard, and consequently, if they were equally effective, patients would prefer to receive standard. If w is some measure of a patient's willingness-to-pay to receive standard rather than treatment, assuming they were equally effective, then $b(\lambda) = \sum_{k=1}^{p} \mu^{(k)} \lambda_k - v - w$. If patients preferred treatment, then $-w$ is the willingness-to-pay to receive it rather than standard. Methods for estimating willingness-to-pay are discussed in the literature [22, 23].

INB's advantages notwithstanding, further work is required. The model needs to be modified to account for discounting for future benefits (effectiveness) and costs as is common with economic models. Additionally, statistical methods need to be extended to the case of censored quality-adjusted survival. Although there are some models for estimating the mean and variance for censored quality-adjusted survival [24], the covariance with the estimator of mean costs needs to be derived.

The issue of covariate adjustment when estimating expected effectiveness and costs needs to be addressed, particularly for non-randomized studies. For uncensored data, parametric regression models could be used for either effectiveness or costs. Parametric regression models could also be use for censored survival data. Lin [25] proposes linear regression model for covariate adjustment for censored cost data.

APPENDIX

A1. The variance estimator for $\hat{\mu}_j$

$$\hat{V}(\hat{\mu}_j) = \sum_{k\,:\,t_{1j}\,\leqslant\,t_{jk}\,<\,\tau} \left[\frac{A_j^2(t_{jk})\,d_{jk}}{n_{jk}(n_{jk} - d_{jk})} \right]$$

where

$$A_j(t_{jk}) = \int_{t_{jk}}^{\tau} \hat{S}_j(u)\,\mathrm{d}u = \sum_{l\,:\,t_{jk}\,\leqslant\,t_{jl}\,<\,\tau} \{\hat{S}_j(t_{jl})[\min(\tau, t_{j,\,l+1}) - t_{jl}]\}$$

A2. The estimators of the variance of $\hat{v}_j$ and the covariance between $\hat{\mu}_j$ and $\hat{v}_j$

The quantity e_{ji} are the patients' survival times, which may or may not be observed. Let U_{ji} be the censoring time for patient i on therapy j. Consider the following definitions:

$$x_{ji} = \min\{e_{ji}, U_{ji}\}$$

$$\delta_{ji} = I(e_{ji} < U_{ji}), \text{ where } I \text{ is the indicator function}$$

$$Y_{jki} = I[x_{ji} \geqslant \alpha_k \text{ and } (x_{ji} \geqslant \alpha_{k+1} \text{ or } \delta_{ji} = 1)]$$

$$R_{ji} = \sum_{l=1}^{n_j} I(x_{jl} \geqslant x_{ji})$$

$$c_{jki} \quad \text{is the cost for patient } i \text{ during interval } k$$

Then $\hat{V}(\hat{v}_j) = \sum_{i=1}^{n_j} \sum_{k=1}^{K} \sum_{l=1}^{K} W_{jki} W_{jli}$ where

$$W_{jki} = \frac{\hat{S}_j(\alpha_k) Y_{jki}(c_{jki} - c_{jk})}{m_{jk}} - \hat{S}_j(\alpha_k) c_{jk} \left(\frac{I(x_{ji} \leqslant \alpha_k)\delta_{ji}}{R_{ji}} - \sum_{l\,:\,x_{jl}\,\leqslant\,\min(\alpha_k, x_{ji})} \frac{\delta_{jl}}{R_{jl}^2} \right)$$

Recall that $m_{jk} = \sum_{i=1}^{n_j} Y_{jki}$ and $c_{jk} = \frac{1}{m_{jk}} \sum_{i\,:\,Y_{jki}\,=\,1} c_{jki}$.

Let $\hat{C}(\hat{\mu}_j, \hat{v}_j)$ denote the estimator of the covariance between $\hat{\mu}_j$ and $\hat{v}_j$. Then

$$\hat{C}(\hat{\mu}_j, \hat{v}_j) = -\sum_{i=1}^{n_j} \left(Q_{ji} \sum_{k=1}^{K} W_{jki} \right)$$

where

$$Q_{ji} = \frac{I(x_{ji} \leqslant \tau)\delta_{ji} A_j(x_{ji})}{R_{ji}} - \sum_{l=1}^{n_j} \frac{I(x_{jl} \leqslant x_{ji}) I(x_{jl} \leqslant \tau)\delta_{jl} A_j(x_{jl})}{R_{jl}^2}$$

and

$$A_j(t) = \int_t^{\tau} \hat{S}_j(u)\mathrm{d}u$$

A3. The derivation of $\hat{C}(\hat{\mu}_j, \hat{v}_j)$

From Lin *et al.* [17]

$$n_j^{1/2}(\hat{v}_j - v_j) = n_j^{-1/2} \sum_{i=1}^{n_j} \sum_{k=1}^{K} \xi_{jki} + o_p(1) \tag{A1}$$

where

$$\xi_{jki} = \frac{S_j(\alpha_k) Y_{jki}[c_{jki} - E(c_{jki}|Y_{jki}=1)]}{E(Y_{jki})} - E(c_{jki}|Y_{jki}=1) S_j(\alpha_k) \int_0^{\alpha_k} \frac{\mathrm{d}M_{ji}(t)}{\mathrm{Pr}(x_{ji} \geqslant t)}$$

where $M_{ji}(t) = \delta_{ji} I(x_{ji} \leqslant t) - \int_0^t I(x_{ji} \geqslant t) \, \mathrm{d}\Lambda_j(t)$ and $\Lambda_j(t)$ is the cumulative hazard function

$$\int_0^t \frac{\mathrm{d} \log S_j(u)}{\mathrm{d}u} \mathrm{d}u$$

By the arguments of Gill [26]

$$n_j^{1/2}(\hat{\mu}_j - \hat{\mu}_j) = - n_j^{-1/2} \sum_{i=1}^{n_j} \int_0^\tau \frac{a_j(t)}{\mathrm{Pr}(e_{ij} \geqslant t)} \, \mathrm{d}M_{ji}(t) + o_p(1) \tag{A2}$$

where $a_j(t) = \int_t^\tau S_j(u) \, \mathrm{d}u$ and $S_j(t) = \mathrm{Pr}(e_{ji} \geqslant t)$.

Because the right-hand side of both equations (A1) and (A2) are normalized sums of n_j i.i.d. random variables, if follows from the multivariate central limit theorem that $n_j^{1/2}(\hat{\mu}_j - \mu_j, \hat{v}_j - v_j)$ is asymptotically zero-mean bivariate normal with covariance equal

$$-E \left\{ \left[\int_0^\tau \frac{a_j(t)}{\mathrm{Pr}(e_{ji} \geqslant t)} M_{ji}(t) \right] \sum_{k=1}^{K} \xi_{jki} \right\}$$

By replacing all the unknown parameters by the respective sample estimators, we obtain $\hat{C}(\hat{\mu}_j, \hat{v}_j)$.

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. Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Economics* 1996; **5**:297–305.
3. Chaudhary MA, Stearns SC. Confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statistics in Medicine* 1996; **15**:1447–1458.
4. Mullahy J, Manning W. Statistical issues of cost-effectiveness analysis. In *Valuing Health Care*, Sloan F (ed.). Cambridge University Press: Cambridge, 1994; 149–184.
5. 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.
6. Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost effectiveness analysis. In *Cost Effectiveness in Health and Medicine*, Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Oxford University Press: New York, 1996.

 7. 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.
 8. 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.
 9. Wakker P, Klaassen MP. Confidence intervals for cost/effectiveness ratios. *Health Economics* 1995; **4**:373–381.
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. *International Journal of Technology 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, Mallahy J. Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998; **18** (Suppl):S68–S80.
14. Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment* 1998; **14**:467–471.
15. Coldman AJ, Elwood MJ. Examining survival data. *Canadian Medical Journal* 1979; **121**:1065–1071.
16. Collet D. *Modeling Survival Data in Medical Research*. Chapman & Hall: London, 1997.
17. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**:419–434.
18. Tannock IF, Osoba D, Stockler MR, Ernst DS, Neville AJ, Moore MJ, Armitage GR, Wilson JJ, Venner PM, Coppin CM, Murphy KC. Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: A Canadian trial with palliative endpoints. *Journal of Clinical Oncology* 1996; **14**:1756–1764.
19. Bloomfield DJ, Krahn MD, Neogi T, Panzarella T, Warde P, Willan AR, Ernst S, Moore MJ, Neville A, Tannock IF. Economic evaluation of chemotherapy with mitoxantrone plus prednisone for symptomatic hormone-resistant prostate cancer: Based on a Canadian trial with palliative endpoints. *Journal of Clinical Oncology* 1998; **16**:2272–2279 (1998).
20. Laska EM, Meisner M, Siegel C. Power and sample in cost-effectiveness analysis. *Medical Decision Making* 1999; **19**:339–343.
21. Willan AR. Power function arguments in support of an alternative approach for analyzing management trials. *Controlled Clinical Trials* 1994; **15**:211–219.
22. O'Brien BJ. Willingness-to-pay in cost-benefit analysis of health care programs. In *Encyclopedia of Biostatistics*, *Volume 6*, Colton T, Armitage P (eds). Wiley: Chichester, 1998; 4750–4755.
23. O'Brien BJ, Gafni A. When do the 'dollars' make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making* 1996; **16**:288–299.
24. Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* 1997; **84**:339–348.
25. Lin DY. Linear regression analysis of censored medical costs. *Biostatistics* 2000; **1**:35–47.
26. Gill RD. Large sample behavior of the product-limit estimator on the whole line. *Annals of Statistics* 1983; **11**:49–58.