Statistical Methods in Medical Research 2002; **11**: 539–551

Using inverse-weighting in cost-effectiveness analysis with censored data

AR Willan Program in Population Health Sciences, Research Centre, Hospital for Sick Children, Toronto, Canada, **DY Lin** Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA and **RJ Cook** and **EB Chen** Department of Statistics & Actuarial Science, University of Waterloo, Waterloo, Canada

Due to induced dependent censoring, estimating mean costs and quality-adjusted survival in a cost-effectiveness analysis using standard life-table methods leads to biased results. In this paper we propose methods for estimating the difference in mean costs and the difference in effectiveness, together with their respective variances and covariance in the presence of dependent censoring. We consider the situation in which the measure of effectiveness is either the probability of patients surviving a duration of interest, mean survival time over a duration of interest or mean quality-adjusted survival time over a duration of interest. The method of inverse-weighting is used for censored cost and quality of life data. The methods are illustrated in an example using an incremental net benefit analysis.

1 Introduction

It is becoming increasingly common for researchers to collect patient-level cost data in addition to effectiveness outcomes in randomized clinical trials. As a result there have been many publications[1–28] regarding the development of statistical methods for the design and analysis of cost-effectiveness studies. Initially, authors concentrated on providing inference on incremental cost-effectiveness ratios (ICER), but emphasis has recently shifted to incremental net benefit (INB),[21–28] of which ICER can be thought of as a special case.[25–28] To conduct a cost-effectiveness analysis, using either an ICER or an INB approach, one need only estimate the difference in effectiveness ($\hat{\Delta}_e$) and costs ($\hat{\Delta}_c$), along with their respective variances and covariance ($\hat{V}(\hat{\Delta}_e)$, $\hat{V}(\hat{\Delta}_c)$, $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$). If there is no censoring (i.e., all patients are followed for the duration of interest), sample means, variances and covariances are all that are required.[26,27] However, for censored data, calculations become considerably more complex.[27–37] Since patients accumulate cost at different rates, there is a positive correlation between the amounts accumulated to death and censoring. Because of the positive correlation, the application of standard survival methods would provide an over-estimate of mean cost.[27–32] The same issue exits for estimating mean quality-adjusted survival time.[28,34–40]

Based on what is known as the direct method (Lin *et al.*[29]), Willan and Lin[27] describe an approach for inference about INB in the presence of censoring when the measure of

Address for correspondence: Andrew R Willan, Hospital for Sick Children, 555 University Avenue, Toronto, ON, M5G 1X8, Canada. E-mail: andrew.willan@sickkids.ca

 10.1191/0962280202sm308ra

effectiveness is mean survival time. Willan *et al.*[28] extend these methods to provide solutions for inference when the measure of effectiveness is either the probability of surviving or the mean quality-adjusted survival time. Recently,[30–37] other methods have been proposed to account for censoring. Based on a method know as inverse-weighting, Zhao and Tsiatis[34–36] provide estimates for $\hat{\Delta}_e$ and $V(\hat{\Delta}_e)$, Bang and Tsiatis[31] and Zhao and Tian[33] for $\hat{\Delta}_c$ and $V(\hat{\Delta}_c)$, and Lin[30] for $\hat{\Delta}_c$ and $V(\hat{\Delta}_c)$ with covariate adjustment. Simulation results[31,32,34] indicate that the direct and inverse-weighting methods have similar bias and variance. The inverse-weighting methods provide consistent estimators for general censoring patterns, whereas the direct method requires that censoring occur only at a limited number of time points. In this paper we use the method of inverse-weighting to provide specific solutions for estimating Δ_e, Δ_c, $V(\hat{\Delta}_e)$, $V(\hat{\Delta}_c)$ and $C(\hat{\Delta}_e, \hat{\Delta}_c)$ for the situation in which censoring is present and the measure of effectiveness is either (i) the probability of surviving, (ii) mean survival time, or (iii) mean quality-adjusted survival time. The statistical model and methods are given in Section 2, followed by illustrative applications in Section 3.

2 Methods

2.1 Cost-effectiveness analysis

Consider a trial designed to compare the cost-effectiveness of a new treatment, denoted by T, with a standard treatment, denoted by S. Patient-level cost data are collected along with some measure of effectiveness. In this paper the three measures of effectiveness considered are (i) the probability of surviving the duration of interest, (ii) the mean survival time over the duration of interest, and (iii) the mean quality-adjusted survival time over the duration of interest. If the analysis of the data is performed before all patients have been followed for the duration of interest, or if some patients are lost to follow-up during the intended period of observation, then right-censoring is said to occur. Consequently, cost and the measure of effectiveness are not observed for some patients. Standard survival methods can be used in the presence of censoring to provide valid estimates for the probability of surviving and mean survival time. However, more complex methods are required for the estimation of mean costs and mean quality-adjusted survival time.[27–34]

An assessment of cost-effectiveness in a clinical trial can be expressed as the incremental cost-effectiveness ratio (ICER), which is defined as Δ_c/Δ_e, and measures the cost of achieving an extra unit of effectiveness from using T rather than S. For the probability of surviving, the ICER is a measure of the cost of preventing one death over the duration of interest and for mean survival time it is the cost per year of life gained. For mean quality-adjusted survival time, it is the cost per quality-adjusted year (QALY) gained. The ICER is estimated by $\hat{\Delta}_c/\hat{\Delta}_e$, and the $100(1-\alpha)\%$ Fieller[5,6] confidence limits are given by:

$$\left(\frac{\hat{\Delta}_c}{\hat{\Delta}_e}\right)\left[\frac{1 - z^2_{1-(\alpha/2)}c \pm z_{1-(\alpha/2)}\sqrt{a + b - 2c - z^2_{1-(\alpha/2)}(ab - c^2)}}{1 - z^2_{1-(\alpha/2)}a}\right]$$

where $a = \hat{V}(\hat{\Delta}_e)/\hat{\Delta}_e^2$, $b = \hat{V}(\hat{\Delta}_c)/\hat{\Delta}_c^2$, $c = \hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)/(\hat{\Delta}_e\hat{\Delta}_c)$ and $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal random variable. If the willingness-to-pay for a unit of effectiveness, denoted as λ, is less than the lower limit, then there is evidence that T is not cost-effective. If, on the other hand, λ is greater than the upper limit, there is evidence that T is cost-effective.

Analytical and interpretability issues have been raised regarding the ICER.[27] The confidence interval can include undefined values or may even be completely undefined. Two totally opposite results can have the same ICER. Also, negative ICERs are difficult to interpret and are not properly ordered. In response to these issues more attention is being paid to the incremental net benefit (INB). The INB, given as a function of the willingness-to-pay, is defined as $b(\lambda) = \lambda\Delta_e - \Delta_c$ and is estimated by $\hat{b}(\lambda) = \lambda\hat{\Delta}_e - \hat{\Delta}_c$. T is said to be cost-effective if $b(\lambda)$ is greater than 0. 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 $H_0: b(\lambda) = 0$, versus the alternative hypothesis $H_1: b(\lambda) > 0$, can be rejected at the level α if $\hat{b}(\lambda)/\sqrt{[\hat{V}(\hat{b}(\lambda))]}$ exceeds $z_{1-\alpha}$, in which case there is evidence that T is cost-effective. In addition, the $100(1-\alpha)\%$ confidence limits for $b(\lambda)$ are given by $\hat{b}(\lambda) \pm z_{1-\alpha/2}\sqrt{[\hat{V}(\hat{b}(\lambda))]}$. Most often an INB analysis is performed as a function of λ and illustrated in a plot of $\hat{b}(\lambda)$ versus λ. The slope of the plot is $\hat{\Delta}_e$ and the vertical intercept is $-\hat{\Delta}_c$. The horizontal intercept of $\hat{b}(\lambda)$ is the estimate of the ICER, i.e., $\hat{b}(\hat{\Delta}_c/\hat{\Delta}_e) = 0$. Furthermore, the horizontal intercepts of the confidence limits for $b(\lambda)$ provide the Fieller confidence limits for the ICER.[25–28]

To carry out a statistical cost-effectiveness analysis the following five parameters must be estimated: $\Delta_e, \Delta_c, V(\hat{\Delta}_e), V(\hat{\Delta}_c), C(\hat{\Delta}_e, \hat{\Delta}_c)$. In the remainder of this section estimators for these parameters are derived for censored data for the situation in which the measure of effectiveness is either (i) the probability of surviving the duration of interest, (ii) the mean survival time over the duration of interest, and (iii) the mean quality-adjusted survival time over the duration of interest. These methods are illustrated with an example in Section 3.

2.2 Estimating mean cost

In a two-arm randomized control trial let C_{ji} be the total cost for the duration of interest for patient i on therapy j, where $j = T$ (treatment), S (standard); $i = 1, 2, \ldots, n_j$ = the number of patients randomized to therapy j. Let the expected value of $C_{ji} = v_j$ and let $\Delta_c = v_T - v_S$. For patient i on therapy j, let D_{ji} and U_{ji} be the times from randomization to death and censoring, respectively. Let $G_j(t) = \Pr(U_{ji} \geq t)$, $X_{ji} = \min(D_{ji}, U_{ji})$, $\delta_{ji} = I\{D_{ji} < U_{ji}\}$ and $\bar{\delta}_{ji} = 1 - \delta_{ji}$, where I is the indicator function. If $D_{ji} > U_{ji}$ (i.e., $\bar{\delta}_{ji} = 1$) then C_{ji} is not observed.

A method for estimating Δ_c is proposed by Bang and Tsiatis[31] and Lin.[30] In this method the duration of interest $(0, \tau]$ is divided into the K intervals $[a_k, a_{k+1})$, where

$0 = a_1 < a_2 < \cdots a_{k+1} = \tau$. Let C_{jki} be the observed cost for patient i in group j during interval k. Let $X^*_{jki} = \min(X_{ji}, a_{k+1})$. Then v_j is estimated by

$$\hat{v}_j = \sum_{k=1}^{K} \bar{C}_{jk}$$

where

$$\bar{C}_{jk} = \left(\sum_{i=1}^{n_j} \frac{Y_{jki}}{\hat{G}_j(X^*_{jki})} \right)^{-1} \sum_{i=1}^{n_j} \frac{Y_{jki} C_{jki}}{\hat{G}_j(X^*_{jki})}$$

where $\hat{G}_j(t)$ is the product-limit estimate of $G_j(t)$ and $Y_{jki} = \delta_{ji} + \bar{\delta}_{ji} I\{X_{ji} \geq a_{k+1}\}$. The quantity $\bar{C}_{jk}$ is a weighted average of the observed cost incurred in the kth interval (i.e., of the C_{ijk}s). The weight is zero if the patient is censored before or during the interval. Otherwise, the weight is one over the probability of not being censored at the end of the interval if the patient survives the interval or at the time of death, if not. Seen as a special case of the derivation given in Lin,[30] the estimator of the variance of $\hat{v}_j$ is given by

$$\hat{\mathrm{V}}(\hat{v}_j) = \sum_{i=1}^{n_j} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \hat{Z}^{(c)}_{jki} \hat{Z}^{(c)}_{j\ell i}$$

where

$$\hat{Z}^{(c)}_{jki} = \frac{1}{n_j} \left(\frac{Y_{jki}(C_{jki} - \bar{C}_{jk})}{\hat{G}_j(X^*_{jki})} + \bar{\delta}_{ji} B_{jki} - \sum_{\ell=1}^{n_j} \frac{\bar{\delta}_{j\ell} I\{X_{j\ell} \leq X_{ji}\} B_{jk\ell}}{R_{j\ell}} \right)$$

$$R_{ji} = \sum_{\ell=1}^{n_j} I\{X_{j\ell} \geq X_{ji}\}$$

and

$$B_{jki} = \frac{1}{R_{ji}} \sum_{\ell=1}^{n_j} \frac{I\{X^*_{jk\ell} > X_{ji}\} Y_{jk\ell} (C_{jk\ell} - \bar{C}_{jk})}{\hat{G}_j(X^*_{jk\ell})}$$

Thus Δ_c is estimated by $\hat{\Delta}_c = \hat{v}_T - \hat{v}_S$ and the estimated variance is given by $\hat{\mathrm{V}}(\hat{\Delta}_c) = \hat{\mathrm{V}}(\hat{v}_T) + \hat{\mathrm{V}}(\hat{v}_S)$.

2.3 Probability of surviving the duration of interest

If the measure of effectiveness is the probability of surviving the duration of interest, then the parameter of interest in the comparison of therapies is $\Delta_e = S_T(\tau) - S_S(\tau)$, where $S_j(t)$ is the survival curve defined as $S_j(t) = \Pr(D_{ji} > t)$. Let t_{jh}, where $t_{j,h-1} < t_{jh}$, $h = 1, 2, \ldots L_j$, be unique times to death for patients on therapy j. Then the product-

limit estimator of $S_j(\tau)$ is given by $\hat{S}_j(\tau) = \prod_{h/t_{jh}<\tau}(1 - d_{jh}/n_{jh})$, where d_{jh} are the number of deaths on therapy j at time t_{jh}, and n_{jh} is the number of patients still at risk at time t_{jh}. The estimator of the variance of $\hat{S}_j(\tau)$ is given by

$$\hat{\mathrm{V}}[\hat{S}_j(\tau)] = \sum_{i=1}^{n_j}(\hat{Z}_{ji}^{(p)})^2$$

where

$$\hat{Z}_{ji}^{(p)} = -\hat{S}_j(\tau)\left[\frac{I\{X_{ji} \leq \tau\}\delta_{ji}}{R_{ji}} - \sum_{\ell=1}^{n_j}\frac{I\{X_{j\ell} \leq \min(\tau, X_{ji})\}\delta_{j\ell}}{R_{j\ell}^2}\right]$$

(see Willan *et al.*[28]).

Thus the estimator of Δ_e is given by $\hat{\Delta}_e = \hat{S}_T(\tau) - \hat{S}_S(\tau)$ with estimated variance given by $\hat{\mathrm{V}}(\hat{\Delta}_e) = \hat{\mathrm{V}}(\hat{S}_T(\tau)) + \hat{\mathrm{V}}(\hat{S}_S(\tau))$. The estimated covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is determined by an identical derivation provided by Willan *et al.*,[28] and is given by

$$\begin{aligned}\hat{\mathrm{C}}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{\mathrm{C}}(\hat{S}_T(\tau), \hat{v}_T) + \hat{\mathrm{C}}(\hat{S}_S(\tau), \hat{v}_S)\\ &= \sum_{j=T,S}\sum_{i=1}^{n_j}\left(\hat{Z}_{ji}^{(p)}\sum_{k=1}^{K}\hat{Z}_{jki}^{(c)}\right)\end{aligned}$$

2.4 Mean survival time

Defining $t_{j0} = 0$, the mean survival time is estimated by the area under the product-limit estimate of the survival curve $S(t)$ from 0 to τ, and is given by

$$\int_0^\tau \hat{S}_j(t)\mathrm{d}t = \sum_{h/t_{jh}<\tau}[\hat{S}_j(t_{jh})(\min(t_{j,h+1}, \tau) - t_{j,h})]$$

The estimated variance is given by

$$\hat{\mathrm{V}}\left(\int_0^\tau \hat{S}_j(t)\mathrm{d}t\right) = \sum_{i=1}^{n_j}(\hat{Z}_{ji}^{(m)})^2$$

where

$$\hat{Z}_{ji}^{(m)} = -\left[\frac{I\{X_{ji} \leq \tau\}\delta_{ji}A_j(X_{ji})}{R_{ji}} - \sum_{\ell=1}^{n_j}\frac{I\{X_{j\ell} \leq \min(X_{ji}, \tau)\}\delta_{j\ell}A_j(X_{j\ell})}{R_{j\ell}^2}\right]$$

and

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

(see Willan *et al.*[28]). Thus the estimator of Δ_e is given by

$$\hat{\Delta}_e = \int_0^\tau \hat{S}_T(t)\mathrm{d}t - \int_o^\tau \hat{S}_S(t)\mathrm{d}t$$

with estimated variance given by

$$\hat{V}(\hat{\Delta}_e) = \hat{V}\left(\int_0^\tau \hat{S}_T(t)\mathrm{d}t\right) + \hat{V}\left(\int_o^\tau \hat{S}_S(t)\mathrm{d}t\right)$$

The estimated covariance between $\hat{\Delta}_e$ and $\hat{\Delta}_c$ is determined by an identical derivation provided by Willan *et al.*,[28] and is given by

$$\begin{aligned}\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) &= \hat{C}\left(\int_0^\tau \hat{S}_T(t)\mathrm{d}t, \hat{v}_T\right) + \hat{C}\left(\int_0^\tau \hat{S}_S(t)\mathrm{d}t, \hat{v}_S\right) \\ &= \sum_{j=T,S}\sum_{i=1}^{n_j}\left(\hat{Z}_{ji}^{(m)}\sum_{k=1}^{K}\hat{Z}_{jki}^{(c)}\right)\end{aligned}$$

2.5 Mean quality-adjusted survival time

The standard application of survival table methods to quality-adjusted survival leads to biased estimators of the mean quality-adjusted time.[28,30,38] Consequently, we propose using methods similar to those used in Section 2.1. Let q_{ji} be the quality-adjusted survival for the duration of interest for patient i on therapy j. Let the expected value of $q_{ji} = \mu_j$ and let $\Delta_e = \mu_T - \mu_S$. Recall that the duration of interest is divided into K intervals $[a_k, a_{k+1})$, where $0 = a_1 < a_2 \cdots a_{K+1} = \tau$. Let q_{jki} be the observed quality of life experienced by the patient i on therapy j during time interval k. The quantities q_{jki} can be determined in the following way. Suppose patient i on therapy j has m_{ji} quality of life measures taken at $t_{ji1}, t_{ji2}, \ldots, t_{jim_{ji}}$, with respective scores $Q_{ji1}, Q_{ji2}, \ldots, Q_{jim_{ji}}$. Then $q_{jki} = \int_{a_k}^{a_{k+1}} Q(t)\mathrm{d}t$, where

$$Q(t) = \begin{cases} Q_{ji1} & 0 \le t < t_{ji1} \\ Q_{jih} + \dfrac{(Q_{ji,h+1} - Q_{jih})(t - t_{jih})}{t_{ji,h+1} - t_{jih}} & t_{jih} \le t < t_{ji,h+1} \\ Q_{jim_{ji}} & t_{jim_{ji}} \le t < X_{ji} \\ 0 & t \ge X_{ji} \end{cases}$$

The estimator of μ_j is given by

$$\hat{\mu}_j = \sum_{k=1}^{K} \bar{q}_{jk}$$

where

$$\bar{q}_{jk} = \left(\sum_{i=1}^{n_j} \frac{Y_{jki}}{\hat{G}_j(X^*_{jki})} \right)^{-1} \sum_{i=1}^{n_j} \frac{Y_{jki} q_{jki}}{\hat{G}_j(X^*_{jki})}$$

The estimator of the variance of $\hat{\mu}_j$ is given by

$$\hat{\mathrm{V}}(\hat{\mu}_j) = \sum_{i=1}^{n_j} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \hat{Z}^{(q)}_{jki} \hat{Z}^{(q)}_{j\ell i}$$

where

$$\hat{Z}^{(q)}_{jki} = \frac{1}{n_j} \left(\frac{Y_{jki}(q_{jki} - \bar{q}_{jk})}{\hat{G}_j(X^*_{jki})} + \bar{\delta}_{ji} B_{jki} - \sum_{\ell=1}^{n_j} \frac{\bar{\delta}_{j\ell} I\{X_{j\ell} \leq X_{ji}\} B_{jk\ell}}{R_{j\ell}} \right)$$

and

$$B_{jki} = \frac{1}{R_{ji}} \sum_{\ell=1}^{n_j} \frac{I\{X^*_{jk\ell} > X_{ji}\} Y_{jk\ell}(q_{jk\ell} - \bar{q}_{jk})}{\hat{G}_j(X^*_{jk\ell})}$$

This is the same formulation as given in Section 2.2 for cost, with q_{jki} and $\bar{q}_{jki}$ substituted for C_{jki} and $\bar{C}_{jki}$, respectively.

Thus $\hat{\Delta}_{\mathrm{e}} = \hat{\mu}_{\mathrm{T}} - \hat{\mu}_{\mathrm{S}}$ and $\hat{\mathrm{V}}(\hat{\Delta}_{\mathrm{e}}) = \hat{\mathrm{V}}(\hat{\mu}_{\mathrm{T}}) + \hat{\mathrm{V}}(\hat{\mu}_{\mathrm{S}})$. The estimated covariance between $\hat{\Delta}_{\mathrm{e}}$ and $\hat{\Delta}_{\mathrm{c}}$ is determined by an identical derivation provided by Willan *et al.*,[28] and is given by

$$\hat{\mathrm{C}}(\hat{\Delta}_{\mathrm{e}}, \hat{\Delta}_{\mathrm{c}}) = \sum_{j=\mathrm{T,S}} \sum_{i=1}^{n_j} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \hat{Z}^{(q)}_{jki} \hat{Z}^{(c)}_{j\ell i}$$

3 Example

3.1 Probability of surviving the duration of interest

In a trial of patients at risk of cardiac arrest, 659 patients were randomized between amiodarone (S) and implantable cardioverter defibrillator (T). The duration of interest is 6.42 years (77 months), that is, $\tau = 6.42$. Cost data were collected prospectively on

Table 1 Sample sizes and parameter estimates with proportion surviving 77 months as the measure of effectiveness

	n_j	$\hat{S}_j(\tau)$	$\hat{v}_j$	$\hat{V}(\hat{S}_j(\tau))$	$\hat{V}(\hat{v}_j)$	$\hat{C}(\hat{S}_j(\tau), \hat{v}_j)$
Treatment	212	0.587	87072	0.00284	8500084	13.621
Standard	218	0.567	38825	0.00196	6497939	−5.142

$\hat{\Delta}_e = 0.0207$; $\hat{\Delta}_c = 48247$; $\hat{V}(\hat{\Delta}_e) = 0.0480$; $\hat{V}(\hat{\Delta}_c) = 14998022$; $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) = 8.479$.

430 patients from the centers participating in the economic analysis. Censoring is present since not all patients were followed for 6.42 years. Because the measure of effectiveness is the proportion surviving, the willingness-to-pay is expressed as CAD$ per life saved.

The relevant parameters, estimated using the procedures given in Sections 2.2 and 2.3, can be found in Table 1. Cost is given in CAD$. $\hat{b}(\lambda)$ and the corresponding 95% confidence limits are plotted as a function of λ in Figure 1. For $\lambda =$ CAD$100 000 per life saved, the estimate of INB is −CAD$46 177 with confidence interval −CAD$61 533 to −CAD$30 822. Because the confidence limits are both negative, the investigators can

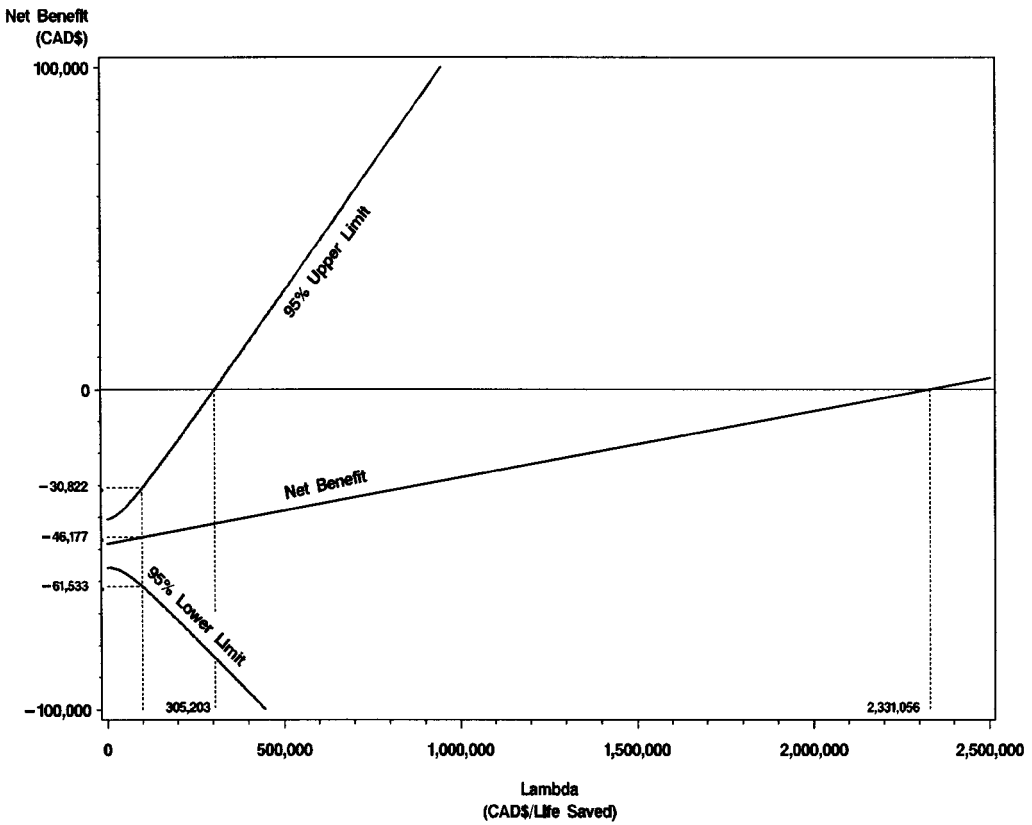

Figure 1 Incremental net benefit as a function of WTP for probability of surviving.

conclude that there is strong evidence that if the willingness-to-pay is CAD$100 000 per life saved, implantable cardioverter defibrillators are not cost-effective. The slope of $\hat{b}(\lambda)$ is positive (i.e., $\hat{\Delta}_e = 0.0207 > 0$), illustrating that Treatment was observed to increase the proportion surviving 77 months. The vertical intercept is negative (i.e., $\hat{\Delta}_c = 48\,247 > 0$), illustrating that Treatment was observed to increase cost. The graphs provide a confidence interval for Δ_c of CAD$40 657 to CAD$55 838 (the negative of the vertical intercepts). The estimate of the ICER, given by the horizontal intercept, is 2 331 056 CAD$/life saved. The Fieller lower confidence limit for the ICER, given by the horizontal intercept of the upper confidence limit for $b(\lambda)$, is CAD$305 203/life saved. 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.

3.2 Mean survival time

The relevant parameters, estimated using the procedures given in Sections 2.2 and 2.4, can be found in Table 2. Because the measure of effectiveness is mean survival, the willingness-to-pay is expressed as CAD$ year of life. $\hat{b}(\lambda)$ and the corresponding 95% confidence limits are plotted as a function of λ in Figure 2. For $\lambda =$ CAD$50 000 per year of life, the estimate of INB is −CAD$20 810 with confidence interval −CAD$40 741 to −CAD$879. Because the confidence limits are both negative, the investigators can conclude that there is strong evidence that if the willingness-to-pay is CAD$50 000 per year of life, implantable cardioverter defibrillators are not cost-effective. The slope of $\hat{b}(\lambda)$ is positive (i.e., $\hat{\Delta}_e = 0.549 > 0$), illustrating that Treatment was observed to increase mean survival over the 77 month duration of interest. The estimate of the ICER, given by the horizontal intercept, is CAD$87 923/year of life. The Fieller confidence interval for the ICER, defined by the horizontal intercept of the confidence limits for $\hat{b}(\lambda)$, is CAD$50 957 to 311 430/year of life.

3.3 Mean quality-adjusted survival time

Utility-based quality of life data were not collected in this trial. Therefore, for illustration we randomly generated the quality of life scores q_{jki} as $(a_{k+1} - a_k)\Phi(W_{jki})$, where W_{jki}, is a randomly generated normally distributed random variable with mean $(1 + 1.75 * I\{j = T\})$ and unit variance and Φ is the cumulative probability distribution for the standard normal random variable, i.e., $\Phi(z_{1-\alpha}) = 1 - \alpha$. Thereby, we have made treatment artificially superior to standard with respect to quality-adjusted survival.

The relevant parameters, estimated using the procedures given in Sections 2.2 and 2.5, can be found in Table 3. Because the measure of effectiveness is quality-adjusted

Table 2 Sample sizes and parameter estimates with mean survival as the measure of effectiveness

	n_j	$\int_0^\tau \hat{S}_j(t)dt$	$\hat{v}_j$	$\hat{V}(\int_0^\tau \hat{S}_j(t)dt)$	$\hat{V}(\hat{v}_j)$	$\hat{C}(\int_0^\tau \hat{S}_j(t)dt, \hat{v}_j)$
Treatment	212	4.832	87 072	0.02377	8 500 084	125.78
Standard	218	4.283	38 825	0.01737	6 497 939	18.79

$\hat{\Delta}_e = 0.549$; $\hat{\Delta}_c = 48\,247$; $\hat{V}(\hat{\Delta}_e) = 0.04114$; $\hat{V}(\hat{\Delta}_c) = 14\,998\,022$; $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) = 144.5$.

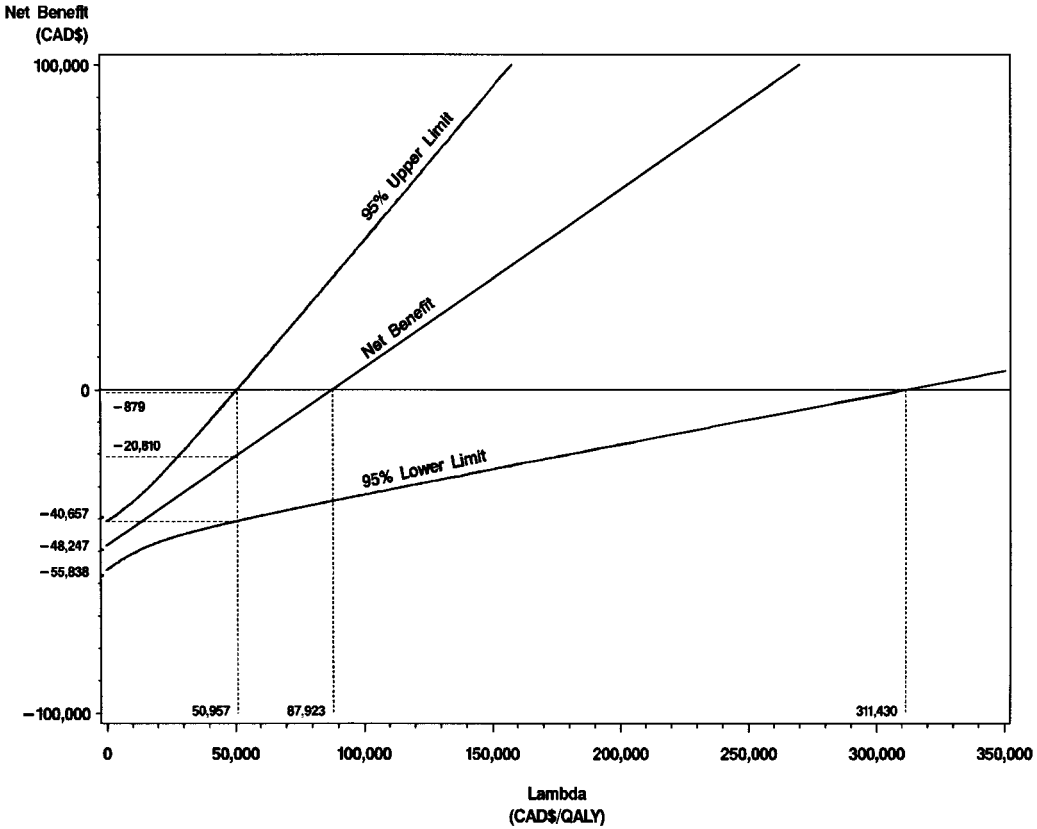

Figure 2 Incremental net benefit as a function of WTP for mean survival time.

survival, the willingness-to-pay is expressed as CAD$ per quality-adjusted life-year (QALY). $\hat{b}(\lambda)$ and the corresponding 95% confidence limits are plotted as a function of λ in Figure 3. For $\lambda =$ CAD$50 000 per QALY, the estimate of INB is CAD$10 040 with confidence interval −CAD$9366 to CAD$29445. Because the confidence interval includes 0, the investigators cannot reject the null hypothesis for a willingness-to-pay of CAD$50 000 per QALY, implying that there is not strong evidence to conclude that implantable cardioverter defibrillators are or are not cost-effective. The slope of $\hat{b}(\lambda)$ is positive (i.e., $\hat{\Delta}_e = 1.166 > 0$), illustrating that Treatment was observed to increase quality-adjusted survival over the 77 month duration of interest. The estimate of the

Table 3 Sample sizes and parameter estimates with quality-adjusted survival as the measure of effectiveness

	n_j	$\hat{\mu}_j$	$\hat{v}_j$	$\hat{V}(\hat{\mu}_j)$	$\hat{V}(\hat{v}_j)$	$\hat{C}(\hat{\mu}_j, \hat{v}_j)$
Treatment	212	4.865	87 072	0.02352	8 500 084	124.96
Standard	218	3.700	38 825	0.01501	6 497 939	8.136

$\hat{\Delta}_e = 1.166$; $\hat{\Delta}_c = 48 247$; $\hat{V}(\hat{\Delta}_e) = 0.0385$; $\hat{V}(\hat{\Delta}_c) = 14 998 022$; $\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c) = 133.09$.

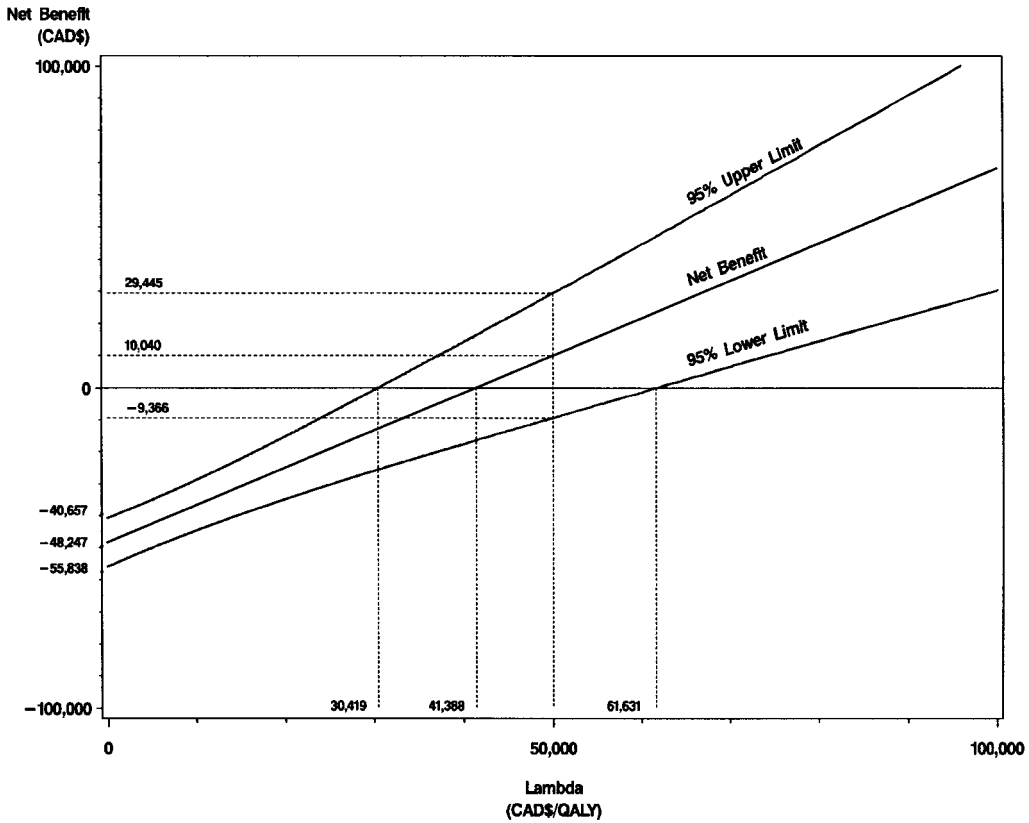

Figure 3 Incremental net benefit as a function of WTP for mean quality-adjusted survival.

ICER, given by the horizontal intercept, is CAD$41 388/QALY. The Fieller confidence interval for the ICER, defined by the horizontal intercept of the confidence limits for $\hat{b}(\lambda)$, is CAD$30 419 to 61 631/QALY.

4 Discussion

In this paper the method of inverse-weighting is proposed for estimating the parameters required for performing a cost-effectiveness comparison of two groups when the measure of effectiveness is some function of survival and censoring is present. The parameter estimates can be used for inference about the ICER or INB. Willan and Lin[27] and Willan *et al.*,[28] using the direct method, propose estimators for the parameters in which the mean cost is estimated by a weighted sum of the average cost in each of the intervals (defined by $[a_k, a_{k+1})$) of those patients who are alive at the beginning of the interval and are not censored before the end. The weight for each interval is the estimated survival function evaluated at the beginning of the interval. For consistency, the direct method requires that censoring occur at the beginning or the end of the

Table 4 Comparison of parameter estimates

	$\hat{\Delta}_e$	$\hat{V}(\hat{\Delta}_e)$	$\hat{\Delta}_c$	$\hat{V}(\hat{\Delta}_c)$	$\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$
Direct	1.166	0.0385	48 247	14 998 022	133.09
Inverse-weighting	1.167	0.0379	48 244	14 979 146	130.25

intervals, whereas inverse-weighting allows arbitrary censoring. As a comparison, the estimates from the example using both methods are contained in Table 4. Little difference is seen. The near equivalence of the two methods is discussed in more detail in Strawderman[37] and O'Hagan and Stevens.[41]

The formulation for INB used in this paper assumes that the value of λ, the willingness-to-pay (WTP), is insensitive to the value of Δ_e. In particular, when Δ_e is negative, λ becomes the willingness-to-accept (WTA) compensation for fore-going health benefit, and there exist evidence that the WTA is greater than WTP (see O'Brien *et al.*[42]). Willan *et al.*[43] provide a statistical analysis of INB for the situation when WTA > WTP.

Acknowledgements

The authors wish to thank Dr Gary Foster for producing the figures and Ms Susan Tomlinson for her meticulous proof reading.

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 health care. *Medical Care* 1994; **32**: 150–63.

2 Mullahy J, Manning W. Statistical issues of cost-effectiveness analysis. In Sloan F, ed. *Valuing Health Care.* Cambridge: Cambridge University Press, 1994.

3 van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E ratios alongside a clinical trial. *Health Economics* 1994; **3**: 309–19.

4 Wakker P, Klaassen MP. Confidence intervals for cost/effectiveness ratios. *Health Economics* 1995; **4**: 373–81.

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**: 1447–58.

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 eds. *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**: 327–40.

10 Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: A comparison of four methods. *Health Economics* 1997; **6**: 243–52.

11 Severens JL, De Boo TM, Konst EM. Uncertainty of incremental cost-effectiveness ratios. *International Journal of Technology Assessment in Health Care* 1999; **15**: 608–14.

12 Briggs AH, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics* 1998; **7**: 723–40.

13 Briggs AH, Gray AM. Power and sample size calculations for cost-effectiveness studies. *Medical Decision Making* 1998; **18** (Suppl): S81–S92.

14 Laska EM, Meisner M, Siegel C. Power and sample in cost-effectiveness analysis. *Medical Decision Making* 1999; **19**: 339–43.

15 Briggs AH, Mooney CZ, Wonderling DE. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Statistics in Medicine* 1999; **18**: 3245–62.

16 Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. *British Medical Journal* 1999; **319**: 635–38.

17 Heitjan DF, Moskowitz AJ, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics* 1999; **8**: 191–201.

18 Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics* 1999; **8**: 257–61.

19 Willan AR, O'Brien BJ. Sample size and power issues in estimating incremental cost-effectiveness ratios from clinical trials data. *Health Economics* 1999; **8**: 203–11.

20 Gardiner JC, Huebner M, Jetton J, Bradley CJ. Power and sample size assessments for test of hypotheses on cost-effectiveness ratios. *Health Economics* 2000; **9**: 227–34.

21 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.

22 Ament A, Baltussen R. The interpretation of results of economic evaluation: Explicating the value of health. *Health Economics* 1997; **6**: 625–35.

23 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.

24 Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment* 1998; **14**: 467–71.

25 Heithan DF. Fieller's method and net health benefit. *Health Economics* 2000; **9**: 327–35.

26 Willan AR. Analysis, sample size and power for estimating incremental net benefit from clinical trials data. *Controlled Clinical Trials* 2001; **22**: 228–37.

27 Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Statistics in Medicine* 2001; **20**: 1563–74.

28 Willan AR, Chen EB, Cook RI, Lin DY. Incremental net benefit in randomized clinical trials with quality-adjusted survival. *Statistics in Medicine*, accepted for publication.

29 Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**: 419–34.

30 Lin DY. Linear regression analysis of censored medical costs. *Biostatistics* 2000; **1**: 35–47.

31 Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika* 2000; **87**: 329–43.

32 Carides GW, Heyse JF, Iglewicz B. A regression-based method for estimating mean treatment cost in the presence of right-censoring. *Biostatistics* 2000; **1**: 229–313.

33 Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics* 2001; **57**: 1002–08.

34 Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* 1997; **84**: 339–48.

35 Zhao H, Tsiatis AA. Efficient estimation of the distribution of quality adjusted survival time. *Biometrics* 1999; **55**: 1101–07.

36 Zhao H, Tsiatis AA. Testing equality of survival functions of quality adjusted lifetime. *Biometrics* 2001; **57**: 861–67.

37 Strawderman RL. Estimating the mean of an increasing stochastic process at a censored stopping time. *JASA* 2000; **95**: 1192–1208.

38 Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Statistics in Medicine* 1990; **9**: 1259–76.

39 Hwang JS, Tsauo JY, Wang JD. Estimation of expected quality adjusted survival by cross-sectional survey. *Statistics in Medicine* 1996; **15**: 93–102.

40 Shen LZ, Pulkstenis E, Hoseyni M. Estimation of mean quality adjusted survival time. *Statistics in Medicine* 1999; **18**: l541–54.

41 O'Hagan A, Stevens JW. On estimators of medical costs with censored data. Research Report No. 511/01, Department of Probability and Statistics, University of Sheffield.

42 O'Brien BJ, Willan AR, Gersten K, Faulkner L. Is there a kink in the consumers' threshold value for cost-effectiveness in health care? *Health Economics* 2002; **11**: 175–180.

43 Willan AR, O'Brien BJ, Leyva RA. Cost-effectiveness analysis when the WTA is greater than the WTP. *Statistics in Medicine* 2001; **20**: 3251–59.