STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:131–145
Published online 28 October 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1794

Regression methods for cost-effectiveness analysis with censored data

Andrew R. Willan[1,2,*,†], D. Y. Lin[3] and Andrea Manca[4]

[1]*Program in Population Health Sciences, Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada M5G 1X8*
[2]*Department of Public Health Sciences, University of Toronto, Canada*
[3]*Department of Biostatistics, University of North Carolina, 3101E McGavran-Greenberg Hall, CB #7420 Chapel Hill, NC 27599-7420, U.S.A.*
[4]*Centre for Health Economics, University of York, Heslington, York, YO10 5DD, U.K.*

SUMMARY

A system of seemingly unrelated regression equations is proposed for prognostic factor adjustment and subgroup analysis when comparing two groups in a cost-effectiveness analysis with censored data. Because of the induced dependent censoring on costs and quality-adjusted survival, inverse probability weighting is employed for parameter estimation. The method is illustrated with data from two recent examples using both survival time and quality-adjusted survival time as the measures of effectiveness.

KEY WORDS: regression; censored data; cost-effectiveness; incremental net benefit; inverse probability weighting

1. INTRODUCTION

The increasingly common practice of collecting individual patient cost data in randomized clinical trials has motivated the development of new statistical methodology designed to answer questions regarding the cost-effectiveness of new therapies. Early development focussed on making inference related to the incremental cost-effectiveness ratio (ICER) [1–10]. More recently, due to the problems associated with ratio statistics, attention has shifted to incremental net benefit (INB) [11–23]. The analysis of INB requires the specification of the decision

*Correspondence to: Andrew R. Willan, Population Health Sciences, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada.
†E-mail: willana@mcmaster.ca

Contract/grant sponsor: National Science and Engineering Research Council of Canada; contract/grant number: 44868-03
Contract/grant sponsor: National Cancer Institute; contract/grant number: R01 CA082659

Received October 2003
Accepted November 2003

maker's willingness-to-pay (WTP) for an additional 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. Regardless of whether an ICER or an INB approach is taken, the between-treatment differences with respect to cost and effectiveness, together with their respective variances and covariance, must be estimated.

In this article, we use a system of seemingly unrelated regression equations to provide a general method for prognostic factor adjustment and subgroup analysis in cost-effectiveness studies with censored data. A similar approach was recently applied to uncensored data [24]. The method can be used in either an ICER approach or an INB approach, and does not require that the set of covariates for costs and effectiveness be the same. The use of covariate adjustment is essential for observational studies, since patient groups are likely to have baseline differences with respect to factors affecting both costs and clinical outcomes. In clinical trials, however, regression models are used most effectively for identifying treatment by prognostic factor interactions in the examination of subgroup effects. In the current paper, we propose to use the method of inverse probability weighting to provide regression models for covariate adjustment and subgroup analysis in cost-effectiveness analysis in the presence of censoring. The measure of effectiveness can be either survival time or quality-adjusted survival time. Inverse probability weighting is used because, even when censoring is completely at random, censoring on the cost scale and the quality-adjusted survival scale will be informative [21–23, 25–37]. There exists a positive correlation between a patients costs accumulated to death and his or her costs accumulated to censoring. This results because patients accrue costs at different rates. The same problem exists for quality-adjusted survival. Consequently, standard life-table methods applied to cost or quality-adjusted survival data will provide biased results.

The paper is structured as follows. The model and general methodology is outlined in Section 2. Two examples are given in Section 3. In the first the measure of effectiveness is survival time and focuses on the identification of interactions for subgroup analyses. In the second the measure of effectiveness is quality-adjusted survival time. A discussion follows in Section 4.

2. METHODS

2.1. Costs

Consider a two-arm randomized controlled trial in which patients are randomized into either a treatment group or a standard group. For patient i let D_i and U_i be the times from randomization to death and censoring, respectively. Let $G(t) = \Pr(U_i \geqslant t)$, $X_i = \min(D_i, U_i)$, $\delta_i = I\{D_i \leqslant U_i\}$ and $\bar{\delta}_i = 1 - \delta_i$, where $I\{\cdot\}$ is the indicator function. The function $G(t)$ is the probability of not being censored in the interval $[0, t)$. Let the time period of interest $(0, \tau]$ be divided into the K intervals $(a_k, a_{k+1}]$, $k = 1, 2, \ldots, K$, where $0 = a_1 < a_2 < \cdots < a_{K+1} = \tau$. Typically the intervals $(a_k, a_{k+1}]$ are the shortest intervals for which disaggregated cost data are available, and usually correspond to the intervals between data collection visits. If cost histories are not available, and cost data are aggregated for each patient over the entire duration on trial, then $K = 1$, $a_1 = 0$ and $a_{K+1} = \tau$. Note that $U_i > \tau$ for all patients followed alive for the period of interest. Let C_{ki} be the observed cost for patient i during interval k. Naturally,

$C_{ki}=0$ for all k where $a_k \geqslant D_i$. Let $X_{ki}^* = \min(X_i, a_{k+1})$. Let $C_i = \sum_{k=1}^{K} C_{ki}$ and consider the linear regression model $E(C_i) = \beta_C' Z_{Ci}$, $i=1,2,\ldots,n$, where E is the expectation function, Z_{Ci} is a vector of covariates whose effects on costs are of interest and β_C is a vector of unknown regression parameters. The first component of Z_{Ci} is set to 1 to provide an intercept, and the second component equals 1 if the ith patient is in the treatment group, 0 otherwise. Therefore, the second component of β_C is the mean difference in cost between randomization groups, adjusted for the other covariates, and is denoted by Δ_c.

From Lin [25], β_C can be estimated by $\hat{\beta}_C = \sum_{k=1}^{K} \hat{\beta}_{Ck}$, where

$$\hat{\beta}_{Ck} = \left(\sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{G}(X_{ki}^*)} Z_{Ci} Z_{Ci}' \right)^{-1} \sum_{i=1}^{n} \frac{\delta_{ki}^* C_{ki}}{\hat{G}(X_{ki}^*)} Z_{Ci}$$

$\hat{G}(t)$ is the product-limit estimator of $G(t)$, and $\delta_{ki}^* = \delta_i + \bar{\delta}_i I\{X_i \geqslant a_{k+1}\}$. Note that $\delta_{ki}^* = 1$ if the ith patient died or survived to at least a_{k+1}, and 0 otherwise. The second component of $\hat{\beta}_C$, denoted by $\hat{\Delta}_c$, is the estimator of the mean difference between randomization groups, adjusted for the other covariates. The variance–covariance matrix of $\hat{\beta}_C$ is estimated by

$$\hat{V}(\hat{\beta}_C) = n^{-1} \hat{A}_C^{-1} \hat{B}_C \hat{A}_C^{-1}$$

where

$$\hat{A}_C = n^{-1} \sum_{i=1}^{n} Z_{Ci} Z_{Ci}', \quad \hat{B}_C = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{m=1}^{K} \hat{\xi}_{Cki} \hat{\xi}_{Cmi}'$$

$$\hat{\xi}_{Cki} = \left(\frac{\delta_{ki}^* (C_{ki} - \hat{\beta}_{Ck}' Z_{Ci})}{\hat{G}(X_{ki}^*)} Z_{Ci} + \bar{\delta}_i F_{Cki} - \sum_{j=1}^{n} \frac{\bar{\delta}_j I\{X_j \leqslant X_i\}}{R_j} F_{Ckj} \right)$$

$$R_i = \sum_{j=1}^{n} I\{X_j \geqslant X_i\}$$

and

$$F_{Cki} = \frac{1}{R_i} \sum_{j=1}^{n} \frac{I\{X_{kj}^* > X_i\} \delta_{kj}^* (C_{kj} - \hat{\beta}_{Ck}' Z_{Cj})}{\hat{G}(X_{kj}^*)} Z_{Cj}$$

The element in the second row and second column of $\hat{V}(\hat{\beta}_C)$, denoted $\hat{\sigma}_{\Delta_c}^2$, is the estimator of the variance of $\hat{\Delta}_c$.

2.2. Quality-adjusted survival

Recall that the period of interest is divided into the K intervals $(a_k, a_{k+1}]$ $(k=1,2,\ldots,K)$, where $0 = a_1 < a_2 < \cdots < a_{K+1} = \tau$. Let q_{ki} be the observed quality of life experienced by the ith patient during time interval k. The quantities q_{ki} can be determined in the following way. Suppose patient i has m_i quality of life measures taken at $t_{i1}, t_{i2}, \ldots, t_{im_i}$, with respective scores

$Q_{i1}, Q_{i2}, \ldots, Q_{im_i}$. Then $q_{ki} = \int_{a_k}^{a_{k+1}} Q(t)\,\mathrm{d}t$, where

$$Q(t) = \begin{cases} Q_{i1} & 0 \leqslant t < t_{i1} \\ Q_{ih} + \dfrac{(Q_{i,h+1} - Q_{ih})(t - t_{ih})}{t_{i,h+1} - t_{ih}} & t_{ih} \leqslant t < t_{i,h+1} \\ Q_{im_i} & t_{im_i} \leqslant t < X_i \\ 0 & t \geqslant X_i \end{cases}$$

Let $q_i = \sum_{k=1}^{K} q_{ki}$ and consider the linear regression model $E(q_i) = \beta_E' Z_{Ei}$, $i = 1, 2, \ldots, n$, where Z_{Ei} is a vector of covariates whose effects on quality-adjusted survival times are of interest and β_E is a vector of unknown regression parameters. The covariates included in Z_{Ei} need not be the same as those included in Z_{Ci} from Section 2.1. Again, the first component of Z_{Ei} is set to 1 to provide an intercept, and the second component equals 1 if the patient is in the treatment group, 0 otherwise. Therefore, the second component of β_E is the mean difference in quality adjusted survival time between randomization groups, adjusted for the other covariates, and is denoted by Δ_e.

Using the same methodology from Lin [25], β_E can be estimated by $\hat{\beta}_E = \sum_{k=1}^{K} \hat{\beta}_{Ek}$, where

$$\hat{\beta}_{Ek} = \left(\sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{G}(X_{ki}^*)} Z_{Ei} Z_{Ei}' \right)^{-1} \sum_{i=1}^{n} \frac{\delta_{ki}^* q_{ki}}{\hat{G}(X_{ki}^*)} Z_{Ei}$$

The second component of $\hat{\beta}_E$, denoted by $\hat{\Delta}_e$, is the estimator of the mean difference between randomization groups, adjusted for the other covariates. The variance–covariance matrix of $\hat{\beta}_E$ is estimated by $\hat{V}(\hat{\beta}_E) = n^{-1} \hat{A}_E^{-1} \hat{B}_E \hat{A}_E^{-1}$, where

$$\hat{A}_E = n^{-1} \sum_{i=1}^{n} Z_{Ei} Z_{Ei}', \quad \hat{B}_E = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{m=1}^{K} \hat{\xi}_{Eki} \hat{\xi}_{Emi}'$$

$$\hat{\xi}_{Eki} = \left(\frac{\delta_{ki}^* (q_{ki} - \hat{\beta}_{Ek}' Z_{Ei})}{\hat{G}(X_{ki}^*)} Z_{Ei} + \bar{\delta}_i F_{Eki} - \sum_{j=1}^{n} \frac{\bar{\delta}_j I\{X_j \leqslant X_i\}}{R_j} F_{Ekj} \right)$$

and

$$F_{Eki} = \frac{1}{R_i} \sum_{j=1}^{n} \frac{I\{X_{kj}^* > X_i\} \delta_{kj}^* (q_{kj} - \hat{\beta}_{Ek}' Z_{Ej})}{\hat{G}(X_{kj}^*)} Z_{Ej}$$

The element in the second row and second column of $\hat{V}(\hat{\beta}_E)$, denoted by $\hat{\sigma}_{\Delta_e}^2$, is the estimator of the variance of $\hat{\Delta}_e$. The covariance matrix between $\hat{\beta}_C$ and $\hat{\beta}_E$ is estimated by $\hat{C}(\hat{\beta}_C, \hat{\beta}_E) = n^{-1} \hat{A}_C^{-1} \hat{B}_{CE} \hat{A}_E^{-1}$, where

$$\hat{B}_{CE} = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{m=1}^{K} \xi_{Cki} \xi_{Emi}'$$

The element in the second row and second column of $\hat{C}(\hat{\beta}_C, \hat{\beta}_E)$, denoted by $\hat{\sigma}_{\Delta_c \Delta_e}$, is the estimator of the covariance between $\hat{\Delta}_c$ and $\hat{\Delta}_e$.

2.3. Survival time

Let $D_i^* = \min(D_i, \tau)$. Consider the linear regression model $E(D_i^*) = \beta_E' Z_{Ei},\ i = 1, 2, \ldots, n$, where Z_{Ei} is a vector of covariates whose effects on survival times are of interest and β_E is a vector of unknown regression parameters. Again, the first component of Z_{Ei} is 1, and the second component is the treatment indicator. Therefore, the second component of β_E is the mean difference in mean survival between randomization groups, adjusted for the other covariates, and is denoted by Δ_e. Let $X_i^* = \min(X_i, \tau)$.

Using the same methodology from Lin [25], β_E can be estimated by

$$\hat{\beta}_E = \left(\sum_{i=1}^{n} \frac{\delta_i^*}{\hat{G}(X_i^*)} Z_{Ei} Z_{Ei}' \right)^{-1} \sum_{i=1}^{n} \frac{\delta_i^* X_i^*}{\hat{G}(X_i^*)} Z_{Ei}$$

where $\delta_i^* = \delta_i + \bar{\delta}_i I\{U_i \geqslant \tau\}$. Note that $\delta_i^* = 1$ if the ith patient died or survived the entire duration of interest, and 0 otherwise. The second component of $\hat{\beta}_E$, denoted by $\hat{\Delta}_e$, is the estimator of the mean differencec between randomization groups, adjusted for the other covariates. The variance–covariance matrix of $\hat{\beta}_E$ is estimated by

$\hat{V}(\hat{\beta}_E) = n^{-1} \hat{A}_E^{-1} \hat{B}_E \hat{A}_E^{-1}$, where

$$\hat{A}_E = n^{-1} \sum_{i=1}^{n} Z_{Ei} Z_{Ei}' \quad \text{and} \quad \hat{B}_E = n^{-1} \sum_{i=1}^{n} \hat{\xi}_{Ei} \hat{\xi}_{Ei}'$$

$$\xi_{Ei} = \left(\frac{\delta_i^* (X_i^* - \hat{\beta}_E' Z_{Ei})}{\hat{G}(X_i^*)} Z_{Ei} + \bar{\delta}_i F_{Ei} - \sum_{j=1}^{n} \frac{\bar{\delta}_j I\{X_j \leqslant X_i\}}{R_j} F_{Ej} \right)$$

and

$$F_{Ei} = \frac{1}{R_i} \sum_{j=1}^{n} \frac{I\{X_j^* > X_i\} \delta_j^* (X_j^* - \hat{\beta}_E' Z_{Ej})}{\hat{G}(X_j^*)} Z_{Ej}$$

The element in the second row and second column of $\hat{V}(\hat{\beta}_E)$, denoted $\hat{\sigma}_{\Delta_e}^2$, is the estimator of the variance of $\hat{\Delta}_e$. The covariance matrix between $\hat{\beta}_C$ and $\hat{\beta}_E$ is estimated by $\hat{C}(\hat{\beta}_C, \hat{\beta}_E) = n^{-1} \hat{A}_C^{-1} \hat{B}_{CE} \hat{A}_E^{-1}$, where

$$\hat{B}_{CE} = n^{-1} \sum_{i=1}^{n} \left(\sum_{k=1}^{K} \xi_{Cki} \right) \xi_{Ei}'$$

The element in the second row and second column of $\hat{C}(\hat{\beta}_C, \hat{\beta}_E)$, denoted $\hat{\sigma}_{\Delta_c \Delta_e}$, is the estimator of the covariance between $\hat{\Delta}_c$ and $\hat{\Delta}_e$.

2.4. Incremental cost-effectiveness ratio

The adjusted ICER is estimated by $\hat{\Delta}_c/\hat{\Delta}_e$, with the Fieller [5, 6] solution for the $100(1-\alpha)$ per cent confidence limits given by

$$(\hat{\Delta}_c/\hat{\Delta}_e)((1-z^2_{1-\alpha/2}a_{ec} \pm z_{1-\alpha/2}\sqrt{a_e+a_c-2a_{ec}-z^2_{1-\alpha/2}(a_e a_c-a^2_{ec})})/(1-z^2_{1-\alpha/2}a_e))$$

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

2.5. Incremental net benefit

Letting λ denote the willingness-to-pay for a unit of effectiveness, the adjusted INB(λ) is estimated by $b_\lambda=\lambda\hat{\Delta}_e-\hat{\Delta}_c$, with variance $\hat{\sigma}^2_\lambda=\lambda^2\hat{\sigma}^2_{\Delta_e}+\hat{\sigma}^2_{\Delta_c}-2\lambda\hat{\sigma}_{\Delta_c\Delta_e}$. The $100(1-2\alpha)$% confidence limits for INB(λ) are given by $b_\lambda \pm z_{1-\alpha}\hat{\sigma}_\lambda$, and if $b_\lambda/\hat{\sigma}_\lambda$ exceeds $z_{1-\alpha}$, the hypothesis $\mathrm{INB}(\lambda)\leqslant 0$ can be rejected in favour of $\mathrm{INB}(\lambda)>0$, at the level α, leading to the conclusion that treatment is cost-effective compared to standard.

2.6. Subgroup analysis

The regression model allows for sub-group analysis. Suppose investigators are interested in determining whether a certain sub-group, say males for sake of argument, have the same incremental net benefit as females. To accomplish this let $Z_{Ci}=Z_{Ei}=(1,z_{i1},z_{i2},z_{i3})'$, where z_{i1} is the treatment variable (1 if patient i is randomized to treatment, 0 otherwise), z_{i2} is 1 if the ith patient is male and 0 if female, and z_{i3} equal $z_{i1}\times z_{i2}$. The covariate z_{i3} is the interaction between sex and treatment. The parameter of interest is $\lambda\langle\beta_E\rangle_4-\langle\beta_C\rangle_4$, where $\langle b\rangle_j$ is the jth component of the vector b. If $\lambda\langle\hat{\beta}_E\rangle_4-\langle\hat{\beta}_C\rangle_4$ is positive, then males (since they are coded 1) are observed to have greater INB(λ). The opposite is true if $\lambda\langle\hat{\beta}_E\rangle_4-\langle\hat{\beta}_C\rangle_4$ is negative. The hypothesis that $\mathrm{INB}_{\mathrm{males}}(\lambda)=\mathrm{INB}_{\mathrm{females}}(\lambda)$ can be rejected at the level α, if

$$\frac{|\lambda\langle\hat{\beta}_E\rangle_4-\langle\hat{\beta}_C\rangle_4|}{\sqrt{\lambda^2\langle\hat{V}(\hat{\beta}_E)\rangle_{4,4}+\langle\hat{V}(\hat{\beta}_C)\rangle_{4,4}-2\lambda\langle\hat{C}(\hat{\beta}_C,\hat{\beta}_E)\rangle_{4,4}}}$$

exceeds $z_{1-\alpha/2}$, where $\langle\beta\rangle_{i,j}$ is the (i,j)th element of matrix B. When testing for interactions, both terms should be included. Observing that $\langle\hat{\beta}_C\rangle_4$, $\langle\hat{\beta}_E\rangle_4$ or both are not statistically significant does not mean that they are known to be zero with 100 per cent certainty, and the additional test of the hypothesis $\lambda\langle\beta_E\rangle_4-\langle\beta_C\rangle_4=0$ is required to determine if there is a significant interaction with respect to INB between the variable in question and randomization group.

3. EXAMPLES

3.1. The Canadian implantable defibrillator study (CIDS)

In a trial of patients at risk of cardiac arrest, a total of 659 patients with resuscitated ventricular fibrillation or sustained ventricular tachycardia or with unmonitored syncope were randomized

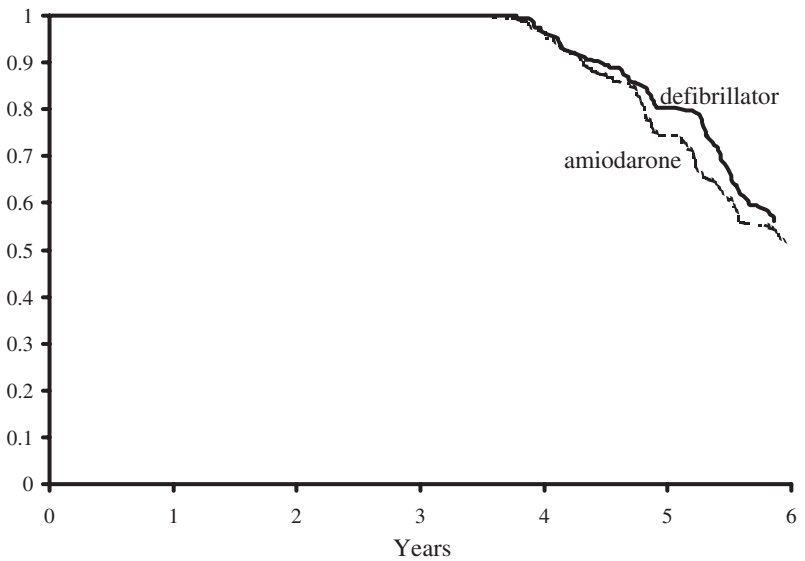

Figure 1. Censoring-free survival curves for CIDS example.

between amiodarone and implantable cardioverter defibrillator during the period from October 1990 to January 1997. Due to budgetary constraints, the costs were collected on the first 430 patients only. The analysis illustrated below includes only those 430 patients. The primary outcome measure was all-cause mortality. The clinical results are reported in Reference [38], and the economic evaluation in Reference [39]. A non-significant reduction in the risk of death was observed in the defibrillator arm, from 10.2 per cent per year to 8.3 per cent per year (19.7 per cent relative risk reduction; 95 per cent confidence interval, -7.7 to 40 per cent; $P = 0.142$). For the economic results reported below, the duration of interest is 6.5 years (78 months), i.e. $\tau = 6.5$. Censoring is present since not all patients were followed for 78 months. The censoring-free survival curves are shown by treatment arm in Figure 1. The censoring pattern is essentially the same in both arms and begins around year 4, reflecting the administrative censoring done at the time of analysis.

The covariates left ventricular ejection fraction (F) and sex (X) were considered. Ejection fraction was dichotomized into greater than 35 per cent ($F = 0$) and equal to or less than 35 per cent ($F = 1$). The covariate X was coded 1 for males and 0 for females. The treatment variable is coded 1 for defibrillator and 0 for amiodarone. The number and proportion of males and those whose ejection fraction was equal to or less than 35 per cent is shown in Table I. For any particular regression model, we used only those observations for which there were no missing data for the covariates included. Using survival time as the measure of effectiveness, the parameter estimates, standard errors and p-values are displayed in Table II, for the model with no covariates, the model with each covariate alone and the model with both covariates. Costs are given in Canadian dollars.

Ejection fraction was statistically significant for both cost and effectiveness, whether or not sex was included in the model. Sex was not significant for cost or effectiveness, regardless

Table I. Covariates by treatment, CIDS example.

		Treatment	
		Defibrillator ($N=212$)	Amiodarone ($N=218$)
Ejection fraction* $\leqslant 35$ per cent	n	132	147
	per cent	64.7	68.7
Males	n	184	174
	per cent	86.8	79.8

*12 missing values.

Table II. Parameter estimates, standard errors and p-values, CIDS example.

		Model			
Parameter		No covariates	Ejection fraction alone	Sex alone	Ejection fraction and sex
$\hat{\Delta}_c$	Estimate	48 668	49 666	48 488	49 586
	SE	3956	3997	4083	4120
	p-value	<0.0001	<0.0001	<0.0001	<0.0001
$\hat{\beta}_{C,F}$	Estimate		9012		8857
	SE	n.i.m.	3871	n.i.m.	3901
	p-value		0.0199		0.0232
$\hat{\beta}_{C,X}$	Estimate			2876	1579
	SE	n.i.m.	n.i.m.	5530	5579
	p-value			0.6029	0.7771
$\hat{\Delta}_e$	Estimate	0.4536	0.2958	0.4493	0.2912
	SE	0.4131	0.4030	0.4197	0.4012
	p-value	0.2722	0.4629	0.2844	0.4683
$\hat{\beta}_{E,F}$	Estimate		-0.8974		-0.9079
	SE	n.i.m.	0.3891	n.i.m.	0.3885
	p-value		0.0211		0.0195
$\hat{\beta}_{E,X}$	Estimate			0.3834	0.3999
	SE	n.i.m.	n.i.m.	0.5607	0.5075
	p-value			0.494	0.4307
$\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$		359.3	365.1	366.4	370.7

n.i.m., not in model.

of whether ejection fraction was included in the model. The signs for the ejection fraction coefficients were positive for cost and negative for effectiveness, indicating that patients with ejection fraction equal to or less than 35 per cent cost on average \$9012 more and lived an average of 0.8974 years less. Because of this and the fact that there were slightly more

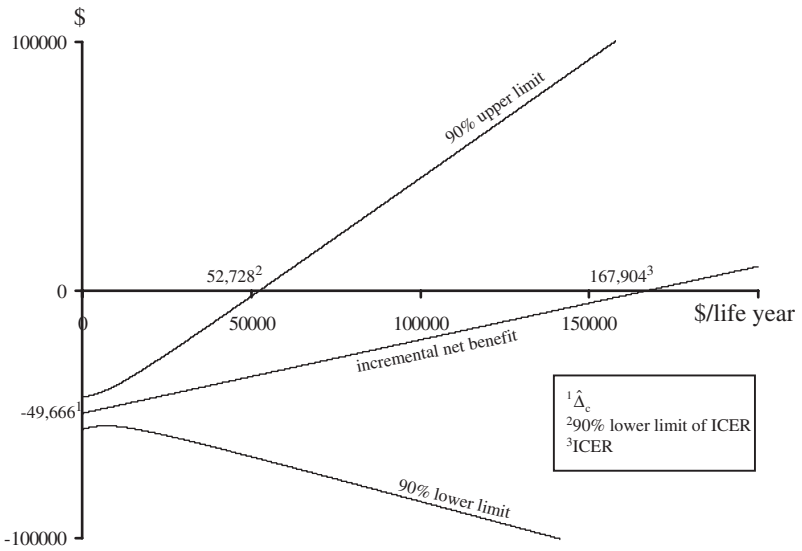

Figure 2. Plot of INB versus lambda for CIDS example including ejection fraction alone.

patients in the amiodarone arm whose ejection fraction was equal to or less than 35 per cent, when ejection fraction is included in the model, the treatment effect increases for cost and decreases for effectiveness. Therefore, the unadjusted ICER is \$107 293/life-year which is less than the adjusted ICER equal to \$167 904/life-year. Ignoring the possibility of an interaction between treatment and ejection fraction for the moment, one might choose to use the parameter estimates from the model with ejection fraction alone in the cost-effectiveness analysis. The adjusted 90 per cent Fieller [5, 6] confidence limits for the ICER are 52 728 and $-130\,609$. Because the lower limit lies in the North–East quadrant and the upper limit lies in the North–West quadrant of the cost-effectiveness plane, the confidence interval includes undefined values (the positive vertical axis), indicating that, because $\hat{\Delta}_e$ is 'statistically' close to zero, there is no upper limit to the ICER, and the proper representation of the confidence interval is $[52,728, \infty)$.

The plot of the adjusted estimate of INB and its 90 per cent confidence limits by λ can be found in Figure 2. Minus $\hat{\Delta}_c$ and the corresponding 90 per cent confidence limits are given where the plots meet the vertical axis. The horizontal intercepts provide the estimate of the ICER and its Fieller confidence limits. The 90 per cent lower limit of the INB never crosses the horizontal axis, indicating that no value of the willingness-to-pay would lead us to reject the hypothesis $\text{INB}(\lambda) \leqslant 0$ in favour of $\text{INB}(\lambda) > 0$, at the 5 per cent level. For the same reason the 90 per cent confidence interval for the ICER includes undefined values.

A model was fit which included treatment, ejection fraction and their interaction (not shown in table). The estimated interaction parameter associated with cost was -1772 ($p = 0.818$) and with effectiveness was $0.2851 (p = 0.711)$. The negative estimate for cost implies that the increase in cost due to the defibrillator was less for patients with a low ejection fraction. The positive estimate for effectiveness implies that the increase in effectiveness due to the defibrillator was greater for patients with a low ejection fraction. Neither interaction term is

Table III. INB interaction between treatment and ejection fraction, CIDS example.

λ	$\hat{\beta}_{E,T\times F}\lambda - \hat{\beta}_{C,T\times F}$	S.E.	p-value
50 000	16 028	37 254	0.667
125 000	37 411	94 578	0.692
250 000	73 050	190 730	0.702
500 000	144 328	383 257	0.706
100 000	286 885	768 421	0.709

Table IV. Treatment by type of hysterectomy, EVALUATE example.

	Treatment	
	Laparoscopic	Standard
Vaginal ($N = 487$)	324	163
Abdominal ($N = 859$)	573	286

significant, however this is not sufficient to conclude that the interaction on the INB scale is non-significant for all values of λ of interest, although in this particular case since the p-values are quite high, it is unlikely that there exist any value of λ for which the interaction is significant on the INB scale. Thus we tested the hypothesis $H : \beta_{E,T\times F}\lambda - \beta_{C,T\times F} = 0$, where $T \times F$ indicates the treatment by ejection fraction interaction. The estimate $\hat{\beta}_{E,T\times F}\lambda - \hat{\beta}_{C,T\times F}$, its standard error and associated p-value are given in Table III for several values of λ. The estimate will always be positive for positive values of λ, but for no values in this range was the interaction significant.

3.2. *The EVALUATE trial*

The EVALUATE trial is a randomized comparison of laparoscopic-assisted hysterectomy versus standard hysterectomy. Detailed reports of the trial can be found elsewhere [40, 41]. Patients were divided into two groups, those undergoing vaginal hysterectomy and those undergoing abdominal hysterectomy, and were analysed and reported separately.

Randomization was 2-to-1 in favour of laparoscopy. Allocation by treatment group and type of hysterectomy is given in Table IV. Effectiveness was expressed in quality-adjusted life years, and measured using the EQ-5D questionnaire [42]. EQ-5D measurements were taken at randomization, and at three follow-up visits: 6 weeks, 4 months and 1 year. These follow-up visits defined the three time intervals (i.e. $K = 3$) with $a_1 = 0$, $a_2 = 6$, $a_3 = 17$ and $a_4 = 52$, expressed in weeks. Since the boundaries of the time interval coincided with the EQ-5D measurements (as they often would), q_{ik}, the quality-adjusted life-year score for interval k, equals $(Q_{i,k+1} + Q_{ik})/2 \times (a_{k+1} - a_k)/52$, where Q_k is the EQ-5D measurement at a_k. Total costs, expressed in U.K. pounds, were collected for these time intervals also. Although there were no deaths, several patients failed to compete all follow-up visits and were considered censored. Over all treatment arms, all patients had observations for the first interval, 97.5 per cent had observations for the first two intervals, and 86.5 per cent had observations

Table V. Potential continuous covariates by treatment and type of hysterectomy, EVALUATE.

		Vaginal			Abdominal		
		Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
Baseline EQ-5D	Laparoscopic	0.746	0.256	0.760	0.716	0.266	0.760
	Standard	0.758	0.232	0.796	0.691	0.266	0.725
Age	Laparoscopic	41.0	6.83	41.1	41.8	7.00	41.9
	Standard	41.1	6.45	41.0	41.5	7.63	40.9
BMI	Laparoscopic	26.5	5.20	25.5	26.6	5.08	26.0
	Standard	26.6	4.87	26.2	26.1	5.49	24.9

Table VI. Potential binary covariates by treatment and type of hysterectomy, EVALUATE.

		Vaginal (per cent)	Abdominal (per cent)
Smoker	Laparoscopic	37.8	41.7
	Standard	40.3	47.1
Previous pelvic surgery	Laparoscopic	1.62	0.54
	Standard	0	0.73

for all three intervals. The censoring pattern was essentially the same in all treatment arms. The following variables were considered as potential covariates in the regression models: EQ-5D at randomization (Q), age (A), body mass index (B), smoking status (K), and whether or not the patient had previous pelvic surgery (P). $K = 1$ if the patient smoked, 0 otherwise, and $P = 1$ if the patient had previous pelvic surgery, 0 otherwise. Descriptive statistics of the covariates by treatment arm are given in Tables V and VI.

The first regression models included the treatment variable alone. The final regression models included a potential covariate if its corresponding significance level was less than or equal to 0.05 in the model including the treatment variable and all other significant covariates. The covariate parameter estimates and their corresponding significance levels are given in Table VII. A 1 year increase in age is observed to increase cost by £24 for a vaginal hysterectomy and by £22 for an abdominal hysterectomy. Smoking is observed to reduce cost by £186 for a vaginal hysterectomy. Not surprisingly, higher EQ-5D measures at randomization are associated with higher QALYs during follow-up for both types of hysterectomies. For abdominal hysterectomy, a one unit increase in BMI is observed to increase cost by £18. Also for abdominal hysterectomy, previous surgery is observed to increase cost by £2397 and reduce QALYs in follow-up by 0.1394. When the significant covariates are adjusted for, the treatment difference in effectiveness more than doubles for vaginal hysterectomies and almost halves for abdominal hysterectomies. These differences have a substantial effect on the ICER but almost no effect on the INB evaluated at either £10 000 or £20 000, see Table VIII.

Table VII. Parameter estimates, standard errors and p-values, EVALUATE.

Parameter		Vaginal		Abdominal	
		No covariates	Final model*	No covariates	Final model*
$\hat{\Delta}_c$	Estimate	400.8	425.2	185.8	177.7
	SE	69.79	70.72	100.8	102.8
	p-value	<0.0001	<0.0001	0.065	0.084
$\hat{\beta}_{C,A}$	Estimate		24.05		21.93
	SE	n.i.m.	5.399	n.i.m.	9.744
	p-value		<0.0001		0.024
$\hat{\beta}_{C,B}$	Estimate				18.21
	SE	n.i.m.	n.i.m.	n.i.m.	8.169
	p-value				0.026
$\hat{\beta}_{C,K}$	Estimate		-185.5		
	SE	n.i.m.	73.700	n.i.m.	n.i.m.
	p-value		0.012		
$\hat{\beta}_{C,P}$	Estimate				2397
	SE	n.i.m.	n.i.m.	n.i.m.	1201
	p-value				0.046
$\hat{\Delta}_e$	Estimate	0.001542	0.003831	0.009148	0.005077
	SE	0.01031	0.009778	0.01008	0.009726
	p-value	0.881	0.695	0.364	0.602
$\hat{\beta}_{E,Q}$	Estimate		0.1505		0.1656
	SE	n.i.m.	0.02405	n.i.m.	0.01918
	p-value		<0.0001		<0.0001
$\hat{\beta}_{E,K}$	Estimate				-0.0242
	SE	n.i.m.	n.i.m.	n.i.m.	0.009110
	p-value				0.010
$\hat{\beta}_{E,P}$	Estimate				-0.1394
	SE	n.i.m.	n.i.m.	n.i.m.	0.06391
	p-value				0.029
$\hat{C}(\hat{\Delta}_e, \hat{\Delta}_c)$		-0.08368	-0.1034	-0.2285	-0.2411

*Includes only significant ($p<0.05$) variables.
n.i.m., not in model.

The only significant covariate by treatment interaction was with smoking in the cost model for the vaginal group. The coefficient was -505.31, with a p-value of 0.0003. The negative coefficient indicates that the increase in cost associated with the laparoscopic procedure was observed to be £505.30 more for non-smokers (£624.35 versus £119.05). To examine the interaction between treatment and smoking on the INB scale, we added to the final model

Table VIII. Unadjusted and adjusted ICER and INB with 90 per cent confidence limits, EVALUATE.

		Unadjusted	Adjusted
Vaginal	ICER	259 983	111 005
	(confidence limits)	(19 981, −25 910)	(19 502, −35 144)
	INB(10 000)	−385	−387
	(confidence limits)	(−600, −169)	(−599, −175)
	INB(20 000)	−369	−348
	(confidence limits)	(−740, 0.325)	(−706, 9.47)
Abdominal	ICER	20 312	34 998
	(confidence limits)	(1438, −25 649)	(900, −14 404)
	INB(10 000)	−94.3	−127
	(confidence limits)	(−353, 165)	(−386, 132)
	INB(20 000)	−2.86	−76.2
	(confidence limits)	(−405, 400)	(−473, 320)

Table IX. INB interaction between treatment and smoking, EVALUATE—vaginal.

λ	$\hat{\beta}_{E,T\times K}\lambda - \hat{\beta}_{C,T\times K}$	S.E.	p-value
5000	182.4	544.2	0.0028
10 000	583.1	258.3	0.0239
15 000	621.9	346.3	0.0724
20 000	660.8	439.2	0.1323
25 000	699.7	534.4	0.1904

given in Table VII the smoking by treatment interaction in the cost model and smoking and the smoking by treatment interaction in the effectiveness model. Then we tested the hypothesis $H: \beta_{E,T\times K}\lambda - \beta_{C,T\times K} = 0$, where $T \times K$ indicates the treatment by smoking interaction. The estimate $\hat{\beta}_{E,T\times K}\lambda - \hat{\beta}_{C,T\times K}$, its standard error and associated p-value are given in Table IX for a several values of λ. For all values of λ the estimate is positive, indicating that the INB is observed to be greater among smokers, reaching significance for WTP values less than £10 000/QALY. Since there was no *a priori* reason to suspect an interaction of this type, and because we explored several possible interactions, these results should be considered exploratory at best.

4. DISCUSSION

We have provided a simple but valid approach for covariate adjustment for censored cost-effectiveness analysis where the measure of effectiveness is either survival time or quality-adjusted survival time. The regression models are linear and the pattern of censoring is assumed to be purely random. The proposed estimators, which are based on simple inverse probability of censored weighting may not be efficient. Robbins and Rotnitzky [43] provide

a general theory for efficient estimation based on inverse probability of censoring weighting. However, due to the complex nature of the cost accumulation process, it would be difficult, if not impossible, to derive asymptotically efficient estimators for our setting.

We have specified linear regression models for both the measures of cost and effectiveness. For some applications linear models may be inappropriate. Recently, Lin [37] proposed a class of generalized linear models for cost data and developed the corresponding estimators. He argued that multiplicative models, which are special cases of generalized linear models, may be more appropriate than linear models in some applications. Combining the ideas of Lin [37] with those of this paper, we may develop methods for covariate adjusted cost-effectiveness analyses using generalized linear models or other non-linear models.

For simplicity of presentation we have assumed that the censoring arises in a purely random fashion. If this is not true, the censoring time distribution may be modelled using proportional hazards models, see Lin [25, 37], however the resulting estimators are far more complicated. Lin [37] discusses the conditions under which the inverse probability weighting can be applied.

ACKNOWLEDGEMENTS

The authors wish to thank Ms Susan Tomlinson for her meticulous proof reading. Of course, all remaining errors are our own responsibility. ARW is funded through the Discovery Grant Program of the Natural Science and Engineering Research Council of Canada (Grant Number 44868-03), and DYL is supported by the National Cancer Institute Grant R01 CA082659, however the opinions expressed are those of the authors 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 *Valuing Health Care*, Sloan F (ed.). Cambridge University Press: Cambridge, 1994; 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 *Cost Effectiveness in Health and Medicine*, Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Oxford University Press: New York, 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. *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. 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. Willan AR, Briggs AH, Hoch JS. Regression methods for non-censored cost-effectiveness data. *Health Economics* 2004; **13**:461–475.
25. Lin DY. Linear regression analysis of censored medical costs. *Biostatistics* 2000; **1**:35–47.
26. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997; **53**:419–434.
27. Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika* 2000; **87**:329–343.
28. 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.
29. Zhao H, Tian L. On estimating medical cost and incremental cost-effectiveness ratios with censored data. *Biometrics* 2001; **57**:1002–1008.
30. Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* 1997; **84**:339–348.
31. Zhao H, Tsiatis AA. Efficient estimation of the distribution of quality adjusted survival time. *Biometrics* 1999; **55**:1101–1107.
32. Zhao H, Tsiatis AA. Testing equality of survival functions of quality adjusted lifetime. *Biometrics* 2001; **57**: 861–867.
33. Strawderman RL. Estimating the mean of an increasing stochastic process at a censored stopping time. *Journal of the American Statistical Association* 2000; **95**:1192–1208.
34. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Statistics in Medicine* 1990; **9**: 1259–1276.
35. Hwang JS, Tsauo JY, Wang JD. Estimation of expected quality adjusted survival by cross-sectional survey. *Statistics in Medicine* 1996; **15**:93–102.
36. Shen LZ, Pulkstenis E, Hoseyni M. Estimation of mean quality adjusted survival time. *Statistics in Medicine* 1999; **18**:1541–1554.
37. Lin DY. Regression analysis of incomplete cost data. *Statistics in Medicine* 2003; **22**:1181–1200.
38. Connolly SJ, Gent M, Roberts RS, Dorian P, Roy D, Sheldon RS, Mitchell LB, Green MS, Klein GJ, O'Brien B. Canadian implantable defibrillator study (CIDS): a randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation* 2000; **101**:1297–1302.
39. O'Brien BJ, Connolly SJ, Goeree R, Blackhouse G, Willan AR, Yee R, Roberts RS, Gent M. Cost-effectiveness of the implantable cardioverter defibrillator results from the Canadian implantable defibrillator study (CIDS). *Circulation* 2000; **103**:1416–1421.
40. Garry R, Hawe J, Abbott J *et al.* The eVALuate study: two parallel randomised trials, one comparing laparoscopic with abdominal hysterectomy, the other comparing laparoscopic with vaginal hysterectomy. *British Medical Journal* 2004; **328**:129.
41. Sculpher M, Manca A, Abbott J, Fountain J, Mason S, Garry R. Cost-effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. *British Medical Journal* 2004; **328**:134.
42. Kind P. The EuroQol instrument: an index of health-related quality of life. In *Quality of Life and Pharmacoeconomics in Clinical Trials* (2nd edn), Spilker B (ed.). Lippincott-Raven: Philadelphia, 1996; 191–201.
43. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*: *Methodological Issues*, Jewell NP, Dietz K, Farewell VT (eds). Birkhauser: Boston, 1992; 297–331.

