UNIVERSITY OF WISCONSIN

DEPARTMENT OF BIOSTATISTICS AND MEDICAL INFORMATICS

Technical Report **April 2010**
 # 210

A novel semiparametric regression method
for interval-censored data

Seungbong Han
Adin-Cristian Andrei
Kam-Wah Tsui

UNIVERSITY OF WISCONSIN
DEPARTMENT OF BIOSTATISTICS
AND MEDICAL INFORMATICS
K6/446 Clinical Science Center
600 Highland Avenue
Madison, Wisconsin 53792-4675
(608) 263-1706

A novel semiparametric regression method for interval-censored data

Seungbong Han [1], Adin-Cristian Andrei [2,*], and Kam-Wah Tsui [1]

[1]Department of Statistics, University of Wisconsin-Madison, MSC,

1300 University Avenue, Madison, WI, 53706, U.S.A.

[2]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison,

K6/428 CSC 600 Highland Avenue, Madison, WI, 53792-4675, U.S.A.

**email:* andrei@biostat.wisc.edu

SUMMARY: In many medical studies, event times are recorded in an interval-censored (IC) format. For example, in numerous cancer trials, time to disease relapse is only known to have occurred between two consecutive clinic visits. Many existing modeling methods in the IC context are computationally intensive and usually require numerous assumptions that could be unrealistic or difficult to verify in practice. We propose a novel, flexible and computationally efficient modeling strategy based on jackknife pseudo-observations (POs). The POs obtained based on nonparametric estimators of the survival function are employed as outcomes in an equivalent, yet simpler regression model that produces consistent covariate effect estimates. Hence, instead of operating in the IC context, the problem is translated into the realm of generalized linear models, where numerous options are available. Outcome transformations via appropriate link functions lead to familiar modeling contexts such as the proportional hazards and proportional odds. Moreover, the methods developed are not limited to these settings and have broader applicability. Simulations studies show that the proposed methods produce virtually unbiased covariate effect estimates, even for moderate sample sizes. An example from the International Breast Cancer Study Group (IBCSG) Trial VI further illustrates the practical advantages of this new approach.

KEY WORDS: Interval-censoring; Pseudo-observations; Regression; Semiparametric; Survival analysis.

1. Introduction

Interval-censored (IC) data are naturally and frequently encountered in numerous medical studies involving event times. Instead of being observed exactly, the event of interest is known to have occurred within a certain time interval, such as that between two consecutive patient examinations. For example, in the IBCSG Trial VI (International Breast Cancer Study Group, 1996), time to cancer progression is known to have occurred between the latest cancer progression-free visit and the date when disease-progression has been detected.

There is a rich literature in the area of modeling IC data, but a pervasive lack of software availability prevents their routine use in practice. The proportional hazards setting appears to be a favorite structural framework for methodology development (see, for example, Finkelstein, 1986, Alioum and Commenges, 1996 or Satten, 1996). More recently, Betensky et al. (2002) develop local likelihood estimation methods involving smoothing of the baseline hazard function, while Cai and Betensky (2003) assume a piecewise-linear baseline hazard and use penalized likelihood estimation methods. A useful alternative to the proportional hazards model is the proportional odds model. In this context, Murphy et al. (1997) carry out parameter and variance estimation via profile likelihood methods. Shen (1998) develops consistent estimators for the baseline hazard and the regression parameters using sieve maximum likelihood and monotone splines. In addition, Rabinowitz et al. (2000) propose an approximate conditional likelihood-based model and Sun et al. (2007) develop goodness-of-fit tests. Accelerated failure time models provide yet another useful methodological context, such as in Betensky et al. (2001) and Li and Pu (2003). An asymptotically efficient sieve maximum likelihood estimator is developed by Xue et al. (2006). For more on these and other existing methods, one might consult Tian and Cai (2006) or Sun (2006). In practice, middle-point imputation is used to substitute interval-censoring with right-censoring. Despite its

appeal, this oftentimes produces biased results, as discussed by Lindsey and Ryan (1998) and illustrated in our simulations.

We propose a novel and easy to implement semiparametric IC data modeling strategy in which the jackknife pseudo-observations (POs) (Quenouille, 1949 and Tukey, 1958) play a fundamental role. Employed in linear regression analyses (Wu, 1986, Simonoff and Tsai, 1986), the POs use in event time settings is relatively recent and remains specific to right-censored data contexts (see Andersen et al. (2003), Andersen et al. (2004) or Klein and Andersen (2005)). For completeness, we briefly describe how POs are obtained. In general, given an estimator $\hat{\theta}$ of θ based on an *i.i.d.* sample $T_1, \cdots, T_n$ and its version $\hat{\theta}_{(-i)}$ obtained using the reduced sample $T_1, \cdots, T_{i-1}, T_{i+1}, \cdots, T_n$, the i^{th} pseudo-observation is defined as $\nu_i = n\hat{\theta} - (n-1)\hat{\theta}_{(-i)}$, where $i = 1, \cdots, n$. For example, if $\theta = E(T)$ and $\hat{\theta} = n^{-1}\sum_{i=1}^{n} T_i$ is the sample mean, then $\nu_i = T_i$. However, when T_i is subject to interval-censoring, $\hat{\theta}$ estimators have no closed-form and $\nu_i \neq T_i$.

Let Z_i and $S(t|Z_i) = P(T_i > t|Z_i)$ be the covariate vector and the conditional survival function for individual i, respectively. Assume that one intends to construct a regression model $S(t|Z_i) = A(t) - \beta^T Z_i$, where t is fixed and $A(t)$ is a function of time. Let $\hat{S}(t)$ be a consistent estimator of the marginal survival function $S(t) = P(T_i > t)$ of T_i, obtained in nonparametric fashion or otherwise, and let $\hat{S}_{(-i)}(t)$ be its reduced sample version with the i^{th} observation excluded. Using pseudo-observations, a simple, yet very direct modeling strategy could be devised. We sketch a brief heuristic argument and recommend Andersen et al. (2004) for further justification. Noting that $S(t) = E_Z\{E[I(T > t)|Z]\} = E_Z[P(T > t|Z)]$, it is natural to estimate $S(t)$ by $\hat{S}(t) = n^{-1}\sum_{j=1}^{n} \hat{S}(t|Z_j)$ in which $\hat{S}(t|Z_j)$ is a consistent estimator for $S(t|Z_j)$ based on N samples. The leave-one-out estimator of $S(t)$ is $\hat{S}_{(-i)}(t) = (n-1)^{-1}\sum_{j=1,j\neq i}^{n} \tilde{S}(t|Z_j)$ where $\tilde{S}(t|Z_j)$ is a consistent estimator for $S(t|Z_j)$ based on $n-1$ reduced samples. When n is large, $\hat{S}(t|Z)$ and $\tilde{S}(t|Z)$ will be approximately equal. Therefore,

the i^{th} PO $\nu_i = n\hat{S}(t) - (n-1)\hat{S}_{(-i)}(t)$ approximates $\hat{S}(t|Z_i)$. Hence, i^{th} PO ν_i might be used as a substitute for $S(t|Z_i)$ in the desired regression model $S(t|Z_i) = A(t) - \beta^T Z_i$. As detailed at length in the next section, nonparametric estimation of $S(t)$ will provide the POs used in modeling. The implementation effort is minimal, therefore this method has an increased potential for immediate use in biomedical data analyses. Considering the lack of public software availability for more recent interval-censored data methodologies, this approach has a obvious advantage.

The rest of this article is organized as follows: we present the IC data modeling strategy in Section 2, while extensive simulation results showing excellent method performance in moderate sample sizes are shown in Section 3. An application to the IBCSG Trial VI data example constitutes Section 4, while further remarks are part of Section 5.

2. Pseudo-observation-based regression for IC data

Recently, POs have been used successfully in right-censored data modeling; examples include Graw et al. (2009), Andersen et al. (2004), Andrei and Murray (2007), Liu et al. (2008) or Logan et al. (2008). However, the use of POs in the interval-censored (IC) data context is novel and constitutes the central theme of this research. Assume that the observed data are an *i.i.d.* sample $\{(L_i, R_i], Z_i; i = 1, \cdots, n\}$, where $(L_i, R_i]$ is the observation window (censoring interval) for event time T_i and Z_i is a p-dimensional vector of covariates for subject i. Suppose that T_i is independent of $(L_i, R_i]$. By convention, $L_i = R_i$ yields the exact failure time T_i; $L_i = 0$ produces left-censoring, while $R_i = \infty$ creates right-censoring. Let $S(t|Z_i) = P(T_i > t|Z_i)$ be the conditional survival function of T_i, given Z_i, computed at a fixed time $t > 0$. Our primary goal is to model a large class of outcomes $g[S(t|Z_i)]$, where $g(\cdot)$ is a smooth link function. Assume that, for all $t > 0$ following model holds:

$$g[S(t|Z_i)] \quad = \quad A(t) - \beta^T Z_i$$

where β is a p-dimensional covariate-effect vector and the intercept $A(t)$ is a function of time. For example, $g(t)=log[-log(t)]$ leads to the proportional hazards model, while $g(t)=log(\frac{t}{1-t})$ yields the proportional odds model. Needed to define the pseudo-observations (POs), let $\hat{S}(t)$ be a consistent nonparametric estimator of the marginal survival function $S(t)$ and $\hat{S}_{-i}(t)$ be its leave-i^{th}-out version. Then, we define the i^{th} PO as

$$\nu_i(t) \;\; = \;\; ng\{\hat{S}(t)\} - (n-1)g\{\hat{S}_{-i}(t)\}.$$

It follows that $g\{\hat{S}(t)\}$ is a consistent estimator of $g\{S(t)\}$ and, instead of regressing $g\{S(t|Z_i)\}$ on Z_i, we will regress $\nu_i(t)$ on Z_i, thus regarding the POs as a valid substitute for $g\{S(t|Z_i)\}$. As pointed out by Tukey (1958), Andersen et al. (2004) or Graw et al. (2009), the POs could be regarded as independent, for all practical purposes. Importantly, the function $g(\cdot)$ is part of the PO definition. Authors such as Andersen et al. (2003) define the i^{th} PO as $\eta_i(t) = n\hat{S}(t)-(n-1)\hat{S}_{-i}(t)$ and fit the generalized linear model $\eta_i(t) = g^{-1}[A(t)-\beta Z_i]$, to estimate β. We have not altered the definition of the PO, but simply applied the transformation to the outcome $g\{S(t|Z_i)\}$ directly.

One may consider $J \geqslant 2$ distinct time points $0 < t_1 < \ldots < t_J < \infty$ and construct vector $\nu_i = \{\nu_i(t_1), \ldots, \nu_i(t_J)\}^T$ of POs, for each individual $i = 1, \ldots, n$. Recall that $Z_i = (Z_{i1}, \ldots, Z_{ip})^T$ is a p-dimensional covariate vector and let $\alpha = \{A(t_1), \ldots, A(t_J)\}^T$, $\gamma = (\alpha^T, \beta^T)^T$ and B be a $J \times p$-dimensional matrix whose rows are all equal to β^T. Then, one could obtain an estimate $\hat{\gamma}$ for γ by solving the following generalized estimating equations:

$$U(\gamma) = \sum_{i=1}^{n} U_i(\gamma) = \sum_{i=1}^{n} \left\{ \frac{\partial(\alpha + BZ_i)^T}{\partial \gamma^T} \right\}^T V_i^{-1}\{\nu_i - (\alpha + BZ_i)^T\} = 0,$$

where V_i is the usual working covariance matrix accounting for the correlation in ν_i.

The standard "sandwich" variance-covariance matrix of $\hat{\gamma}$ is obtained as

$$\widehat{Var}(\hat{\gamma}) \;\; = \;\; I(\hat{\gamma})^{-1}\widehat{var}\{U(\hat{\gamma})\}I(\hat{\gamma})^{-1},$$

where

$$I(\gamma) \;=\; \sum_{i=1}^{n} \left\{ \frac{\partial(\alpha + BZ_i)^T}{\partial \gamma^T} \right\}^T V_i^{-1} \left\{ \frac{\partial(\alpha + BZ_i)^T}{\partial \gamma^T} \right\}$$

and

$$\widehat{var}\{U(\hat{\gamma})\} \;=\; \sum_{i=1}^{n} U_i(\hat{\gamma}) U_i(\hat{\gamma})^T.$$

Note that when a single time-point t is considered, the equations above simplify a lot since $\alpha = A(t)$, $B = \beta^T$ and $V_i = 1$. See Andersen and Klein (2007) and Scheike and Zhang (2007) for further examples of generalized estimating equations use in modeling based on POs.

To obtain the POs, a consistent estimator of the survival function $S(t)$ is thus needed. In the following subsections, we describe two ways to achieve this: (i) Turnbull's self-consistent estimator and (ii) the EM iterative convex minorant (EMICM) algorithm proposed by Wellner and Zhan (1997). We begin by briefly describing the two algorithms that form the basis of our PO-based modeling strategy.

2.1 *Turnbull's self-consistent estimator of $S(t)$*

Turnbull (1976) developed an algorithm for estimating $S(t)$ as a solution to the so-called self-consistency equation. Assume $\{\tau_j\}_{j=0}^m$ are the increasingly ordered, unique elements of $\{0, L_i, R_i; i = 1, \cdots, n\}$, where $\tau_0 = 0$. Define $\alpha_{ij} = I\{\tau_j \in (L_i, R_i]\}$, $p_j = S(\tau_{j-1}) - S(\tau_j)$ and $\mathbf{p} = (p_1, \cdots, p_m)^T$. The likelihood function is equal to

$$\begin{aligned} L(\mathbf{p}) &= \textstyle\prod_{i=1}^{n} [S(L_i) - S(R_i)] \\ &= \textstyle\prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} p_j \end{aligned} \tag{1}$$

Naturally, the search for a non-parametric maximum likelihood estimator (NPMLE) of $S(t)$ is restricted to the space of piecewise constant, non-increasing functions with jumps only at times $\{\tau_j\}_{j=1}^m$. Thus, maximization of $L(\mathbf{p})$ with respect to $\mathbf{p}$ is subject to these constraints: $\sum_{j=1}^m p_j = 1$ and $p_j \geqslant 0$ for all j. Since exact failure times are unknown, at each τ_j and for the current estimates of $\mathbf{p}$, one recursively computes the expected number d_j of events and

the expected number Y_j of individuals "at-risk". As such, d_j is estimated by $\sum_{i=1}^{n} \frac{\alpha_{ij} p_j}{\sum_{k=1}^{m} \alpha_{ik} p_k}$ and Y_j by $\sum_{k=j}^{m} d_k$. One estimates $\mathbf{p}$ iteratively until a pre-specified convergence criterion is met. This estimator could also be regarded as an application of the EM algorithm (for details, see Sun (2006)). Maximization of $L(\mathbf{p})$ encounters practical difficulties when the set $\{\tau_j\}_{j=0}^{m}$ of support points is too sparse. An unfavorable choice of the starting value of $\mathbf{p}$ could result in non-convergence or convergence to a local maximum, instead of the desired global one (Wellner and Zhan, 1997). Besides, self-consistency equation roots are not automatically NPMLEs. However, an estimate $\hat{\mathbf{p}}$ is an NPMLE if $\sum_{i=1}^{n} \frac{\alpha_{ij} \hat{p}_j}{\sum_{k=1}^{m} \alpha_{ik} \hat{p}_k} = n$ for all $j = 1, \cdots, m$ (see Gentlemen and Geyer (1994)).

2.2 *The EMICM estimator of* $S(t)$

To address the shortcomings of the Turnbull's procedure, Groeneboom and Wellner (1992) have developed the iterative convex minorant (ICM) algorithm. Their approach recasts the likelihood maximization problem as a linear programming one. In related work, Li et al. (1997) and Jongbloed (1998) have provided improved algorithms and studied some of the properties of the NPMLE. Particularly important, Wellner and Zhan (1997) develop the EMICM algorithm, an approach that uses the ICM algorithm to search for the NPMLE among all self-consistent estimates produced by the EM algorithm. Through a composite mapping of the ICM and the EM algorithms, the algorithmic mapping of the EM iteration guarantees the ascent likelihood function in the modified ICM algorithm. Therefore, the EMICM algorithm is guaranteed to achieve global convergence rapidly and efficiently. Pan (1999) argues that ICM is a particular case of the generalized gradient projection method. In related work, Böhning et al. (1996), Goodall et al. (2004), Hudgens (2005), Pan and Chappell (1998) and Braun et al. (2005) discuss ways to construct confidence intervals based on the NPMLE when truncation is also present in the data. For a better understanding of the EMICM algorithm, we begin by briefly describing the ICM algorithm.

2.2.1 *The Iterative Convex Minorant Algorithm.* Groeneboom and Wellner (1992) propose an isotonic regression-based method to maximize $L(\mathbf{p})$, given in equation (1). The NPMLE is obtained via a linear programming problem, the Fenchel duality for convex optimization. Let $\gamma_j = F(\tau_j)$, $j = 1, \cdots, m$ and $\gamma = (\gamma_1, \cdots, \gamma_{m-1})$, where $\gamma_0 = 0$ and $\gamma_m = 1$. Consider $C_{\mathbf{x}}$ to be a subspace of $\Re^{m-1}$ and

$$C_{\mathbf{x}} \;=\; \{\mathbf{x} = (x_1, \cdots, x_{m-1}); 0 \leqslant x_1 \leqslant \cdots \leqslant x_{m-1} \leqslant 1\}.$$

The likelihood function is equal to

$$L(\gamma) \;=\; \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij}(\gamma_j - \gamma_{j-1}).$$

Assume C is a convex cone in $\Re^{m-1}$ and $\phi(\mathbf{x})$ is a continuous, differentiable and concave mapping from $\Re^{m-1}$ to $\Re$. Let $\hat{\mathbf{x}} = argmax_{\mathbf{x}\in C}\phi(\mathbf{x})$. Following Groeneboom and Wellner (1992), when $\hat{\mathbf{x}}$ is known, the maximization of $\phi(\mathbf{x})$ is equivalent to the maximization of the following quadratic form:

$$\phi^{\star}(\mathbf{x}|\mathbf{y}, \mathbf{W}) = -\tfrac{1}{2}(\mathbf{x} - \mathbf{y})^T \mathbf{W}(\mathbf{x} - \mathbf{y}), \tag{2}$$

for some fixed $\mathbf{y} \in \Re^{m-1}$ and a positive definite diagonal $(m-1) \times (m-1)$ matrix $\mathbf{W} = diag\,(w_1, \cdots, w_{m-1})$. The idea of the ICM algorithm is to approximate the concave function $\phi(\mathbf{x})$ locally by the given quadratic form $\phi^{\star}(\mathbf{x}|\mathbf{y}, \mathbf{W})$. From Robertson et al. (1988), one can maximize the quadratic function by exploiting the connection between the solution of the isotonic regression problem and the derivative of the convex minorant. Define $B_0 = (0, 0)$ and

$$B_k \;=\; \left(\sum_{i=1}^{k} w_i, \sum_{i=1}^{k} w_i y_i\right),\; 1 \leqslant k \leqslant m-1,$$

points in $\Re^2$ for some fixed $\mathbf{y}$= $(y1, \cdots, y_{m-1})^T \in \Re^{m-1}$. The set of points $\{B_k; k = 0, \cdots, m-1\}$ is called a cumulative sum (CUSUM) diagram. Then, $\hat{x}_j$ is attained by the derivative of the convex minorant of the points $\{B_k; 0 \leqslant k \leqslant m-1\}$ evaluated at B_j. Because $\hat{\mathbf{x}}$ is unknown, the value $\hat{\mathbf{x}}$ that maximizes $\phi^{\star}(\mathbf{x}|\mathbf{y}, \mathbf{W})$ given in equation (2),

is obtained iteratively. At step l, let $\hat{\gamma}^{(l)} = (\hat{\gamma}_1^{(l)}, \cdots, \hat{\gamma}_{m-1}^{(l)})^T$ and $\nabla log\{L(\hat{\gamma}^{(l)})\}$ be the vector of first-order derivatives of the log-likelihood function, computed at $\hat{\gamma}^{(l)}$. Denote $\mathbf{y} = \hat{\gamma}^{(l-1)} - \mathbf{W}^{-1}(\hat{\gamma}^{(l-1)})\nabla log\{L(\hat{\gamma}^{(l-1)})\}$ and $\mathbf{W}(\gamma^{(l-1)}) = diag\left(w_1(\gamma^{(l-1)}), \cdots, w_{m-1}(\gamma^{(l-1)})\right)$. The ICM algorithm is described below:

(1) Select an initial estimate $\hat{\gamma}^{(0)}$ of γ

(2) Let $w_i^{(l-1)}$ is the i^{th} diagonal element of $\mathbf{W}(\hat{\gamma}^{(l-1)})$. Then we obtain the updated estimate $\hat{\gamma}^{(l)}$ at iteration step l by maximizing $\phi^{\star}(\mathbf{x}|\mathbf{y}, \mathbf{W}(\hat{\gamma}^{(l-1)}))$. That is, $\hat{\gamma}^{(l)}$ is obtained by taking the derivative of the convex minorant of the cumulative sum diagram $\{B_k; 0 \leqslant k \leqslant m-1\}$ which is defined by $B_0 = (0,0)$ and

$$B_k = \left(\sum_{i=1}^{k} w_i^{(l-1)}, \sum_{i=1}^{k}\left(w_i^{(l-1)}\hat{\gamma}_i^{(l-1)} - \frac{\partial}{\partial\gamma_i} log(L(\hat{\gamma}^{(l-1)}))\right)\right),\ 1 \leqslant k \leqslant m-1.$$

The natural choice of $w_j(\gamma^{(l-1)})$, $j = 1, \cdots, m-1$ is to take $-\frac{\partial^2}{\partial\gamma_j^{(l-1)}\partial\gamma_j^{(l-1)}} log\{L(\gamma^{(l-1)})\}$ (Jongbloed, 1998).

(3) Repeat Step 2 until $\hat{\gamma}$ satisfies a convergence criterion.

However, the ICM algorithm may not attain global convergence. To address this shortcoming, Jongbloed (1998) proposes an improved line search method. Suppose $\mathbf{z}$ maximizes $\phi^{\star}(\mathbf{z}|\mathbf{y},\mathbf{W})$ with $\mathbf{y} = \mathbf{x} - W^{-1}(\mathbf{x})\nabla log\{L(\mathbf{x})\}$. For $\mathbf{x} \neq \hat{\mathbf{x}}$ and sufficiently small $\lambda > 0$, $\phi(\mathbf{x} + \lambda(\mathbf{z} - \mathbf{x})) > \phi(\mathbf{x})$, which means the log likelihood function can be increased by an additional line search algorithm. Therefore, for $0 < \epsilon < 0.5$, the following line search algorithm (Step 2.1) can be carried out between step (2) and (3) of the ICM algorithm:

Step (2.1). If $log\{L(\hat{\gamma}^{(l)})\} > log\{L(\hat{\gamma}^{(l-1)})\} + (1-\epsilon)[\nabla log\{L(\hat{\gamma}^{(l-1)})]'(\hat{\gamma}^{(l)} - \hat{\gamma}^{(l-1)})$, then go to step (3). Otherwise, find $\mathbf{z} = \hat{\gamma}^{(l-1)} + \lambda(\hat{\gamma}^{(l)} - \hat{\gamma}^{(l-1)})$ for $0 \leqslant \lambda \leqslant 1$, and $\mathbf{z}$ should satisfy the inequality $\epsilon[\nabla log\{L(\hat{\gamma}^{(l-1)})\}]^T(\mathbf{z} - \hat{\gamma}^{(l-1)}) \leqslant log\{L(\mathbf{z})\} - log\{L(\hat{\gamma}^{(l-1)})\} \leqslant (1-\epsilon)[\nabla log\{L(\hat{\gamma}^{(l-1)})\}]^T(\mathbf{z} - \hat{\gamma}^{(l-1)})$.

2.2.2 *The EM Iterative Convex Minorant Algorithm.* To obtain the NPMLE at a greater speed and to achieve global convergence, Wellner and Zhan (1997) have proposed the EMICM algorithm that combines EM and ICM. EMICM does not require a line search, but updates the estimates of $\mathbf{p}$ by alternating between the EM and the ICM algorithms, until convergence. Wellner and Zhan (1997) propose three convergence criteria. Given $\epsilon > 0$, a fixed permissible error, first criterion is based on the maximum coordinate difference between consecutive estimates $\gamma^{(l)}$ and $\gamma^{(l-1)}$, namely

$$max_j \left|\hat{\gamma}_j^{(l)} - \hat{\gamma}_j^{(l-1)}\right| < \epsilon, \ for\ j = 1, \cdots, m-1.$$

The second criterion, based on the log-likelihood function, declares convergence if

$$\left|log(L(\hat{\gamma}^{(l)})) - log(L(\hat{\gamma}^{(l-1)}))\right| < \epsilon.$$

The last criterion, based on the Fenchel conditions (Robertson et al., 1988), may require more iterations, but it is useful in evaluating whether local or global convergence has occurred. Thus, $\hat{\gamma}^{(l)}$ is an NPMLE if

$$\left|\sum_{j=1}^{m-1} \hat{\gamma}_j^{(l)} \frac{\partial}{\partial \gamma_j} log(L(\hat{\gamma}^{(l)}))\right| < \epsilon,$$

and

$$max\left\{\sum_{k=j}^{m-1} \frac{\partial}{\partial \gamma_k} log(L(\hat{\gamma}^{(l)}))\right\} < \epsilon,$$

for $j = 1, \cdots, m-1$.

2.3 *Pseudo-observations-based regression*

We can use either Turnbull's (TB) self-consistent estimator or the EMICM algorithm to estimate the survival function $S(t)$ that will then allow us to obtain pseudo-observations at a single or at multiple timepoints. In simulated and real data examples, the TB approach is denoted by PO-TB, while that using the EMICM algorithm is denoted by PO-EMICM. In all numerical calculations, we use two **R** functions: (i) *EMICM* in the **Icens** package (Gentleman

and Vandal, 2009) for EMICM algorithm and (ii) *geese* from the **geepack** package (Yan, 2002) for GEE implementation.

3. Simulation Studies

We apply PO-EMICM and PO-TB to simulated data generated from: (1) proportional hazards, (2) proportional odds and (3) accelerated failure time models. Under each scenario, samples of size $n = 100$ are generated and $1,000$ replications are performed. Suppose that T_i is the survival time for subject i and let IQR and $t_{(q)}$ be its inter-quartile range and q^{th} percentile, respectively.

In all scenarios, POs are obtained both at single point $t_{(50)}$, as well as at three time points $(t_{(45)}, t_{(50)}, t_{(55)})$. In the later situation, a first-order autoregressive working correlation matrix V_i is assumed. Throughout, we consider 2-dimensional covariates $Z_i = (Z_{i1}, Z_{i2})$. For $i = 1, \ldots, n$, we create random-width censoring intervals $(L_i, R_i]$ as follows: for a fixed $W > 0$, define $L_i = max(T_i - c_{iL},\ 0)$ and $R_i = T_i + c_{iR}$, with c_{iL} and c_{iR} randomly drawn from the uniform $U(0,\ 2W)$ distribution. As W increases, the width of these intervals increases, on average.

In the first scenario (proportional hazards), the underlying model is

$$log\left[-log\left\{ S(t|Z) \right\}\right] = log\left\{ \int_0^t h_0(u)du \right\} - \beta^T Z,$$

where $h_0(\cdot)$ is the baseline hazard function. This model corresponds to the $g(t) = log\{-log(S(t)\}$ transformation function. We fix $W = IQR$ and compare PO-EMICM and PO-TB covariate effect estimates to those based on the middle-point Cox proportional hazards model (COX-MPI). Specifically, one imputes $(L_i + R_i)/2$ for T_i, when both R_i and L_i are finite. Given the lack of public software for semi-parametric IC data methods, it is routine to resort to this in practice. The **R** function *coxph* is used to fit the COX-MPI. Two baseline hazard functions $h_0(t) = t^3$ or $h_0(t) = t^4$ are considered. For $\beta_1 = -0.5$ and $\beta_2 = 1$, covariate vector

$Z_i = (Z_{i1}, Z_{i2})$ is such that Z_{i1} and Z_{i2} are independent, $Bernoulli(0.5)$ and $N(3, \sigma^2 = 0.5)$-distributed, respectively. Summary results presented in Table 1 include: the empirical mean of the $\hat{\beta}$ estimates, the average of the estimated $\hat{\beta}$ standard errors, the empirical standard error of the $\hat{\beta}$ estimates and the coverage probability of the true β by the nominal 95% confidence intervals.

[Table 1 about here.]

PO-EMICM and PO-TB perform very well in all scenarios, producing virtually unbiased estimators for (β_1, β_2), with appropriate coverage probabilities. Resulting parameter estimates when POs are obtained at multiple timepoints are more efficient than their single-point counterparts. PO-EMICM appears to be more efficient than PO-TB, mainly due to the benefits conferred by the ICM algorithm. Although more efficient, since it uses information across the entire survival curve, the middle-point imputation approach, COX-MPI, produces substantially biased $\hat{\beta}_1$ and $\hat{\beta}_2$ estimates, fact also reflected in the coverage probabilities below the nominal level.

The second simulation set is devised for the proportional odds model

$$logit \left\{ S(t|Z) \right\} = logit \left\{ S_0(t) \right\} - \beta^T Z,$$

with baseline survival function $S_0(t)$. The appropriate link function is obviously $g(t) = logit(t)$. We let $S_0(t) = e^{-t}$, $W = 0.3IQR$, and generate $Z_i = (Z_{i1}, Z_{i2})$ as follows: for $(\beta_1, \beta_2) = (1, 2)$, we generate Z_{i1} and Z_{i2} from independent uniform $U(1, 2)$ and $U(2, 3)$ distributions, respectively; when $(\beta_1, \beta_2) = (1, 1)$, Z_{i1} and Z_{i2} are generated independently from $U(1, 2)$ and $Bernoulli(0.5)$ distributions, respectively. For comparison, we also employ the method developed by Martinussen and Scheike (2006) for right-censored data obeying a proportional odds model. As such, for all finite intervals $(L_i, R_i]$, we resort once more to middle-point imputation and present our results under MS-MPI. This method is implemented using the **R** function *prop.odds* in the **timereg** (Scheike et al., 2009) package. Results shown

in Table 2, indicate that the proposed PO-EMICM and PO-TB methods perform very well, in terms of bias and coverage probabilities. MS-MPI appears to underestimate both β_1 and β_2, while being more efficient.

[Table 2 about here.]

The third simulation study is designed for a particular accelerated failure time model

$$T = exp(\beta_1 Z_{i1} + \beta_2 Z_{i2} + U),$$

where $exp(U)$ follows the $Gompertz(1,1)$ distribution. The conditional survival function $S(t|Z_i)$ follows the model

$$S(t|Z_i) = exp\left[1 - exp\left\{te^{-(\beta_1 Z_{i1} + \beta_2 Z_{i2})}\right\}\right].$$

Thus, the transformation function is $g(t) = log[log\{1 - log(t)\}]$. We let $W = IQR/2$ and generate data as follows: when $(\beta_1, \beta_2) = (1, -0.5)$, Z_{i1} and Z_{i2} are generated independently from $U(2,3)$ and $Bernoulli(0.5)$ distributions, respectively; when $(\beta_1, \beta_2) = (1, 0.5)$, we generate Z_{i1} and Z_{i2} independently from $U(2,3)$ and $N(2, \sigma^2 = 0.5)$, respectively. For comparison, we use a parametric accelerated failure time model based on the middle point imputations (AFT-MPI). To implement the parametric approach, we use the function *survreg* from the **R** package **survival** (Therneau and Lumley, 2009) and assume a Weibull distribution. Simulation results displayed in Table 3 indicate that both PO-EMICM and PO-TB produce good estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$, and the 95% confidence intervals' coverage probabilities are very close to the nominal level. On the other hand, AFT-MPI produces biased parameter estimates in all scenarios.

[Table 3 about here.]

4. The IBCSG Trial VI Study

Breast cancer is one of the most common types of cancer in the United States. Bonadonna et al. (1976) has proposed a prolonged cyclic adjuvant chemotherapy for breast cancer,

showing that a 12-month postoperative combination chemotherapy of cyclophosphamide, methotrexate and fluorouracil (CMF) is associated with a decreased risk of breast cancer recurrence of in women with positive axillary lymph nodes. The International Breast Cancer Study Group (IBCSG) Trial VI has investigated the optimal duration and timing of this combination chemotherapy.

Of interest was how CMF, with or without subsequent re-introduction, is associated with time to breast cancer recurrence, as well as overall survival. Between July 1986 to April 1993, eligible patients were randomly assigned, with equal probability, to one of the following four regimens: (i) CMF for six initial consecutive courses on months 1-6 (CMF6); (ii) CMF for six initial consecutive courses on months 1-6 plus three single courses of re-introduction CMF given on months 9, 12, 15 (CMF6+3); (iii) CMF for three initial consecutive courses on month 1-3 (CMF3) and (iv) CMF for three initial consecutive courses on months 1-3 plus three single courses of re-introduction CMF given on months 6, 9, 12 (CMF3+3).

Based on repeated clinic visits, recurrence time was assessed as the difference between the randomization date and the relapse date, defined as the visit when cancer relapse was established. The inter-visit average duration was 6.9 months, thus pinning down the precise time of cancer recurrence is problematic. Conceivably, cancer relapse does not occur exactly at the visit time when it is established, but sometime before that. Therefore, it might be more appropriate to treat the time-to-relapse as an interval-censored event occurring between the visit when recurrent disease was established and the last cancer-free visit. Accordingly, we model disease free survival (DFS) time as interval-censored data. To compare DFS under the four CMF regimens, we assume an underlying proportional hazards model. It is well-established (see International Breast Cancer Study Group (1996) and Gruber et al. (2008)) that patient age at baseline and node group are significantly associated with DFS. We focus attention on the group of 131 women age 41 years or younger at baseline, with four or more

cancer nodules. Figure 1 depicts the DFS estimates in the four groups obtained based on the EMICM algorithm. The highest risk of relapse is associated with CMF3, while the best prognosis is for patients receiving the standard CMF6 regimen.

To compare the four CMF regimens, we present unadjusted, as well as adjusted models. Adjustment factors are: tumor grade (1 (reduced), 2, 3(increased)), tumor size ($\leqslant$ 2 vs. $>$ 2 cm across), vessel invasion (yes/no), estrogen receptor (ER) status (negative/positive), progesterone receptor (PR) status (negative/positive).

We present least squares estimates when the POs used in PO-EMICM are computed only at the median timepoint 1.75. When POs computed at three timepoints (1.65, 1.75 and 1.98, representing the 45^{th}, 50^{th} and 55^{th} percentiles of the EMICM-based relapse time distribution estimator), we employ GEEs with a first-order autoregressive working correlation matrix. In addition, employ the middle-point imputation approach (COX-MPI) described in simulations, thus replacing the interval-censored with a right-censored data structure. Results presented in Table 4 include the hazard ratio (HR) estimates, together with corresponding 95% confidence intervals and p-values (P).

[Figure 1 about here.]

[Table 4 about here.]

Analyses employing the two POs-based method, labeled PO-EMICM(1) and PO-EMICM(3) reveal some interesting findings. Mortality in the CMF3 arm is significantly higher than in the CMF6 (reference) group, both in adjusted and unadjusted models. Importantly, in adjusted models, the estimated hazard rate in more than twice as high in the CMF3 group, compared to the CMF6 standard regimen (HR=2.18 (p-value=0.006) in PO-EMICM(1) and HR=2.37 (p-value=0.009) in EM-ICM(3)). Using the middle-point imputation strategy (COX-MPI), in which data are treated as right-censored, no significant differences are found between CMF6 and any of the other three regimens. For example, the estimated hazard ration for CMF3 in

the adjusted model is equal to 1.93 (p-value of 0.054), thus not significantly different from 1 at a 5% significance level. In conclusion, recognizing the true nature of the data (interval-censored, in this case) is important and can have major implications. Although convenient, the practice of imputing the middle point and then treating the data as right-censored, may lead to biased results or, as seen here, to a lack of statistical significance.

5. Discussion

This article presents a novel, PO-based regression method, for modeling interval-censored event times. Many existing nonparametric or semiparametric methods for IC data do not seem to be used routinely, likely due to a lack of software availability. The proposed method is computationally simple, thus convenient to implement using standard software. POs are constructed using an NPMLE of the survival function. Because the NPMLE does not have a closed-form in interval-censored data, two iterative algorithms (Turnbull's method and EMICM) are used in this methodological development. However, we emphasize an important distinction. Existing estimation and testing methods for interval-censored data usually employ EM-type algorithms to estimate covariate effects, with inherited potential problems, such as local convergence or lack thereof. Henschel et al. (2007) have indicated similar problems with algorithm convergence. By contrast, our PO-based approach leads to covariate effect estimates that are obtained in a direct fashion, using GEE or least-squares regression. Iteration is only required to estimate the survival function and EMICM guarantees global convergence to the NPMLE. Furthermore, using our approach, robust variance estimates are readily available, thus facilitating significance testing and confidence intervals construction. Importantly, model misspecification will lead to biases covariate effect estimates. For example, if the true underlying model obeys proportional hazards, yet the fitted model assumed proportional odds, the resulting parameter estimates will be incorrect, although their statistical significance may be preserved.

Acknowledgments

The authors would like to thank the IBCSG for permission to use their data. ACA's research is supported in part by following grants: P30 CA014520-36, UL1 RR025011-03, R21 CA132267-02 and W81XWH-08-1-0341. KWT's research is supported in part by the NSF grant DMS-0604931.

References

Alioum, A. and Commenges, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics* **52,** 512–524.

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Life Time Data Analysis* **10,** 335–350.

Andersen, P. K. and Klein, J. P. (2007). Regression analysis for multistate models based on a pseudo-value approach with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics* **34,** 3–16.

Andersen, P. K., Klein, J. P., and Rosthϕj, S. (2003). Generalized linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90,** 15–27.

Andrei, A. C. and Murray, S. (2007). Regression models for the mean of quality-of-life-adjusted restricted survival time using pseudo-observations. *Biometrics* **63,** 398–404.

Betensky, R. A., Lindsey, J. C., Ryan, L. M., and Wand, M. P. (2002). A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine* **21,** 263–275.

Betensky, R. A., Rabinowitz, D., and Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika* **88,** 703–711.

Böhning, D., Schlattmann, P., and Dietz, E. (1996). Interval censored data: A note on the nonparametric maximum likelihood estimator of the distribution function. *Biometrika* **83,** 462–466.

Bonadonna, G., Brusamolino, E., Valagussa, P., and et al. (1976). Combination chemotherapy as an adjuvant treatment in operable breast cancer. *New England Journal of Medicine* **294,** 405–410.

Braun, J., Duchesne, T., and Stafford, J. E. (2005). Local likelihood density estimation for interval censored data. *The Canadian Journal of Statistics* **33,** 39–60.

Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* **59,** 570–579.

Finkelstein, D. M. (1986). A proportional hazards model for interval censored failure time data. *Biometrics* **42,** 845–854.

Gentleman, R. and Vandal, A. (2009). *Icens: NPMLE for Censored and Truncated Data*. R package version 1.2.0.

Gentlemen, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika* **81,** 618–623.

Goodall, R. L., Dunn, D. T., and Babiker, A. G. (2004). Interval–censored survival time data: confidence intervals for the nonparametric survivor function. *Statistics in Medicine* **23,** 1131–1145.

Graw, F., Gerds, T. A., and Schumacher, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis* **15,** 241–255.

Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and non–parametric maximum likelihood*. DMV seminar, Band 19, Birkhauser, New York.

Gruber, G., Cole, B. F., Castinglione-Gertsch, M., and et al. (2008). Extracapsular tumor spread and the risk of local, axillary and supraclavicular recurrence in node-positive, premenopausal patients with breast cancer. *Annals of Oncology* **19,** 1393–1401.

Henschel, V., Heiß, C., and Mansmann, U. (2007). *intcox: Iterated Convex Minorant Algorithm for interval censored event data*. R package version 0.9.1.1.

Hudgens, M. G. (2005). On nonparametric maximum likelihood estimation with interval censoring and left truncation. *Journal of the Royal Statistical Society, Series B* **67,** 573–587.

International Breast Cancer Study Group (1996). Duration and reintroduction of adjuvant chemotherapy for node-positive premenopausal breast cancer patients. *Journal of Clinical Oncology* **14,** 1885–1894.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational & Graphical Statistics* **7,** 310–321.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling for competing risks data based on pseudo-values of the cumulative incidence function. *Biometrics* **61,** 223–229.

Li, L. and Pu, Z. (2003). Rank estimation of log-linear regression with interval censored data. *Lifetime Data Analysis* **9,** 57–70.

Li, L., Watkins, T., and Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics* **24,** 531–542.

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics methods for interval-censored data. *Statistics in Medicine* **17,** 219–238.

Liu, L., Logan, B. R., and Klein, J. P. (2008). Inference for current leukemia free survival. *Lifetime data analysis* **14,** 432–446.

Logan, B. R., Nelson, G. O., and Klein, J. P. (2008). Analyzing center specific outcomes in hematopoietic cell transplantation. *Lifetime data analysis* **14,** 389–404.

Martinussen, T. and Scheike, T. (2006). *Dynamic Regression Models for Survival Data.* Springer Verlag.

Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*

92, 968–976.

Pan, W. and Chappell, R. (1998). Estimating survival curves with left truncated and interval censored data via the EMS algorithm. *Communications in Statistics. Theory and Methods* **27,** 777–793.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B* **11,** 18–84.

Rabinowitz, D., Betensky, R. A., and Tsiatis, A. A. (2000). Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics* **56,** 511–518.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. John Wiley: New York.

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika* **83,** 355–370.

Scheike, T., Martinussen, T., and Silver, J. (2009). *timereg: timereg package for flexible regression models for survival data.* R package version 1.2-5.

Scheike, T. and Zhang, M. J. (2007). Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics* **34,** 17–32.

Shen, X. (1998). Proportional odds regression and sieve maximum likelihood estimation. *Biometrika* **85,** 165–177.

Simonoff, J. S. and Tsai, C. L. (1986). Jacknife-based estimators and confidence regions in nonlinear regression. *Technometrics* **28,** 103–112.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data*. Springer-Verlag: New York.

Sun, J., Sun, L., and Zhu, C. (2007). Testing the proportional odds model for interval-censored data. *Lifetime Data Analysis* **13,** 37–50.

Therneau, T. and Lumley, T. (2009). *survival: Survival analysis, including penalised likelihood.* R package version 2.35-7.

Tian, L. and Cai, T. (2006). On the accelerated failure time model for current status and interval censored data. *Biometrika* **93,** 329–342.

Tukey, J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics* **29,** 614.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B* **38,** 290–295.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* **92,** 945–959.

Wu, C. F. J. (1986). Jackknife,bootstrap and other resampling methods in regression analysis. *Annals of Statistics* **14,** 1261–1295.

Xue, H., Lam, K. F., Ben, C., and De Wolf, F. (2006). Semiparametric accelerated failure time regression analysis with application to interval-censored HIV/AIDS data. *Statistics in Medicine* **25,** 3850–3863.

Yan, J. (2002). geepack: Yet another package for generalized estimating equations. *R-News* pages 12–14.

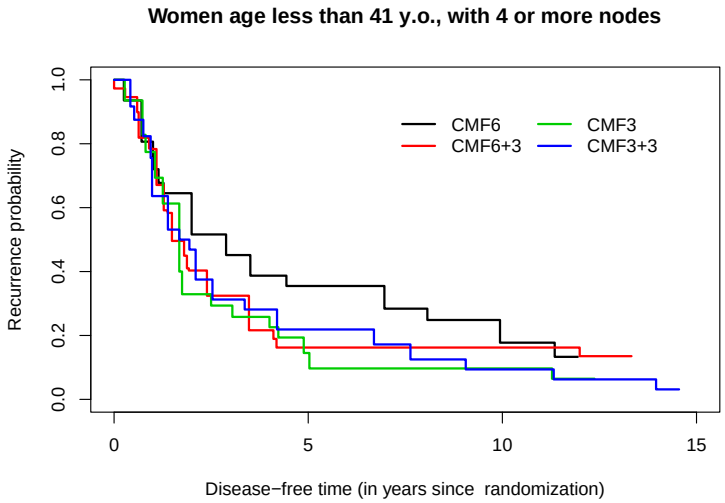

Figure 1. IBCSG trial VI: time to breast cancer recurrence estimates, by treatment group, in patients with four or more nodes and age $\leqslant 40$ years at baseline.

Table 1

Proportional Hazards model, Scenarios A ($h_0(t) = t^3$) and B ($h_0(t) = t^4$): $n = 100$, $\beta_1 = -0.5$, $\beta_2 = 1.0$, $h_0(t)$ =baseline hazard, $\hat{\beta}_i$= empirical mean β_i estimates, SE= empirical mean of $\hat{\beta}_i$ estimated standard errors, ESE= empirical standard error of $\hat{\beta}_i$ estimates, 95.CP= coverage probability of true β_i by the 95% confidence intervals, $i = 1, 2$.

Scenario	Method(J)[a]	$\hat{\beta}_1$	SE	ESE	95.CP	$\hat{\beta}_2$	SE	ESE	95.CP
A	PO-TB(1)	-0.52	0.83	0.92	0.95	1.05	0.79	0.89	0.93
	PO-EMICM(1)	-0.50	0.76	0.81	0.95	1.04	0.76	0.83	0.95
	PO-EMICM(3)	-0.52	0.61	0.62	0.95	1.01	0.58	0.62	0.94
	COX-MPI	-0.42	0.21	0.22	0.92	0.83	0.23	0.22	0.85
B	PO-TB(1)	-0.55	0.82	0.87	0.95	1.05	0.76	0.87	0.94
	PO-EMICM(1)	-0.51	0.76	0.81	0.95	1.04	0.76	0.82	0.96
	PO-EMICM(3)	-0.52	0.61	0.63	0.94	1.01	0.57	0.61	0.93
	COX-MPI	-0.41	0.21	0.22	0.91	0.82	0.22	0.22	0.84

[a] J indicates the number of points $t_1 < \cdots < t_J$ used

Table 2

Proportional Odds model, Scenarios A ($\beta_1 = 1$, $\beta_2 = 2$) and B ($\beta_1 = 1$, $\beta_2 = 1$) : n=100, $\hat{\beta}_i$= empirical mean β_i estimates, SE= empirical mean of $\hat{\beta}_i$ estimated standard errors, ESE= empirical standard error of $\hat{\beta}_i$ estimates, $95.CP$= coverage probability of true β_i by the 95% confidence intervals, $i = 1, 2$.

Scenario	Method(J)[b]	$\hat{\beta}_1$	SE	ESE	95.CP	$\hat{\beta}_2$	SE	ESE	95.CP
A	PO-TB(1)	0.99	1.29	1.44	0.95	2.02	1.26	1.39	0.94
	PO-EMICM(1)	1.00	1.23	1.32	0.95	1.96	1.19	1.34	0.95
	PO-EMICM(3)	0.99	0.95	1.01	0.94	2.00	0.96	0.94	0.93
	MS-MPI	0.92	0.61	0.61	0.94	1.83	0.63	0.64	0.93
B	PO-TB(1)	1.00	1.23	1.33	0.94	1.04	0.74	0.80	0.95
	PO-EMICM(1)	0.97	1.16	1.25	0.94	1.03	0.69	0.73	0.95
	PO-EMICM(3)	0.98	0.91	0.94	0.94	1.00	0.53	0.54	0.95
	MS-MPI	0.94	0.61	0.61	0.94	0.94	0.36	0.36	0.94

[b] J indicates the number of points $t_1 < \cdots < t_J$ used

Table 3

Accelerated Failure Time model, Scenarios A ($\beta_1 = 1$, $\beta_2 = -0.5$) and B ($\beta_1 = 1$, $\beta_2 = 0.5$): n=100,$\hat{\beta}_i$= empirical mean β_i estimates, SE= empirical mean of $\hat{\beta}_i$ estimated standard errors, ESE= empirical standard error of $\hat{\beta}_i$ estimates, 95.CP= coverage probability of true β_i by the 95% confidence intervals, $i = 1, 2$.

Scenario	Method(J)[c]	$\hat{\beta}_1$	SE	ESE	CP	$\hat{\beta}_2$	SE	ESE	95.CP
A	PO-TB(1)	1.00	0.81	0.89	0.95	-0.49	0.47	0.53	0.94
	PO-EMICM(1)	1.01	0.74	0.79	0.95	-0.47	0.42	0.45	0.95
	PO-EMICM(3)	1.00	0.57	0.60	0.95	-0.49	0.33	0.35	0.95
	AFT-MPI	0.90	0.21	0.21	0.92	-0.45	0.21	0.22	0.99
B	PO-TB(1)	1.01	0.81	0.93	0.94	0.53	0.48	0.54	0.93
	PO-EMICM(1)	1.02	0.74	0.84	0.94	0.52	0.44	0.47	0.94
	PO-EMICM(3)	1.00	0.57	0.59	0.94	0.49	0.34	0.34	0.94
	AFT-MPI	0.90	0.21	0.20	0.91	0.46	0.21	0.21	1.00

[c] J indicates the number of points $t_1 < \cdots < t_J$ used

Table 4
The International Breast Cancer Study Group Trial VI example: patients with four or more nodes and age $\leqslant$ 40 years at baseline.

Model type	Unadjusted			Adjusted		
Interval-Censored (PO-EMICM(1))	HR	95 % CI	P	HR	95 % CI	P
CMF6	1.00	reference		1.00	reference	
CMF6+3	1.22	0.64-2.33	0.56	1.62	0.90-2.91	0.11
CMF3	1.65	0.93-2.91	0.09	2.18	1.26-3.78	0.006
CMF3+3	1.48	0.84-2.61	0.18	1.25	0.59-2.62	0.57
Interval-Censored (PO-EMICM(3))						
CMF6	1.00	reference		1.00	reference	
CMF6+3	1.60	0.82-3.11	0.169	1.89	1.04-3.44	0.036
CMF3	1.67	0.85-3.30	0.138	2.37	1.23-4.58	0.009
CMF3+3	1.24	0.61-2.53	0.546	1.15	0.53-2.52	0.714
Right-Censored (Cox Model MPI)						
CMF6	1.00	reference		1.00	reference	
CMF6+3	1.32	0.79-2.22	0.287	1.65	0.86-3.16	0.129
CMF3	1.49	0.89-2.54	0.142	1.93	0.99-3.75	0.054
CMF3+3	1.43	0.85-2.40	0.184	1.05	0.54-2.05	0.894