Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 409-428

A PROPORTIONAL HAZARDS MODEL FOR BIVARIATE SURVIVAL DATA UNDER INTERVAL CENSORING

By HSIAO-CHUAN TIEN
and
PRANAB KUMAR SEN
University of North Carolina at Chapel Hill, USA

SUMMARY. Bivariate failure time models are often encountered in medical research, where each subject under study may experience two types of events or there exists a pairing. There are problems in measuring the association between variables when interval censoring occurs. A proportional hazards model is proposed to analyse such data. Maximum partial likelihood estimators are obtained and their large-sample properties are studied.

1. **Introduction**

In medical research when there are multiple end points, a subject under study may experience multiple types of events or there may be some natural or artificial pairing (ties) of such events. For such multivariate survival analysis (especially for the bivariate case), analytic methods have been developed quite extensively; we refer to Clegg et al. (2000) where other references are cited. Usually, there is some incompleteness in the available data sets due to lack of compliance or withdrawal (drop out) and other forms of truncation of the study. Such censoring may occur due to a planned time period of the trial (known as Type I censoring), a termination following a specified number of failures (Type II censoring), loss to follow-up (random censoring), and other reasons. It is a common practice that observations are recorded in class intervals, so that exact response time may not be known but it is known to have occurred within some time intervals (which need not be of equal width). In the contemporary literature, this is referred to as interval censoring. In clinical trials or longitudinal studies, interval censoring occurs naturally when response times are recorded in periodic follow-ups. For ex-

Paper received October 2000; revised February 2001.
AMS (2000) subject classification. Primary 62N01; secondary 62H12.
Keywords and phrases. Bivariate survival function; censoring, counting processes, semi-parametric models.

ample in pregnancy studies, the time of conception can only be recorded up to a menstrual period.

The focus of this research is on the analysis of bivariate survival data when interval censoring is present on one or both failure times, which are generally not (statistically) independent. When the bivariate model is fully parameterized, covariates can be introduced into the model by allowing the marginal distributions to be functions of the covariates. Aitken and Clayton (1980) parameterized the marginal as exponential, Weibull, and extreme values distributions. Huster, Brookmeyer, and Self (1989) extended the fully parametric model of Clayton (1978) for analysing paired failure time regression data with right censoring to incorporate the covariates. The estimates of the parameters are found by maximizing the likelihood using Newton-Raphson methods. Huster, Brookmeyer, and Self (1989) proposed an independence working model to treat the dependence within a pair as a nuisance parameter. The model leads to inconsistent estimates of the marginal parameters and variances of the estimates when there is dependence between the pair. Wei, Lin and Weissfeld (1989) focused only on the marginal distributions of the failure times and ignored the dependence structure between the failure times. Frydman (1994), Alioum and Commenges (1996) modified Turnbull's (1976) method to accommodate both truncation and censoring. But their methods only focus on univariate outcome. There is a need for development of regression models to accommodate interval censoring in a multivariate setting.

In the present study, a semiparametric method is proposed to incorporate bivariate survival outcomes with interval censoring. A bivariate extension of the Cox (1972) proportional hazards model is used to incorporate covariates wherein the bivariate baseline hazards functions are adapted from Dabrowska (1988). Our proposed approaches are presented in the next section and are illustrated in Section 5 with a data set from Demographic and Health Survey. Theoretical results on consistency and asymptotic normality are outlined in Section 3, and test statistics are introduced in Section 4. Some concluding remarks are appended in Section 6.

The Models

For modeling bivariate survival outcomes, Dabrowska's (1988) bivariate hazards functions are incorporated wherein stochastic dependence within subjects is retained. Further, the Cox (1972) idea of modeling univariate discrete hazards function is extended to cover bivariate discrete hazards functions that are adapted to some continuous versions.

1.1 *Bivariate hazards function.* For a pair of nonnegative random variables X_1, X_2, let $S_{X_1,X_2}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$ be the joint survival function. Following Dabrowska (1988), bivariate conditional hazards functions are defined as

$$\lambda_{12}(x_1, x_2) = \frac{f(x_1, x_2)}{S(x_1-, x_2-)}, \tag{1}$$

$$\lambda_{10}(x_1, x_2) = \frac{\int_0^\infty I\{v > x_2\} f(x_1, v) dv}{S(x_1-, x_2)}, \tag{2}$$

$$\lambda_{20}(x_1, x_2) = \frac{\int_0^\infty I\{u > x_1\} f(u, x_2) du}{S(x_1, x_2-)}, \tag{3}$$

where $I(A)$ stands for the indicator function of a set A. Here, $\lambda_{12}(x_1, x_2)$ is the instantaneous rate of a *simultaneous failure* at times (x_1, x_2), given that the individuals were alive at times (x_1-, x_2-); $\lambda_{10}(x_1, x_2)$ represents the rate of only a *single failure* at time x_1 given that the first individual was alive at time x_1- and second individual survived beyond time x_2. The meaning of $\lambda_{20}(x_1, x_2)$ is analogous.

The hazard functions corresponding to bivariate outcomes can be expressed in terms of the vector

$$\boldsymbol{\lambda}(x_1, x_2) = \Big(\lambda_{12}(x_1, x_2), \lambda_{10}(x_1, x_2), \lambda_{20}(x_1, x_2)\Big)'.$$

In line with the Cox (1972) proportional hazards assumption, it is assumed that in models with covariates $\boldsymbol{z}$, the bivariate hazard functions have a semi-parametric structure as follows:

$$\lambda_{12}(x_1, x_2 \mid \boldsymbol{z}) = \lambda_{12}(x_1, x_2 \mid \mathbf{0}) \exp(\boldsymbol{\gamma}'\boldsymbol{z}(x_1, x_2)), \tag{4}$$

$$\lambda_{10}(x_1, x_2 \mid \boldsymbol{z}_1) = \lambda_{10}(x_1, x_2 \mid \mathbf{0}) \exp(\boldsymbol{\beta}_1'\boldsymbol{z}_1(x_1)), \tag{5}$$

$$\lambda_{20}(x_1, x_2 \mid \boldsymbol{z}_2) = \lambda_{20}(x_1, x_2 \mid \mathbf{0}) \exp(\boldsymbol{\beta}_2'\boldsymbol{z}_2(x_2)), \tag{6}$$

where $\boldsymbol{z}_k' = (z_{k1}, \ldots, z_{kp})$ is the covariates associated with X_k only and $\beta_k' = (\beta_{k1}, \ldots, \beta_{kp})$ is the corresponding regression parameters, $k = 1, 2$, $\boldsymbol{\gamma}_k = (\gamma_{k1}, \ldots, \gamma_{kp})$, and $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')$, $\boldsymbol{z}' = (\boldsymbol{z}_1', \boldsymbol{z}_2')$. The $\lambda_{12}(x_1, x_2 \mid \mathbf{0})$, $\lambda_{10}(x_1, x_2 \mid \mathbf{0})$, and $\lambda_{20}(x_1, x_2 \mid \mathbf{0})$ denote the baseline hazards rates, which are defined as (1), (2), and (3).

1.2 *Definition of Risk Sets and Conditional Hazards Model.* When the survival time is recorded in interval grouping, we have a discrete survival data. Cox (1972), Kalbfleisch and Prentice (1980) considered discrete proportional hazards model in the univariate case. Their idea is extended here to the bivariate case.

In the presence of censoring, let $T_1 = \min(X_1, C_1)$ and $T_2 = \min(X_2, C_2)$ where (C_1, C_2) are potential right censoring time for (X_1, X_2). The failure times and the censoring times are assumed to be independent. Let $\Delta_k = 1$ if $T_k = X_k$, and 0 otherwise. Suppose that the continuous-time bivariate proportional hazards models in (4), (5), and (6) hold but that the observed outcomes are grouped into intervals, $c_{j_1-1} < X_1 \leq c_{j_1}$, $d_{j_2-1} < X_2 \leq d_{j_2}$, $j_1 = 1, \ldots, l$, $j_2 = 1, \ldots, m$. The lengths of intervals vary from subject to subject. Here, we assume the censoring intervals are non-overlapping. That is $0 = c_0 < c_1 \leq \ldots \leq c_l \leq c_{l+1} = \infty$ and $0 = d_0 < d_1 \leq \ldots \leq d_m < d_{m+1} = \infty$. We further assume that the l and m are finite; they may be allowed to increase with n but at a rate slower than $n^{1/2}$.

When the two response variables are comparable we might want to compare the results at the same time. Here we proceed to compare along the diagonal line, and define $R(t) = \{l : T_{1l} > t-, T_{2l} > t-\}$ be the set of subjects at risk just prior to time t. Then it is possible to have simultaneous failures at time t under our model. That is, we allow the same censoring time on both variables. It implies that $l = m$; $c_j = d_j$, for $j = 1, \ldots, m$.

Andersen (1988) treated the problem of interval censoring by point estimating and then treated the estimated values as observed times. We propose to use Andersen's (1988) idea when interval censored data occur. Consider choosing

$$X_k^* = a_{kj}, \quad \text{if } c_{j-1} < X_k \leq c_j, \quad j = 1, \ldots, m, \ k = 1, 2.$$

Under the assumption of finite first order moments for both X_1 and X_2, the a_{kj}s are chosen as

$$a_{kj} = \frac{\displaystyle\int_{c_{j-1}}^{c_j} x dF_{X_k}(x)}{F_{X_k}(c_j) - F_{X_k}(c_{j-1})}.$$

The $F_{X_k}(.)$s are the marginal distribution functions for X_ks.

Define $T = \min(T_1, T_2)$, and $\delta_k(t) = 1$ if t is from T_k, and 0 otherwise, $k = 1, 2$. Note here $\delta_1(T) = \delta_2(T) = 1$ when $T_1 = T_2 = T$. Let $t_{(1)} < t_{(2)} < \ldots < t_{(h)}$ denote the distinct ordered failure times. For notational convenience the parentheses for the ordered statistics are suppressed. Further note that these t_g's $(g = 1, \ldots, h)$ relate to the class intervals where the actual failures preceding censoring occur, so that let n_{kg}, $k = 1, 2$ be the multiplicity of failure at time t_g with respect to X_k only, and n_{12g} be the multiplicity of failure at time t_g with respect to both X_1 and X_2. Here $\tilde{n} = \sum_{g=1}^{h}(n_{1g} + n_{2g} + n_{12g})$ is the actual number of failures preceding censoring occurred.

1.3 *Partial likelihood function.* The bivariate conditional hazards model that we propose to use under interval censoring involves

$$\lambda_{k0}(t, t \mid \boldsymbol{z}_k) = \lambda_{k0}(t, t \mid \boldsymbol{0}) \exp(\boldsymbol{\beta}_k^{'} \boldsymbol{z}_k), \ k = 1, 2, \tag{7}$$

$$\lambda_{12}(t, t \mid \boldsymbol{z}) = \lambda_{12}(t, t \mid \boldsymbol{0}) \exp(\boldsymbol{\gamma}^{'} \boldsymbol{z}), \tag{8}$$

here $\lambda_{12}(t, t) > 0$. The conditional hazards function in (7) considers the covariates corresponding to X_k only. The model may not be appropriate for certain studies. Like in twin or mother-daughter studies, the hazards for twin 1 (daughter) could be affected by the covariates for twin 2 (mother). Under this circumstance, it would be appropriate to change (7) into

$$\lambda_{k0}(t, t \mid \boldsymbol{z}) = \lambda_{k0}(t, t \mid \boldsymbol{0}) \exp(\boldsymbol{\beta}_k^{'} \boldsymbol{z}).$$

The partial likelihood for the individuals failing at t_i under our model is proportional to

$$\prod_{k=1}^{2} \left\{ \exp\left[\boldsymbol{\beta}_k^{'} (\sum_{j=1}^{n_{ki}} \boldsymbol{z}_{kj})\right] \Big/ \sum_{l \in R(t_i)} \exp\left[\boldsymbol{\beta}_k^{'} (\sum_{g=1}^{n_{kl}} \boldsymbol{z}_{g_{ki}})\right] \right\}^{(\delta_{ki} - \delta_{1i}\delta_{2i})\Delta_{ki}}$$
$$\times \left\{ \exp\left[\boldsymbol{\gamma}^{'} (\sum_{j=1}^{n_{12i}} \boldsymbol{z}_j)\right] \Big/ \sum_{l \in R(t_i)} \exp\left[\boldsymbol{\gamma}^{'} (\sum_{g=1}^{n_{12l}} \boldsymbol{z}_{g_i})\right] \right\}^{\delta_{1i}\delta_{2i}\Delta_{1i}\Delta_{2i}} \tag{9}$$

where $\sum_{j=1}^{n_{ki}} \boldsymbol{z}_{kj}$ is the sum of the covariate vectors $\boldsymbol{z}_k$ over all the distinct individuals failing at t_i in X_k only and $\sum_{j=1}^{n_{12i}} \boldsymbol{z}_j$ is the sum of $\boldsymbol{z}$ over all the distinct individuals failing at t_i in both X_1 and X_2. The above partial likelihood actually involves three mutually exclusive components: the subset with failures only in X_1 but not in X_2, the subset with failures only in X_2 but not in X_1, and the subset with failures in both X_1 and X_2. Let A_{1i}, A_{2i} and A_{12i} represent the subsets respectively. The analysis of the unknown parameters will mainly be based on the three disjoint sets. Then the logarithm of the Cox partial likelihood under our bivariate hazards model can be expressed as

$$\sum_{i=1}^{n} \left\{ \sum_{k=1}^{2} (\delta_{ki} - \delta_{1i}\delta_{2i}) \, \Delta_{ki} \boldsymbol{\beta}_k^{'} \boldsymbol{z}_{A_{ki}} \right.$$
$$- \sum_{k=1}^{2} (\delta_{ki} - \delta_{1i}\delta_{2i}) \, \Delta_{ki} \log \Big(\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}_k^{'} \boldsymbol{z}_{A_{kj}}) \Big)$$
$$\left. + \delta_{1i}\delta_{2i}\Delta_{1i}\Delta_{2i} \boldsymbol{\gamma}^{'} \boldsymbol{z}_{A_{12i}} - \delta_{1i}\delta_{2i}\Delta_{1i}\Delta_{2i} \log \Big(\sum_{j \in R(t_i)} \exp(\boldsymbol{\gamma}^{'} \boldsymbol{z}_{A_{12j}}) \Big) \right\}, \tag{10}$$

where

$$\boldsymbol{z}_{A_{ki}} = \sum_{j \in A_{ki}} \boldsymbol{z}_j, \; k = 1, 2, \qquad \boldsymbol{z}_{A_{12i}} = \begin{pmatrix} \boldsymbol{z}_{A_{(1)i}} \\ \boldsymbol{z}_{A_{(2)i}} \end{pmatrix} = \sum_{j \in A_{12i}} \boldsymbol{z}_j.$$

REMARK. The $\lambda_{k0}(t,t)$ here does not represent the marginal hazards function in Wei, Lin and Weissfeld (1989). Further, in our model, the covariates are assumed to be time independent. Extension of the model to incorporate time-dependent covariate effects is possible, albeit at considerable notational complexities.

1.4 *Counting process formulation.* We denote the three counting processes associated at time point t by $N_{ki}(t,t) = I\{T_{ki} \le t,\ T_{(3-k)i} > t,\ \delta_{ki} = 1,\ \delta_{(3-k)i} = 0,\ \Delta_{ki} = 1\}$, $k = 1,2$, and $N_{12i}(t,t) = I\{T_{1i} \le t,\ T_{2i} \le t,\ \delta_{1i} = 1,\ \delta_{2i} = 1,\ \Delta_{1i} = 1,\ \Delta_{2i} = 1\}$, and the *at risk processes* by $Y_{ki}(t,t) = I\{T_{ki} \ge t,\ T_{(3-k)i} > t\}$, $k = 1,2$, and $Y_{12i}(t,t) = I\{T_{1i} \ge t,\ T_{2i} \ge t\}$. The logarithm of the partial likelihood in (10) has the form

$$\begin{aligned} l(\boldsymbol{\theta};t) = \sum_{i=1}^{n} \Bigg\{ & \int_0^t \boldsymbol{\beta}_1' \boldsymbol{z}_{A_{1i}}(u) dN_{1i}(u,t) \\ & - \int_0^t \log\Big\{ \sum_{j=1}^{n} Y_{kj}(u,t) \exp\left[\boldsymbol{\beta}_1' \boldsymbol{z}_{A_{1j}}(u)\right]\Big\} dN_{1i}(u,t) \\ & + \int_0^t \boldsymbol{\beta}_2' \boldsymbol{z}_{A_{2i}}(u) dN_{2i}(t,u) \\ & - \int_0^t \log\Big\{ \sum_{j=1}^{n} Y_{2j}(t,u) \exp\left[\boldsymbol{\beta}_2' \boldsymbol{z}_{A_{2j}}(u)\right]\Big\} dN_{2i}(t,u) \\ & + \int_0^t\int_0^t \boldsymbol{\gamma}' \boldsymbol{z}_{A_{12i}}(u_1,u_2) dN_{12i}(u_1,u_2) \\ & - \int_0^t\int_0^t \log\Big\{ \sum_{j=1}^{n} Y_{12j}(u_1,u_2) \exp\left[\boldsymbol{\gamma}' \boldsymbol{z}_{A_{12j}}(u_1,u_2)\right]\Big\} dN_{12i}(u_1,u_2) \Bigg\}, \end{aligned} \tag{11}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$. It is convenient to introduce the following notations

$$S_k^{(r)}(\boldsymbol{\beta}_k, t) = \frac{1}{n} \sum_{j=1}^{n} Y_{kj}(t,t) \boldsymbol{z}_{A_{kj}}(t)^{\otimes r} \exp\left[\boldsymbol{\beta}_k' \boldsymbol{z}_{A_{kj}}(t)\right], \; r = 0,1,2,$$

$$S_{12}^{(0)}(\boldsymbol{\gamma}, t, t) = \frac{1}{n} \sum_{j=1}^{n} Y_{12j}(t,t) \exp\left[\boldsymbol{\gamma}' \boldsymbol{z}_{A_{12j}}(t,t)\right],$$

$$S_{12}^{(r)}(\boldsymbol{\gamma},t,t)=\frac{1}{n}\sum_{j=1}^{n}Y_{12j}(t,t)\boldsymbol{z}_{A_{12j}}(t,t)^{\otimes r}\exp\left[\boldsymbol{\gamma}'\boldsymbol{z}_{A_{12j}}(t,t)\right],\ r=0,1,2,$$

$$E_k(\boldsymbol{\beta}_k,t)=\frac{S_k^{(1)}(\boldsymbol{\beta}_k,t)}{S_k^{(0)}(\boldsymbol{\beta}_k,t)},\quad E_{12}(\boldsymbol{\gamma},t,t)=\frac{S_{12}^{(1)}(\boldsymbol{\gamma},t,t)}{S_{12}^{(0)}(\boldsymbol{\gamma},t,t)},$$

$$V_k(\boldsymbol{\beta}_k,t)=\frac{S_k^{(2)}(\boldsymbol{\beta}_k,t)}{S_k^{(0)}(\boldsymbol{\beta}_k,t)}-E_k(\boldsymbol{\beta}_k,t)^{\otimes 2},$$

$$V_{12}(\boldsymbol{\gamma},t,t)=\frac{S_{12}^{(2)}(\boldsymbol{\gamma},t,t)}{S_{12}^{(0)}(\boldsymbol{\gamma},t,t)}-E_{12}(\boldsymbol{\gamma},t,t)^{\otimes 2},$$

where $k=1,2$, $\boldsymbol{a}^{\otimes 2}=\boldsymbol{a}\boldsymbol{a}'$. The lowercase s,e, and v are the expectations of the uppercase S,E, and V, respectively. The score functions corresponding to (11) are

$$\boldsymbol{U}(\boldsymbol{\theta};t)=\begin{pmatrix}\boldsymbol{U}_1(\boldsymbol{\beta}_1;t)\\ \boldsymbol{U}_2(\boldsymbol{\beta}_2;t)\\ \boldsymbol{U}_3(\boldsymbol{\gamma};t)\end{pmatrix},$$

where $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are $p\times 1$ vector functions of the unknown parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ respectively, and $\boldsymbol{U}_3$ is the $2p\times 1$ vector functions of the unknown parameters $\boldsymbol{\gamma}$. Specifically

$$\begin{aligned}
\boldsymbol{U}_1(\boldsymbol{\beta}_1,t) &= \sum_{i=1}^{n}\left\{\int_0^t \boldsymbol{z}_{A_{1i}}(u)dN_{1i}(u,t)-\int_0^t\frac{S_1^{(1)}(\boldsymbol{\beta}_1,u)}{S_1^{(0)}(\boldsymbol{\beta}_1,u)}dN_{1i}(u,t)\right\}\\
\boldsymbol{U}_2(\boldsymbol{\beta}_2,t) &= \sum_{i=1}^{n}\left\{\int_0^t \boldsymbol{z}_{A_{2i}}(u)dN_{2i}(t,u)-\int_0^t\frac{S_2^{(1)}(\boldsymbol{\beta}_2,u)}{S_2^{(0)}(\boldsymbol{\beta}_2,u)}dN_{2i}(t,u)\right\},\\
\boldsymbol{U}_3(\boldsymbol{\gamma},t) &= \sum_{i=1}^{n}\left\{\int_0^t\int_0^t \boldsymbol{z}_{A_{12i}}(u_1,u_2)dN_{12i}(u_1,u_2)\right.\\
&\qquad\left.-\int_0^t\int_0^t\frac{S_{12}^{(1)}(\boldsymbol{\gamma},u_1,u_2)}{S_{12}^{(0)}(\boldsymbol{\gamma},u_1,u_2)}dN_{12i}(u_1,u_2)\right\}.
\end{aligned}$$

Define the M_{1i}, M_{2i}, and M_{12i} as

$$M_{1i}(t,t)=N_{1i}(t,t)-\int_0^t Y_{1i}(u,t)\lambda_{10i}(u,t)du, \tag{12}$$

$$M_{2i}(t,t)=N_{2i}(t,t)-\int_0^t Y_{2i}(t,u)\lambda_{20i}(t,u)du, \tag{13}$$

$$M_{12i}(t,t)=N_{12i}(t,t)-\int_0^t\int_0^t Y_{12i}(u_1,u_2)\lambda_{12i}(u_1,u_2)du_1du_2, \tag{14}$$

The M_1, M_2 and M_{12} are all zero mean martingales. M_k and M_{12} are orthogonal martingales for $k = 1, 2$. It follows immediately that

$$\boldsymbol{U}_1(\boldsymbol{\beta}_1, t) = \sum_{i=1}^{n} \left\{ \int_0^t \boldsymbol{z}_{A_{1i}}(u) dM_{1i}(u,t) - \int_0^t \frac{S_1^{(1)}(\boldsymbol{\beta}_1, u)}{S_1^{(0)}(\boldsymbol{\beta}_1, u)} dM_{1i}(u,t) \right\}, \quad (15)$$

$$\boldsymbol{U}_2(\boldsymbol{\beta}_2, t) = \sum_{i=1}^{n} \left\{ \int_0^t \boldsymbol{z}_{A_{2i}}(u) dM_{2i}(t,u) - \int_0^t \frac{S_2^{(1)}(\boldsymbol{\beta}_2, u)}{S_2^{(0)}(\boldsymbol{\beta}_2, u)} dM_{2i}(t,u) \right\}, \quad (16)$$

$$\boldsymbol{U}_3(\boldsymbol{\gamma}, t) = \sum_{i=1}^{n} \left\{ \int_0^t \int_0^t \boldsymbol{z}_{A_{12i}}(u_1, u_2) dM_{12i}(u) - \int_0^t \int_0^t \frac{S_{12}^{(1)}(\boldsymbol{\gamma}, u_1, u_2)}{S_{12}^{(0)}(\boldsymbol{\gamma}, u_1, u_2)} dM_{12i}(u_1, u_2) \right\}. \quad (17)$$

The maximum partial likelihood estimators $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ are defined as the solution to $\boldsymbol{U}(\widehat{\boldsymbol{\theta}}, t) = \mathbf{0}$.

2. Main Results

First, we formulate appropriate regularity assumptions under which we will present the main theoretical results pertaining to proposed model; these conditions are similar to those used by Andersen and Gill (1982) for the univariate case with some modifications to suit the bivariate case. The results are presented with the continuous time interval $(0, \tau)$ in mind, where $\tau (0 < \tau < \infty)$ is a terminal time.

C1. $\lambda_{k0}(t,t) \geq 0$, $k = 1, 2$, $\lambda_{12}(t,t) \geq 0$. $\int_0^\tau \lambda_{10}(t,\tau)dt$, $\int_0^\tau \lambda_{20}(\tau,t)dt$, and $\int_0^\tau \int_0^\tau \lambda_{12}(t_1,t_2)dt_1 dt_2$ are all finite.

C2. There exists a neighbourhood $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ and functions $s_k^{(r)}(\boldsymbol{\beta}_k, t)$, and $s_{12}^{(r)}(\boldsymbol{\gamma}, t_1, t_2)$, $r = 1, 2$, $k = 1, 2$, defined on $\boldsymbol{\Theta} \times [0, \tau]$ such that

$$\sup_{t \in [0,\tau], \boldsymbol{\beta}_k \in \boldsymbol{\Theta}} \left\| S_k^{(r)}(\boldsymbol{\beta}_k, t) - s_k^{(r)}(\boldsymbol{\beta}_k, t) \right\| \xrightarrow{P} 0, \ r = 0, 1, 2, \ k = 1, 2,$$

$$\sup_{t_1 \in [0,\tau] t_2 \in [0,\tau], \boldsymbol{\gamma} \in \boldsymbol{\Theta}} \left\| S_{12}^{(r)}(\boldsymbol{\gamma}, t_1, t_2) - s_{12}^{(r)}(\boldsymbol{\gamma}, t_1, t_2) \right\| \xrightarrow{P} 0, \ r = 0, 1, 2.$$

C3. For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}, t \in [0, \tau]$:

$$\frac{\partial}{\partial \boldsymbol{\beta}_k} s_k^{(0)}(\boldsymbol{\beta}_k, t) = s_k^{(1)}(\boldsymbol{\beta}_k, t), \quad \frac{\partial}{\partial \boldsymbol{\beta}_k'} s_k^{(1)}(\boldsymbol{\beta}_k, t) = s_k^{(2)}(\boldsymbol{\beta}_k, t),$$

$$\frac{\partial}{\partial \boldsymbol{\gamma}} s_{12}^{(0)}(\boldsymbol{\gamma}, t, t) = s_{12}^{(1)}(\boldsymbol{\gamma}, t, t), \quad \frac{\partial}{\partial \boldsymbol{\gamma}'} s_{12}^{(1)}(\boldsymbol{\gamma}, t, t) = s_{12}^{(2)}(\boldsymbol{\gamma}, t, t).$$

$s_k^{(r)}(\boldsymbol{\beta}_k,t)$, and $s_{12}^{(r)}(\boldsymbol{\gamma},t,t)$, $k=1,2$, $r=0,1,2$, are continuous functions of $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ uniformly in $t\in[0,\tau]$ and all are bounded on $\boldsymbol{\Theta}\times[0,\tau]$ and $s_k^{(0)}(\boldsymbol{\beta}_k,t)$, and $s_{12}^{(0)}(\boldsymbol{\gamma},t,t)$ are bounded away from zero on $\boldsymbol{\Theta}\times[0,\tau]$.

C4. The $4p\times 4p$ matrix

$$\boldsymbol{I}(\boldsymbol{\theta}_0)=\begin{pmatrix}\boldsymbol{I}_1(\boldsymbol{\beta}_{10}), & 0, & 0\\ 0, & \boldsymbol{I}_2(\boldsymbol{\beta}_{20}), & 0\\ 0, & 0, & \boldsymbol{I}_3(\boldsymbol{\gamma}_0)\end{pmatrix}, \tag{18}$$

is positive definite, where

$$\begin{aligned}\boldsymbol{I}_1(\boldsymbol{\beta}_{10}) &= \int_0^\tau \left\{v_1(\boldsymbol{\beta}_{10},u)s_1^{(0)}(\boldsymbol{\beta}_{10},u)\lambda_{10}(u,\tau)\right\}du, \text{ is a } p\times p \text{ matrix},\\ \boldsymbol{I}_2(\boldsymbol{\beta}_{20}) &= \int_0^\tau \left\{v_2(\boldsymbol{\beta}_{20},u)s_2^{(0)}(\boldsymbol{\beta}_{20},u)\lambda_{20}(\tau,u)\right\}du, \text{ is a } p\times p \text{ matrix},\\ \boldsymbol{I}_3(\boldsymbol{\gamma}_0) &= \int_0^\tau\int_0^\tau \{v_{12}(\boldsymbol{\gamma}_0,u_1,u_2)\}\, s_{12}^{(0)}(\boldsymbol{\gamma}_0,u_1,u_2)\lambda_{12}(u_1,u_2)du_1du_2,\\ &\qquad \text{is a } 2p\times 2p \text{ matrix}.\end{aligned}$$

C5. (Variance-covariance stability). There exists a $4p\times 4p$ matrix $\boldsymbol{\Sigma}$ such that when $n\to\infty$,

$$\frac{1}{n}\mathrm{Var}(\boldsymbol{G}) \xrightarrow{P} \boldsymbol{\Sigma},$$

where

$$\boldsymbol{G}=\begin{pmatrix}\boldsymbol{G}_1\\ \boldsymbol{G}_2\\ \boldsymbol{G}_{12}\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^n\boldsymbol{G}_{1i}\\ \sum_{i=1}^n\boldsymbol{G}_{2i}\\ \sum_{i=1}^n\boldsymbol{G}_{12i}\end{pmatrix}.$$

with

$$\begin{aligned}\boldsymbol{G}_{1i}(\boldsymbol{\beta}_1) &= \int_0^\tau\left(\boldsymbol{z}_{A_{1i}}(u)-\frac{s_1^{(1)}(\boldsymbol{\beta}_1,u)}{s_1^{(0)}(\boldsymbol{\beta}_1,u)}\right)dM_{1i}(u,\tau), \text{ a } (p\times 1) \text{ vector},\\ \boldsymbol{G}_{2i}(\boldsymbol{\beta}_2) &= \int_0^\tau\left(\boldsymbol{z}_{A_{2i}}(u)-\frac{s_2^{(1)}(\boldsymbol{\beta}_2,u)}{s_2^{(0)}(\boldsymbol{\beta}_2,u)}\right)dM_{2i}(\tau,u), \text{ a } (p\times 1) \text{ vector},\\ \boldsymbol{G}_{12i}(\boldsymbol{\gamma}) &= \int_0^\tau\int_0^\tau\left(\boldsymbol{z}_{A_{12}}(u_1,u_2)-\frac{s_{12}^{(1)}(\boldsymbol{\gamma},u_1,u_2)}{s_{12}^{(0)}(\boldsymbol{\gamma},u_1,u_2)}\right)dM_{12i}(u_1,u_2),\\ &\qquad \text{a } (2p\times 1) \text{ vector}.\end{aligned}$$

REMARK. The conditions that we use here only holds in the interval $\mathcal{I} = [0, \tau)$, for some fixed $\tau \in [0, \infty]$. For the extension of the results to $[0, \infty)$, we need to show the influence from (τ, ∞) can be made arbitrarily small, uniformly in n, by choosing τ large enough, see Gill (1980).

THEOREM 2.1. *The processes given by* $M_{1j}(t,t) = N_{1j}(t,t) - B_{1j}(t)$, $M_{2j}(t,t) = N_{2j}(t,t) - B_{2j}(t)$, $M_{12j}(t,t) = N_{12j}(t,t) - B_{12j}(t)$, *with*

$$\begin{aligned} B_{1j}(t) &= \int_0^t Y_{1j}(u,t)\lambda_{10j}(u,t)du, \\ B_{2j}(t) &= \int_0^t Y_{2j}(t,u)\lambda_{20j}(t,u)du, \\ \textit{and } B_{12j}(t) &= \int_0^t \int_0^t Y_{12j}(u_1,u_2)\lambda_{12j}(u_1,u_2)du_1du_2, \end{aligned}$$

are zero mean martingales with respect to $\mathcal{F}_t$, *a filtration defined on a common probability space, if* $EB_{kj}(t) < \infty$, $k = 1, 2$, *and* $EB_{12j}(t) < \infty$ *for all* $t > 0$, $\forall j$.

PROOF. Since we only focus on a single time parameter t here. $M_{1j}(t,t)$, $M_{2j}(t,t)$, and $M_{12j}(t,t)$ can be rewritten as $M_{1j}(t)$, $M_{2j}(t)$, and $M_{12j}(t)$, and $N_{1j}(t,t)$, $N_{2j}(t,t)$, and $N_{12j}(t,t)$ can be rewritten as $N_{1j}(t)$, $N_{2j}(t)$, and $N_{12j}(t)$. Let τ_n denote a localizing sequence for the local martingale $M = N - B$, so $M(. \wedge \tau_n)$ is a martingale for any n. Since $M(0) = 0$ a.s.

$$EM_{1j}(t \wedge \tau_n) = E\{EM_{1j}(t \wedge \tau_n) \mid \mathcal{F}_0\} = E\{M_{1j}(0)\} = 0$$

for any t, and thus, $EN_{1j}(t \wedge \tau_n) = EB_{1j}(t \wedge \tau_n)$ for any t and n. By the Monotone Convergence Theorem,

$$EN_{1j}(t) = \lim_{n\to\infty} E\{N_{1j}(t \wedge \tau_n)\} = \lim_{n\to\infty} E\{B_{1j}(t \wedge \tau_n)\} = EB_{1j}(t) \qquad (19)$$

for any $t \geq 0$.

To see that $E|M_{1j}(t)| < \infty$ for all $t > 0$, note that

$$E|M_{1j}(t)| \leq E\left|N_{1j}(t)\right| + \int_0^t\int_0^t P\{T_{1j} \leq u_1, T_{2j} > u_2\}\lambda_{10j}(u_1,u_2)du_1du_2 \leq 2.$$

Since $N_{1j}(t)$ is a counting process, then the expectation of $M_{1j}(t)$ is zero based on (19). Since $EN_{1j}(t) < \infty$, then $EB_{1j}(t) < \infty$, for all t. Therefore $M_{1j}(t)$ is a zero mean martingale. The proofs for $M_2(t)$ and $M_{12}(t)$ are similar. □

COROLLARY 2.1 *The* M_1, M_2, *and* M_{12} *defined in (12), (13), and (14) are orthogonal martingales.*

PROOF. Recall M_1, and M_{12} in (12), and (14).

$$\begin{aligned}\langle M_1, M_{12}\rangle &= E\{M_1 M_{12}\} - E\{M_1\}E\{M_{12}\}\\ &= E\left\{\Big(N_1(t,t) - \int_0^t Y_1(u,t)\lambda_1(u,t)du\Big)\right.\\ &\quad \left.\cdot \Big(N_{12}(t,t) - \int_0^t\int_0^t Y_{12}(u_1,u_2)\lambda_{12}(u_1,u_2)du_1du_2\Big)\right\}. \quad (20)\end{aligned}$$

Based on the definitions of N_1, N_{12}, Y_1, Y_{12}, the expectation in (20) equals zero. That is, $\langle M_1, M_{12}\rangle = 0$. By Corollary 1.4.3 in Fleming and Harrington (1991), M_1, and M_{12} are orthogonal. Similarly, we can prove $\langle M_1, M_2\rangle = 0$, $\langle M_2, M_{12}\rangle = 0$. Again, by Corollary 1.4.3 in Fleming and Harrington (1991), M_1, M_2, and M_{12} are orthogonal. □

The following theorems are obtained under our proposed model.

THEOREM 2.2 *Under Conditions C1-C5,* $\widehat{\boldsymbol{\theta}}$ *converges to a well-defined constant vector* $\boldsymbol{\theta}_0$.

THEOREM 2.3 *Under Conditions C1-C5,* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *is asymptotically normal with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{V} = \boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{I}^{-1}(\boldsymbol{\theta})$.

THEOREM 2.4 *Under Conditions C1-C5, the covariance matrix* $\boldsymbol{V}$ *can be consistently estimated by* $\widehat{\boldsymbol{I}}_n^{-1}(\widehat{\boldsymbol{\theta}})\ \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})\ \widehat{\boldsymbol{I}}_n^{-1}(\widehat{\boldsymbol{\theta}})$. *where*

$$\widehat{\boldsymbol{I}}_n(\widehat{\boldsymbol{\theta}}) = -\frac{1}{n}\frac{\partial \boldsymbol{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \quad \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{G}}'(\widehat{\boldsymbol{\theta}}),$$

and $\widehat{\boldsymbol{G}}$ *is the consistent estimator of* $\boldsymbol{G}$ *with* S *substituted for* s.

REMARK. As in section (1.2), $\tilde{n}$ is the total number of distinct failure times preceding censoring occurs, which is a stochastic process. Assume $\lim_{n\to\infty}\frac{\tilde{n}}{n} = \pi$, where π is the probability of failure before censoring, $0 < \pi < 1$. The asymptotic results in Theorems 2.2 and 2.3 may not hold when π closing to zero which corresponds to heavy censoring.

3. Test Statistics

In this section, we generalize the standard test statistics to interval censored bivariate failure time data under our proposed model. Suppose we wish to test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, vs. $H_1 : \theta_j \neq \theta_{j0}$, for some j. The test statistic for testing H_0 vs. H_1 based on the partial likelihood is:

- Likelihood ratio statistic:

$$Q_L = -2\log\Lambda = 2\{l(\widehat{\boldsymbol{\theta}};\tau) - l(\boldsymbol{\theta}_0;\tau)\}.$$

We construct score and Wald's statistics using the asymptotic normality of $\widehat{\boldsymbol{\theta}}$.

- Rao's efficient score statistic:

$$Q_R = n\boldsymbol{U}'(\boldsymbol{\theta}_0)\widehat{\boldsymbol{I}}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{U}(\boldsymbol{\theta}_0)$$

- Wald's statistic:

$$Q_W = n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\widehat{\boldsymbol{I}}_n(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{I}}_n(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

The asymptotic properties of these three statistics are given in the following theorem.

THEOREM 3.1. *Under H_0, each Q_L, Q_R, and Q_W has an asymptotic χ^2 distribution with $4p$ degrees of freedom.*

4. Example

We illustrate the proposed methodology in an application with a data set of a follow-up study of 22909 Indonesian women of ages between 15 to 49 years in a Demographic and Health Survey (DHS). The collected data set records the information that described the timing of the woman's experience with pregnancy, childbearing, contraception use, the duration of breastfeeding, postpartum amenorrhea and other related experience for 69 months. A total of 5037 women without using any contraceptive methods from January 1986 to September 1991 are available to analyse the association between duration of post-partum amenorrhea and breastfeeding. Women who do not breastfeed are amenorrheic for an average of 1.5-2.0 months after a birth, while the mean length of amenorrhea may exceed 20 months in populations with extensive breastfeeding customs (Potter and Kobrin, 1981). All the women were married, aged 15-49 and have 1-16 children. Duration of breastfeeding and duration of post-partum amenorrhea were recorded as discrete outcomes (months). There was a very high proportion of right censored observations. The primary cause of right censoring was that both activities were still ongoing at the time of last interview. Sixty percent of the women were still breastfeeding and thirty three percent of the women were still amenorrheic at the time of last interview.

The range of the duration of breastfeeding is from 0 month to 59 months and the range of the duration of post-partum amenorrhea is from 0 month to 58 months. We treat the data as bivariate discrete outcome. In this study, we include two covariates in the bivariate hazards model: the logarithm of mother's age and the birth order of the child. We believe the birth order of the child has influence on the duration of breastfeeding and the logarithm of mother's age has influence on the duration of post-partum amenorrhea .

First we fit the data with our proposed model (model 1):

$$\lambda_{k0}(t,t \mid \boldsymbol{z}) = \lambda_{k0}(t,t \mid \mathbf{0}) \exp(\beta_k z_k),\ k = 1,2,$$
$$\lambda_{12}(t,t \mid \boldsymbol{z}) = \lambda_{12}(t,t \mid \mathbf{0}) \exp(\gamma_1 z_1 + \gamma_2 z_2),$$

where Z_1 is the birth order of the child and Z_2 is the logarithm of mother's age. Here, β_1 is the effect of the birth order of the child on the duration of breastfeeding conditioned on post-partum amenorrhea, β_2 is the effect of the logarithm of mother's age on the duration of post-partum amenorrhea conditioned on breastfeeding, γ_1 is the effect of the birth order of the child on both the duration of breastfeeding and the duration of post-partum amenorrhea and γ_2 is the effect of the logarithm of mother's age on both the duration of breastfeeding and the duration of post-partum amenorrhea. Further, we perform the following modeling (model 2):

$$\lambda_{k0}(t,t \mid \boldsymbol{z}) = \lambda_{k0}(t,t \mid \mathbf{0}) \exp(\beta_k z_k),\ k = 1,2,$$
$$\lambda_{12}(t,t \mid \boldsymbol{z}) = \lambda_{12}(t,t \mid \mathbf{0}) \exp(\beta_1 z_1 + \beta_2 z_2).$$

Their estimates are listed in table 1. Under our approach, there are 36 distinct observed failure times. The actual number of failures preceding censoring occurred is 3427. Among those, 171 women stopped breastfeeding before their post-partum amenorrhea, 3073 women had post-partum amenorrhea before the end of breastfeeding process, and 183 women had both events happened at the same time.

TABLE 1. REGRESSION COEFFICIENT ESTIMATES WITH BIVARIATE HAZARDS MODEL.

	Model 1		Model 2	
	$\widehat{\theta}$	(SE)	$\widehat{\theta}$	(SE)
birth order (β_1)	–0.0052	(3.8×10^{-6})	–0.0073	(3.0×10^{-3})
log(age) (β_2)	0.0043	(3.8×10^{-6})	0.0043	(5.9×10^{-4})
birth order (γ_1)	–0.0313	(1.1×10^{-3})		
log(age) (γ_2)	0.0266	(1.4×10^{-3})		

The hypothesis of interest is whether there are effect of child's birth order on breastfeeding and effect of mother's age on post-partum amenorrhea. The likelihood ratio statistic is 52.6 from model 1 and 51.2 from model 2, which are both significant compared with chi-square distribution with 4 and

2 degrees of freedom. The Rao's statistic and Wald's statistic give similar results. The estimates of coefficients from model 1 suggest that conditional on post-partum amenorrhea the duration of breastfeeding seemed to increase in accordance with the number of births. Mother's age could reduce the duration of post-partum amenorrhea under the condition of breastfeeding.

4. Discussion

Our proposed bivariate hazards functions in (7) and (8) only model the hazards at time (t, t) as a failure in one variable eliminates the unit as a whole. The advantage of our approach is that it is easy to define the corresponding risk sets, and apply manageable statistical analysis schemes. It is possible to treat the general case (t_1, t_2), with $t_1 \neq t_2$. Further extensions to examine failures at different time (t_1, t_2) will be considered. Definition of the risk sets related to the above bivariate hazards function is needed. Previous procedures can be applied, and similar theoretical work can be carried out to study the behaviour of the large sample properties of the estimators.

When the bivariate failures observes as the same time intervals, the probability for both failures happen at the same time t is zero. This implies that (8) equals zero. Under this circumstance, the partial likelihood would only include the hazards functions from (7). The partial likelihood in (9) can be further simplified. The λ_{k0}s that we propose here are different from the marginal hazards function. Even when $\lambda_{12}(t, t) = 0$, the partial likelihood function is not equivalent to the partial likelihood function based on marginal hazards approach.

The asymptotic results in this study only apply to non-overlapping intervals. More studies are needed to develop methodology on overlapping intervals. The length of the observed interval might influence the effect of the estimation. Further numerical study would provide a way to investigate possible patterns based on the length of observed interval.

Appendix

Proof of theorem **(2.2)** The proof for the consistency of $\widehat{\boldsymbol{\theta}}$ follows similar techniques in Lemma 3.1 of Andersen and Gill (1982) with some modification to consider bivariate setup. Consider the process

$$D(\boldsymbol{\theta}, \tau) = \frac{1}{n}(l(\boldsymbol{\theta}, \tau) - l(\boldsymbol{\theta}_0, \tau))$$

$$= \frac{1}{n}\sum_{i=1}^{n}\Big\{\Big(\int_0^\tau(\boldsymbol{\beta}_1-\boldsymbol{\beta}_{10})'\boldsymbol{z}_{A_{1i}}(u)dN_{1i}(u,\tau)-\int_0^\tau\log\frac{S_1^{(0)}(\boldsymbol{\beta}_1,u)}{S_1^{(0)}(\boldsymbol{\beta}_{10},u)}dN_{1i}(u,\tau)\Big)$$
$$+\Big(\int_0^\tau(\boldsymbol{\beta}_2-\boldsymbol{\beta}_{20})'\boldsymbol{z}_{A_{2i}}(u)dN_{2i}(\tau,u)-\int_0^\tau\log\frac{S_2^{(0)}(\boldsymbol{\beta}_2,u)}{S_2^{(0)}(\boldsymbol{\beta}_{20},u)}dN_{2i}(\tau,u)\Big)$$
$$+\Big(\int_0^\tau\int_0^\tau(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\boldsymbol{z}_{A_{12i}}(u_1,u_2)dN_{12i}(u_1,u_2)$$
$$-\int_0^\tau\int_0^\tau\log\frac{S_{12}^{(0)}(\boldsymbol{\gamma},u_1,u_2)}{S_{12}^{(0)}(\boldsymbol{\gamma}_0,u_1,u_2)}dN_{12i}(u_1,u_2)\Big)\Big\}.$$

Under conditions C1-C3, we can show that $D(\boldsymbol{\theta},\tau)$ converges in probability to the same limit as

$$B(\boldsymbol{\theta},\tau)=\frac{1}{n}\sum_{i=1}^{n}\Big\{\Big(\int_0^\tau(\boldsymbol{\beta}_1-\boldsymbol{\beta}_{10})'\boldsymbol{z}_{A_{1i}}(u)dN_{1i}(u,\tau)$$
$$-\int_0^\tau\log\frac{s_1^{(0)}(\boldsymbol{\beta}_1,u)}{s_1^{(0)}(\boldsymbol{\beta}_{10},u)}dN_{1i}(u,\tau)\Big)$$
$$+\Big(\int_0^\tau(\boldsymbol{\beta}_2-\boldsymbol{\beta}_{20})'\boldsymbol{z}_{A_{2i}}(u)dN_{2i}(\tau,u)-\int_0^\tau\log\frac{s_2^{(0)}(\boldsymbol{\beta}_2,u)}{s_2^{(0)}(\boldsymbol{\beta}_{20},u)}dN_{2i}(\tau,u)\Big)$$
$$+\Big(\int_0^\tau\int_0^\tau(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\boldsymbol{z}_{A_{12i}}(u_1,u_2)dN_{12i}(u_1,u_2)$$
$$-\int_0^\tau\int_0^\tau\log\frac{s_{12}^{(0)}(\boldsymbol{\gamma},u_1,u_2)}{s_{12}^{(0)}(\boldsymbol{\gamma}_0,u_1,u_2)}dN_{12i}(u_1,u_2)\Big)\Big\}.$$

for each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. This indicates that $D(\boldsymbol{\theta},\tau)$ is asymptotically equivalent to $B(\boldsymbol{\theta},\tau)$ as $n\to\infty$.

Under conditions C1-C3, for each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, the compensator of $B(\boldsymbol{\theta},\tau)$ converges in probability to

$$b(\boldsymbol{\theta},\tau)=\int_0^\tau\Big\{(\boldsymbol{\beta}_1-\boldsymbol{\beta}_{10})'s_1^{(1)}(\boldsymbol{\beta}_{10},u)$$
$$-s_1^{(0)}(\boldsymbol{\beta}_{10},u)\log\frac{s_1^{(0)}(\boldsymbol{\beta}_1,u)}{s_1^{(0)}(\boldsymbol{\beta}_{10},u)}\Big\}\lambda_{10}(u,\tau)du$$
$$+\int_0^\tau\Big\{(\boldsymbol{\beta}_2-\boldsymbol{\beta}_{20})'s_2^{(2)}(\boldsymbol{\beta}_{20},u)-s_2^{(0)}(\boldsymbol{\beta}_{20},u)\log\frac{s_2^{(0)}(\boldsymbol{\beta}_2,u)}{s_2^{(0)}(\boldsymbol{\beta}_{20},u)}\Big\}\lambda_{20}(\tau,u)du$$
$$+\int_0^\tau\int_0^\tau\Big\{(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'s_{12}^{(1)}(\boldsymbol{\gamma}_0,u_1,u_2)$$

$$-s_{12}^{(0)}(\boldsymbol{\gamma}_0, u_1, u_2) \log \frac{s_{12}^{(0)}(\boldsymbol{\gamma}, u_1, u_2)}{s_{12}^{(0)}(\boldsymbol{\gamma}_0, u_1, u_2)} \Bigg\} \lambda_{12}(u_1, u_2) du_1 du_2.$$

For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\begin{aligned}
\frac{\partial b(\boldsymbol{\theta}, \tau)}{\partial \boldsymbol{\beta}_1} &= \int_0^\tau \left\{ \frac{s_1^{(1)}(\boldsymbol{\beta}_{10}, u)}{s_1^{(0)}(\boldsymbol{\beta}_{10}, u)} - \frac{s_1^{(1)}(\boldsymbol{\beta}_1, u)}{s_1^{(0)}(\boldsymbol{\beta}_1, u)} \right\} s_1^{(0)}(\boldsymbol{\beta}_{10}, u) \lambda_{10}(u, \tau) du, \\
\frac{\partial b(\boldsymbol{\theta}, \tau)}{\partial \boldsymbol{\beta}_2} &= \int_0^\tau \left\{ \frac{s_2^{(1)}(\boldsymbol{\beta}_{20}, u)}{s_2^{(0)}(\boldsymbol{\beta}_{20}, u)} - \frac{s_2^{(1)}(\boldsymbol{\beta}_2, u)}{s_2^{(0)}(\boldsymbol{\beta}_2, u)} \right\} s_2^{(0)}(\boldsymbol{\beta}_{20}, u) \lambda_{20}(\tau, u) du, \\
\frac{\partial b(\boldsymbol{\theta}, \tau)}{\partial \boldsymbol{\gamma}} &= \int_0^\tau \int_0^\tau \left\{ \frac{s_{12}^{(1)}(\boldsymbol{\gamma}_0, u_1, u_2)}{s_{12}^{(0)}(\boldsymbol{\gamma}_0, u_1, u_2)} - \frac{s_{12}^{(1)}(\boldsymbol{\gamma}, u_1, u_2)}{s_{12}^{(0)}(\boldsymbol{\gamma}, u_1, u_2)} \right\} s_{12}^{(0)}(\boldsymbol{\gamma}_0, u_1, u_2) \\
&\quad \lambda_{12}(u_1, u_2) du_1 du_2,
\end{aligned}$$

are zero at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Furthermore

$$\begin{aligned}
\frac{\partial^2 b(\boldsymbol{\theta}, \tau)}{\partial \boldsymbol{\beta}_1^2} &= -\int_0^t v_1(\boldsymbol{\beta}_1, u) s_1^{(0)}(\boldsymbol{\beta}_{10}, u) \lambda_{10}(u, \tau) du, \\
\frac{\partial^2 b(\boldsymbol{\theta}, \tau)}{\partial \boldsymbol{\beta}_2^2} &= -\int_0^t v_2(\boldsymbol{\beta}_2, u) s_2^{(0)}(\boldsymbol{\beta}_{20}, u) \lambda_{20}(\tau, u) du, \\
\frac{\partial^2 b(\boldsymbol{\theta}, \tau)}{\partial \boldsymbol{\gamma}^2} &= -\int_0^\tau \int_0^\tau v_{12}(\boldsymbol{\gamma}, u_1, u_2) s_{12}^{(0)}(\boldsymbol{\gamma}_0, u_1, u_2) \lambda_{12}(u_1, u_2) du_1 du_2
\end{aligned}$$

which, under condition C4, are negative definite at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Therefore, we can claim that for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $D(\boldsymbol{\theta}, \tau)$ converges in probability to a concave function $b(\boldsymbol{\theta}, \tau)$ of $\boldsymbol{\theta}$ which has a unique maximizer at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. It follows by the theorem II.1 in Andersen and Gill (1982), the maximizing value $\widehat{\boldsymbol{\theta}}$ of $D(\boldsymbol{\theta}, \tau)$ converges in probability to the maximizing value $\boldsymbol{\theta}_0$ of $b(\boldsymbol{\theta}, \tau)$. That is $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ as $n \to \infty$. □

Proof of theorems (2.3) and (2.4). First we are going to show that each element of $\widehat{\boldsymbol{\theta}}$ is asymptotic normal distributed. For each element in $\boldsymbol{U}(\boldsymbol{\theta}, \tau)$, let $U_{1_r}(\boldsymbol{\beta}_1, \tau)$ be the rth component in (15), $r = 1, \ldots, p$. Applying the Taylor expansion to $U_{1_r}(\boldsymbol{\beta}_1, \tau)$ around $\beta_{1_{r0}}$, we get

$$\sqrt{n}(\widehat{\beta}_{1_r} - \beta_{1_{r0}}) = \left(-\frac{1}{n} \frac{\partial U_{1_r}(\boldsymbol{\beta}_1, \tau)}{\partial \beta_{1_r}} \Bigg|_{\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*} \right)^{-1} \frac{1}{\sqrt{n}} U_{1_r}(\boldsymbol{\beta}_{10}, \tau),$$

where $\boldsymbol{\beta}^*$ is on the line of segment between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Based on conditions C2 and C3, we can show that $\frac{1}{\sqrt{n}} U_{1_r}(\boldsymbol{\beta}_{10}, \tau)$ is asymptotic equivalent to

$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{1_{ri}}(\boldsymbol{\beta}_{10},\tau)$, where $G_{1_{ri}}$ is the r-th element in $\boldsymbol{G}_{1_i}$. With condition C5 and martingale central limit theorem, we can claim that the local martingale $\frac{1}{\sqrt{n}}U_{1_r}(\boldsymbol{\beta}_{10},\tau)$ converges to a Gaussian process. Based on conditions C1-C4, we can claim that $-\frac{1}{n}\frac{\partial U_{1_r}(\boldsymbol{\beta}_1,\tau)}{\partial \beta_{1_r}}\Big|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*}$ converges in probability to a nonzero variance function for any random $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1(n)$ such that $\boldsymbol{\beta}_1^* \xrightarrow{P} \boldsymbol{\beta}_{10}$ as $n\to\infty$.

Therefore, we can claim that $\sqrt{n}(\widehat{\beta}_{1_r}-\beta_{1_{r0}})$ is asymptotically normal. Similar, we can show the asymptotic normality for $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{2_r}-\boldsymbol{\beta}_{2_{r0}})$ and $\sqrt{n}(\widehat{\gamma}_m-\gamma_{m0})$, where β_{2_r} is the rth element in $\boldsymbol{\beta}_2$ and γ_m is the mth element in $\boldsymbol{\gamma}$. The linear combination of the elements in $\boldsymbol{U}(\boldsymbol{\theta}_0,t)$ is asymptotic normal distributed. By Cramer-Wold device from lemma 5.2.1 in Fleming and Harrington (1991), and condition C5, we have

$$\frac{1}{\sqrt{n}}\boldsymbol{U}(\boldsymbol{\theta}_0,\tau) \xrightarrow{P} N_{4p}(\mathbf{0},\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)), \tag{A.1}$$

where N_{4p} is a multivariate normal distribution. Since $\frac{1}{\sqrt{n}}\boldsymbol{U}(\boldsymbol{\theta}_0,\tau)$ is asymptotic equivalent to $\frac{1}{\sqrt{n}}\boldsymbol{G}(\boldsymbol{\theta}_0)$, with (A.1) condition C5, and theorem (2.2) the $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be consistently estimated by $\frac{1}{n}\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{G}}'(\widehat{\boldsymbol{\theta}})$.

Let

$$\begin{aligned}
\boldsymbol{I}_{1n}(\boldsymbol{\beta}_1) &= -\frac{1}{n}\frac{\partial \boldsymbol{U}(\boldsymbol{\beta}_1)}{\partial \boldsymbol{\beta}_1'} = \frac{1}{n}\int_0^\tau V_1(\boldsymbol{\beta}_1,t)dN_{1.}(t,\tau),\\
\boldsymbol{I}_{2n}(\boldsymbol{\beta}_2) &= -\frac{1}{n}\frac{\partial \boldsymbol{U}(\boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2'} = \frac{1}{n}\int_0^\tau V_2(\boldsymbol{\beta}_2,t)dN_{2.}(\tau,t),\\
\boldsymbol{I}_{3n}(\boldsymbol{\gamma}) &= -\frac{1}{n}\frac{\partial \boldsymbol{U}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'} = \frac{1}{n}\int_0^\tau\int_0^\tau V_{12}(\boldsymbol{\gamma},t_1,t_2)dN_{12.}(t_1,t_2),
\end{aligned}$$

where $N_{k.}=\sum_{i=1}^{n} N_{ki}, k=1,2$, and $N_{12.}=\sum_{i=1}^{n} N_{12i}$. The difference between $\boldsymbol{I}_{1n}(\beta_1)|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*}$ and $\boldsymbol{I}_1(\boldsymbol{\beta}_{10})$ can be expressed as

$$\begin{aligned}
&\left\|\frac{1}{n}\int_0^\tau V_1(\boldsymbol{\beta}_1^*,t)dN_{1.}(t,\tau)-\int_0^\tau\left\{v_1(\boldsymbol{\beta}_{10},t)s_1^{(0)}(\boldsymbol{\beta}_{10},t)\lambda_{10}(t,\tau)\right\}dt\right\|\\
&\quad\le \left\|\int_0^\tau\left\{V_1(\boldsymbol{\beta}_1^*;t)-v_1(\boldsymbol{\beta}_1;t)\right\}\frac{dN_{1.}(t,\tau)}{n}\right\| &\text{(A.2)}\\
&\qquad+\left\|\int_0^\tau\left\{v_1(\boldsymbol{\beta}_1^*;t)-v_1(\boldsymbol{\beta}_{10};t)\right\}\frac{dN_{1.}(t,\tau)}{n}\right\| &\text{(A.3)}
\end{aligned}$$

$$+\left\|\int_0^\tau v_1(\boldsymbol{\beta}_{10};t)\Big\{\frac{dN_{1\cdot}(t,\tau)}{n}-\frac{\sum_{i=1}^n Y_{1i}(t)\lambda_{10i}(t,\tau)}{n}dt\Big\}\right\| \quad \text{(A.4)}$$

$$+\left\|\int_0^\tau v_1(\boldsymbol{\beta}_{10};t)\Big\{S_1^{(0)}(\boldsymbol{\beta}_{10},t)-s_1^{(0)}(\boldsymbol{\beta}_{10},t)\Big\}\lambda_{10}(t,\tau)dt\right\|. \quad \text{(A.5)}$$

The hazards function has finite interval and is bounded, we can get for all $\eta > 0$

$$\lim_{\eta\uparrow\infty}\lim_{n\to\infty}P\{\frac{N_{1\cdot}(\tau,\tau)}{n}\geq\eta\}=0. \quad \text{(A.6)}$$

Based on conditions C2 and C3, we have

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}_1\in\boldsymbol{\Theta}}\left\|\int_0^\tau\Big\{V_{1_{(r,r)}}(\boldsymbol{\beta}_1^*;t)-v_{1_{(r,r)}}(\boldsymbol{\beta}_1^*;t)\Big\}\frac{dN_{1\cdot}(t,\tau)}{n}\right\|\stackrel{P}{\longrightarrow}0.$$

When $\boldsymbol{\beta}_1^*\stackrel{P}{\to}\boldsymbol{\beta}_{10}$ and (A.6) plus the above result indicate that (A.2) converges in probability to zero. In condition C3, the continuity in $\boldsymbol{\beta}_1$, uniformly in t, together with (A.6) imply that (A.3) converges to zero as well. Using the Lemma 8.2.1 in Fleming and Harrington (1991) on (A.4), for all $\rho,\eta>0$ and any $t\geq 0$,

$$\begin{aligned}P&\left\{\left|\int_0^\tau v_1(\boldsymbol{\beta}_{10};t)\Big\{\frac{dN_{1\cdot}(t,\tau)}{n}-\frac{\sum_{i=1}^n Y_{1i}(t)\lambda_{10i}(t,\tau)}{n}dt\Big\}\right|>\eta\right\}\\ &\leq\frac{\rho}{\eta^2}+P\left\{\int_0^\tau v_{1_{(r,r)}}(\boldsymbol{\beta}_{10};t)^{\otimes 2}S_1^{(0)}(\boldsymbol{\beta}_{10};t)\lambda_{10}(t,\tau)dt>\rho\right\}.\end{aligned}$$

Conditions C1-C3 indicate the left hand side of the above inequality converges in probability to zero. The (A.4) vanishes. Applying conditions C1, C2 to the last term, we have (A.5) disappear.

Therefore, when $\boldsymbol{\beta}_1^*\stackrel{P}{\to}\boldsymbol{\beta}_{10}$, we have

$$\boldsymbol{I}_{n1}(\boldsymbol{\beta}_1;\tau)\Big|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*}\stackrel{P}{\longrightarrow}\boldsymbol{I}_1(\boldsymbol{\beta}_{10})\ \text{ as }\ n\longrightarrow\infty.$$

Following similar techniques, we also have when $\boldsymbol{\beta}_2^*\stackrel{P}{\to}\boldsymbol{\beta}_{20}$ ($\boldsymbol{\beta}_2^*$ within the line segment joining $\widehat{\boldsymbol{\beta}}_2$ and $\boldsymbol{\beta}_{20}$)

$$\boldsymbol{I}_{n2}(\boldsymbol{\beta}_2;\tau)\Big|_{\boldsymbol{\beta}_2=\boldsymbol{\beta}_2^*}\stackrel{P}{\longrightarrow}\boldsymbol{I}_2(\boldsymbol{\beta}_{20})\ \text{ as }\ n\longrightarrow\infty.$$

and when $\boldsymbol{\gamma}^*\stackrel{P}{\to}\boldsymbol{\gamma}$ ($\boldsymbol{\gamma}^*$ within the line segment joining $\widehat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}_0$)

$$\boldsymbol{I}_{n3}(\boldsymbol{\gamma};\tau)\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^*}\stackrel{P}{\longrightarrow}\boldsymbol{I}_3(\boldsymbol{\gamma}_0)\ \text{ as }\ n\longrightarrow\infty.$$

It implies that, for $\boldsymbol{\theta}^*$ sitting on the line of segment connecting $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$,

$$\boldsymbol{I}_n(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \xrightarrow{P} \boldsymbol{I}(\boldsymbol{\theta}_0), \text{ as } \quad n \longrightarrow \infty. \tag{A.7}$$

Based on condition C4, the above result and theorem (2.2), the consistent estimator for $\boldsymbol{I}(\boldsymbol{\theta}_0)$ is $\widehat{\boldsymbol{I}}_n(\widehat{\boldsymbol{\theta}})$. When $\boldsymbol{\theta}^* \xrightarrow{P} \boldsymbol{\theta}_0$ as $n \to \infty$, the local martingale $\frac{1}{\sqrt{n}}\boldsymbol{U}(\boldsymbol{\theta}_0, t)$ converges to a Gaussian process and $\frac{1}{n}\boldsymbol{I}_n(\boldsymbol{\theta})$ converges in probability to a nonsingular matrix. The asymptotic normality of $\widehat{\boldsymbol{\theta}}$ is obvious. The asymptotic covariance $\boldsymbol{V} = \boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{I}^{-1}(\boldsymbol{\theta})$ can be consistently estimated by $\widehat{\boldsymbol{I}}_n^{-1}(\widehat{\boldsymbol{\theta}})\ \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}})\ \widehat{\boldsymbol{I}}_n^{-1}(\widehat{\boldsymbol{\theta}})$. □

PROOF OF THEOREM (3.1.) First consider the following Taylor expansion of $l(\boldsymbol{\theta})$ along on $\boldsymbol{\theta}_0$:

$$\begin{aligned} l(\boldsymbol{\theta};\tau) \quad &= \quad l(\boldsymbol{\theta}_0;\tau) + \sqrt{n}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\frac{\boldsymbol{U}(\boldsymbol{\theta}_0;\tau)}{\sqrt{n}} \\ &\quad -\frac{n}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\left\{\boldsymbol{I}_n(\boldsymbol{\theta};\tau)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0), \end{aligned} \tag{A.8}$$

where $\boldsymbol{\theta}^*$ is on the line of segment between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Then consider the following Taylor expansion of $\boldsymbol{U}(\boldsymbol{\theta}_0;\tau)$:

$$\frac{1}{\sqrt{n}}\boldsymbol{U}(\boldsymbol{\theta}_0;\tau) = \left\{\boldsymbol{I}_n(\boldsymbol{\theta};\tau)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{**}}\right\}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$$

where $\boldsymbol{\theta}^{**}$ belongs to the line segment joining $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. When $\boldsymbol{\theta}^* \xrightarrow{P} \boldsymbol{\theta}_0$ and $\boldsymbol{\theta}^{**} \xrightarrow{P} \boldsymbol{\theta}_0$, under H_0, we can show that

$$\begin{aligned} \boldsymbol{I}_n(\boldsymbol{\theta};\tau)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} &\xrightarrow{P} \boldsymbol{I}(\boldsymbol{\theta}_0;\tau), \\ \boldsymbol{I}_n(\boldsymbol{\theta};\tau)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{**}} &\xrightarrow{P} \boldsymbol{I}(\boldsymbol{\theta}_0;\tau). \end{aligned}$$

Therefore, we can rewrite (A.8) as

$$2\{l(\widehat{\boldsymbol{\theta}};\tau) - l(\boldsymbol{\theta}_0;\tau)\} = n(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\boldsymbol{I}(\boldsymbol{\theta}_0;\tau)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + o_p(1).$$

By Slutsky's and Cochran's theorems, $Q_L \xrightarrow{D} \chi^2_{4p}$. Similarly, the asymptotic distributions for Q_R and Q_W can be obtained. □

Acknowledgments. The authors are grateful to the referees for their most helpful comments on the manuscript. The authors would also like to thank Dr. Chirayath Suchindran for providing the data.

References

Aitkin, M. and Clayton, D. (1980). The fitting of exponential, Weibull, and extreme value distributions to complex censored survival data using GLIM. *Appl. Statist.*, **29**, 159-163.

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

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model and counting processes: a large sample study. *Ann. Statist.*, **10(4)**, 1100-1120.

Andersen, P.K. (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Stat. Med.*, **7**, 661-670.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141-151.

Clegg, L.X., Cai, J. and Sen, P.K. (2000). Modeling multivariate failure time data. In *Handbook of Statistics, Volume 18: Bioenvironmental and Public Health Statistics.* P.K. Sen and C.R. Rao, eds., Elsevier, Amsterdam, 803-838.

Cox, D.R. (1972). Regression models and life-tables (with discussion) *J. Roy. Statist. Soc. Ser. B.*, **34**, 187-220.

Dabrowska, D.M. (1988). Kaplan-Meier estimate on the plane, *Ann. Statist.*, **16**, 1475-1489.

Fleming, T.R. and Harrington, D.P. (1991). *Counting processes and survival analysis.* Wiley, New York.

Frydman, H. (1994). A note on nonparametric estimation of the distribution function from interval- censored and truncated observations. *J. Roy. Statist. Soc. B.*, **56**, 71-74.

Gill, R.D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tracts 124.

Huster, W.J., Brookmeyer, R. and Self, S.G. (1989). Modelling paired survival data with covariates. *Biometrics*, **45**, 145-156.

Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Potter, R.G. and Kobrin, F.E. (1981). Distributions of amenorrhea and anovulation. *Population Studies*, **35**, 85-94.

Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped censored and truncated data. *J. Roy. Statist. Soc. B.*, **38**, 290-296.

Wei, L.J. and Lin, D.Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. *J. Amer. Statist. Assoc.*, **84**, 1065-1073.

Hsiao-Chuan Tien
Department of Biostatistics
School of Public Health
University of North Carolina
Chapel Hill, NC 27599-7420, USA
E-mail: htien@bios.unc.edu

Pranab Kumar Sen
Department of Biostatistics
School of Public Health
University of North Carolina
Chapel Hill, NC 27599-7420, USA
E-mail: pksen@bios.unc.edu