Regression Analysis for Marginal Failure Probability under Competing Risks

Wei-Hwa Chang
Weijing Wang
Email: wjwang@stat.nctu.edu.tw
Institute of Statistics, National Chiao-Tung University
Hsin-Chu, Taiwan, R.O.C.
December, 2006

Under Revision by Statistica Sinica

Abstract. The cumulative incidence function provides intuitive summary information about competing risks data. Via a mixture decomposition of this function, we study covariate effects on the marginal probability of developing a particular cause of failure at a given time point without specifying the corresponding failure time distribution. Several inference methods are constructed based on two principles of handling missing data due to censoring, namely the imputation approach and the weighting approach. Our work extends part of the results in Fine (1999) and Wang (2003). Large sample properties of the proposed estimators are derived and their finite sample performances are examined via simulations. For illustrative purposes, the proposed methods are applied to the well-known Heart Transplant data and is compared with the analysis of Larson and Dinse (1985) which was developed under the framework of a mixture model.

Key words : Cause-specific hazard; Mixture model; Imputation; Cumulative incidence function; Inverse probability of censoring; Logistic regression; Missing Data.

1 Introduction

Multiple events data are commonly seen in biomedical studies. Under the framework of competing risks, subjects may fail from one of several different causes. Let T be the failure time and $\tilde{B}$ be the corresponding cause of failure taking values from the set $\{1, \ldots, J\}$. Competing risks data naturally can be summarized by the following two quantities. One is the cause-specific hazard function defined as

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t + \Delta t), \tilde{B} = j | T \geq t)}{\Delta t}$$

which is the rate of occurrence for type-j failure in presence of all causes of failure. The other is the cumulative incidence function, or the crude failure probability, defined as

$$F_j(t) = \Pr(T \leq t, \tilde{B} = j),$$

which describes the cumulative probability of developing type-j failure by time t. This paper considers regression analysis of $F_j(t)$ which provides more direct information about the cumulative risks and hence can be more easily explained to clinicians.

There is a functional relationship between the cause-specific hazard function and the cumulative incidence function such that

$$F_j(t) = \int_0^t S(u-)\lambda_j(u)du,$$

and $S(t) = \Pr(T > t) = \exp(-\int_0^t \sum_{j=1}^J \lambda_j(u)du)$. Some authors including Cheng, Fine and Wei (1998) have suggested to make inference of the cumulative incidence function by specifying regression models for all causes of hazards. However, since the effect of a covariate on $\lambda_j(t)$ can be very different from its effect on $F_j(t)$, the conclusion based on such an indirect approach may be misleading. Also the parameters in the models for cause-specific hazards may lack simple interpretation in terms of the crude failure probabilities.

There have been growing interests in directly modelling the cumulative incidence function for a particular cause. Suppose that the first type of failure is of main interest. Fine and Gray (1999) and Fine (2001) both considered semi-parametric regression models of the form,

$$g(F_1(t|\mathbf{z})) = h(t) + \mathbf{z}^T\boldsymbol{\theta}, \tag{1}$$

where $\mathbf{z}$ is a $p \times 1$ vector of covariates, $g(\cdot)$ is a link function mapping from $(0, 1)$ to $(-\infty, \infty)$ and $h(t)$ is an unknown monotone function. Fine and Gray (1999) studied estimation in this model with the complementary log-log link function which leads to a proportional hazards assumption on the improper variable T_1=$T \cdot I(\tilde{B} = 1) + \infty \cdot I(\tilde{B} \neq 1)$. Fine (2001) further extends this approach to more general transformation models which include the proportional odds model with the logistic link function.

Another popular analysis suggests a mixture model that expresses the cumulative incidence function as

$$F_j(t) = \pi_j(1 - Q_j(t)), \tag{2}$$

for $j = 1, \ldots, J$, where $\pi_j = \Pr(\tilde{B} = j)$ measures the marginal probability of type-j failure, and $1 - Q_j(t) = \Pr(T \leq t|\tilde{B} = j)$ describes the corresponding latency distribution for the sub-population with $\tilde{B} = j$. Wang (2003) considered nonparametric analysis under a two-path framework which is a special case of model (2) with $J = 2$. In presence of right censoring, Wang (2003) showed that valid estimators of π_j and $Q_j(t)$, the probability of experiencing path-j and the sojourn time for the path, can be obtained only if the follow-up period is long enough to recover the tail information of the first failure time. Specifically let C be the censoring variable. Define the two support points: τ_T =sup$\{t$: $\Pr(T > t) > 0\}$ and τ_C =sup$\{t$: $\Pr(C > t) > 0\}$. Wang (2003) mentioned that π_j and $Q_j(t)$ may not be identifiable if $\tau_T > \tau_C$.

For model (2) with covariates, Larson and Dinse (1985) assumed a multinomial logit model for π_j and a parametric proportional hazard model for $Q_j(t)$, and suggested the EM algorithm in the estimation procedure. Kuk (1992) considered a similar mixture model, in which the latency component follows a semi-parametric proportional hazard model, and took a rank likelihood approach in estimation. Notice that the maximum likelihood approach handles the estimation of π_j and $Q_j(t)$ jointly so that model mis-specification of one component may affect the validity of estimation for the other component. Fine (1999) also considered a binary regression model for π_j and a semi-parametric transformation model for $Q_j(t)$. Instead of using the likelihood approach, he proposed two estimating functions for the two components in (2) separately. This approach is more robust compared with previous methods but faces the issue of non-identifiability if $\tau_T > \tau_C$. To remedy the potential problem, Fine (1999) modeled $\pi_j(\tau) = \Pr(T \le \tau, \tilde{B} = j)$ and $Q_j(t|\tau) = \Pr(T > t | T \le \tau, \tilde{B} = j)$ instead, where τ is a predetermined time point satisfying $\tau < \tau_X$, where $\tau_X = \tau_C \wedge \tau_T$.

In this article, we also study model (2) and focus on regression analysis of the marginal failure probability π_j. The severe acute respiratory syndrome (SARS) provides an example. SARS is an epidemic and life-threatening acute disease that resulted in a global outbreak in 2003. A patient who was discharged from the hospital and was still alive at that time can be considered as being cured. Clinicians and the pubic usually are more interested in finding out which characteristics of a patient would affect his/her probability of being cured. We adopt a similar regression model as in Fine (1999) such that $F_j(t \wedge \tau|\mathbf{Z})$ can be written as

$$\Pr(T \le \tau, \tilde{B} = j|\mathbf{Z})\Pr(T \le t|T \le \tau, \tilde{B} = j, \mathbf{Z}) = \pi(\mathbf{Z}^T\boldsymbol{\beta}(\tau))(1 - Q_{j,\mathbf{Z}}(t|\tau)), \quad (3)$$

where $\mathbf{Z} = [1, \mathbf{z}^T]^T$ is the $(p+1) \times 1$ covariates, $\pi(\cdot)$ is a known function mapping from $(-\infty, \infty)$ to $(0, 1)$, $\boldsymbol{\beta}(\tau)$ is a $(p+1) \times 1$ vector of parameters and τ satisfies $\tau < \tau_X$ to avoid the problem of non-identifiability. Notice that we do not specify the form of $Q_{j,\mathbf{Z}}(t|\tau)$ since our primary concern is the marginal failure probability $\pi_j(\tau)$. In Section 2, we apply two principles of handling missing data to estimate $\boldsymbol{\beta}(\tau)$. The first approach utilizes the technique of weighting to adjust for the censoring bias which has also been considered by Fine (1999). The proposed estimator of of $\boldsymbol{\beta}(\tau)$ is more efficient than Fine's estimator. The second proposal uses the technique of imputation which extends the idea of Wang (2003) from a nonparametric setting to the regression framework. Technical results are summarized in the appendix. Section 3 contains simulation studies and data analysis. In Section 4, we give some concluding remarks.

2 The Proposed Methods

2.1 Preliminary

Without loss of generality, consider only two causes of failures, namely $\tilde{B} = j$ $(j = 1, 2)$. Denote $\{(T_i, \tilde{B}_i, \mathbf{Z}_i)\ (i = 1, \ldots, n)\}$ as a random sample of $(T, \tilde{B}, \mathbf{Z})$ under model (3). To simplify the notation, we will write $\boldsymbol{\beta} = \boldsymbol{\beta}(\tau)$ for model (3). The likelihood function of $\boldsymbol{\beta}$ is given by

$$\tilde{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{\pi(\mathbf{Z}_i^T \boldsymbol{\beta})\right\}^{\Delta_{1i}} \{\bar{\pi}(\mathbf{Z}_i^T \boldsymbol{\beta})\}^{1-\Delta_{1i}}, \tag{4}$$

where $\Delta_{1i} = I(\tilde{B}_i = 1, T_i \leq \tau)$ and $\bar{\pi}(t) = 1 - \pi(t)$. The resulting score function becomes

$$\tilde{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{\Delta_{1i} - \pi(\mathbf{Z}_i^T \boldsymbol{\beta})\right\} \frac{\pi_\phi(\mathbf{Z}_i^T \boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T \boldsymbol{\beta}) \bar{\pi}(\mathbf{Z}_i^T \boldsymbol{\beta})} \mathbf{Z}_i, \tag{5}$$

where $\pi_\phi(t) = \partial \pi(t)/\partial t$.

In presence of right censoring, the value of Δ_{1i} may be unknown. Observed variables become $\{(X_i, B_i, \mathbf{Z}_i)\ (i = 1, \ldots, n)\}$, which are *iid* replications of $(X, B, \mathbf{Z})$ where $X = T \wedge C$ and $B = \tilde{B} \cdot I(T \leq C)$. Under censoring, the likelihood function of $\boldsymbol{\beta}$ becomes very complicated and involves specification of several nuisance functions such as $\Pr(T > t | \tilde{B} = j, T \leq \tau, \mathbf{Z})$ and $\Pr(T > t | \tilde{B} = j, T > \tau, \mathbf{Z})$ for $j = 1, 2$. We propose to directly modify the score function $\tilde{U}(\boldsymbol{\beta})$ by applying two useful principles of handling missing data, namely the weighting approach and the imputation approach. We assume that, given $\mathbf{Z}$, C is independent of $(T, \tilde{B})$ and, to simplify the analysis, T and C are both continuous variables.

2.2 Inverse Probability Weighting

To simplify the notations, let $\delta_{1i} = I(B_i = 1)$, $\delta_{2i} = I(B_i = 2)$ and $\delta_i = \delta_{1i} + \delta_{2i}$. Suppose that the censoring distribution depends on $\mathbf{Z}$ only through some discrete covariates L which means that $\Pr(C \geq t | \mathbf{Z}) = \Pr(C \geq t | L) = G_L(t)$. The observable response

$I(X \le \tau, \delta_1 = 1)$ can be treated as a biased proxy of Δ_1. Moreover

$$E\left(\left.\frac{I(X \le \tau, \delta_1 = 1)}{G_L(X)}\right| \mathbf{Z}\right) = E\left[I(T \le \tau, \tilde{B} = 1) E\left(\left.\frac{I(T \le C)}{G_L(T)}\right| T, \tilde{B}, \mathbf{Z}\right)\right| \mathbf{Z}\right]$$

$$= E\left(\Delta_1 | \mathbf{Z}\right) = \pi(\mathbf{Z}^T\boldsymbol{\beta}), \tag{6}$$

which means that by taking an inverse weighting adjustment of $I(X \le \tau, \delta_1 = 1)$, we can obtain an unbiased proxy of Δ_1. Replacing Δ_{1i} by $I(X_i \le \tau, \delta_{1i} = 1)/\hat{G}_{L_i}(X_i)$ in the score function (5), Fine (1999) proposed the following estimating function

$$U_F(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \frac{I(X_i \le \tau, \delta_{1i} = 1)}{\hat{G}_{L_i}(X_i)} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta}) \right\} \frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})} \mathbf{Z}_i, \tag{7}$$

where $\hat{G}_L(t)$ is the Kaplan-Meier estimator of $G_L(t)$ with

$$\hat{G}_L(t) = \prod_{u<t} \left\{ 1 - \frac{\sum_{k=1}^{n} I(X_k = u, \delta_k = 0, L_k = L)}{\sum_{k=1}^{n} I(X_k \ge u, L_k = L)} \right\}. \tag{8}$$

Although the estimating function $U_F(\boldsymbol{\beta})$ has an intuitive and simple form, it may not be efficient since only a subset of the sample, (X, δ_1), was used in its construction.

As a comparison with (6), it can be shown that

$$E\left[\left.\frac{I(X > \tau)}{G_L(\tau+)} + \frac{I(X \le \tau, \delta_2 = 1)}{G_L(X)}\right| \mathbf{Z}\right] = 1 - \pi(\mathbf{Z}^T\boldsymbol{\beta}) = \bar{\pi}(\mathbf{Z}^T\boldsymbol{\beta}), \tag{9}$$

where $G_L(\tau+) = \Pr(C > \tau | L)$. Equation (9) alone can be used to construct an estimating function of $\boldsymbol{\beta}$ as (7). However possible more efficient estimation can be achieved by incorporating equations (6) and (9). Specifically for $i = 1, \ldots, n$, let

$$h_{1i} = \frac{I(X_i \le \tau, \delta_{1i} = 1)}{G_{L_i}(X_i)} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta}), \quad h_{2i} = \frac{I(X_i > \tau)}{G_{L_i}(\tau+)} + \frac{I(X_i \le \tau, \delta_{2i} = 1)}{G_{L_i}(X_i)} - \bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta}),$$

and $\mathbf{h}_i = [h_{1i}, h_{2i}]^T$. Applying the principles discussed in Heyde (1997, Chap.2), the optimal estimating function of $\boldsymbol{\beta}$ based on $\mathbf{H} = [\mathbf{h}_1^T, \ldots, \mathbf{h}_n^T]^T$ is

$$U^*(\boldsymbol{\beta}) = -E\left(\frac{\partial \mathbf{H}^T}{\partial \boldsymbol{\beta}}\right) \Sigma_{\mathbf{H}}^{-1} \mathbf{H} = \sum_{i=1}^{n} \left[-E\left(\frac{\partial \mathbf{h}_i^T}{\partial \boldsymbol{\beta}}\right)\right] \Sigma_{\mathbf{h}_i}^{-1} \mathbf{h}_i, \tag{10}$$

where $\Sigma_{\mathbf{H}} = E(\mathbf{H}\mathbf{H}^T)$,

$$\Sigma_{\mathbf{h}_i} = E(\mathbf{h}_i\mathbf{h}_i^T) = \begin{bmatrix} v_{1i} & -\pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta}) \\ & v_{2i} \end{bmatrix} \quad \text{and} \quad v_{ji} = Var(h_{ji}) \quad (j = 1, 2).$$

Because analytic expressions of v_{1i} and v_{2i} are not available, they can be replaced by other related quantities. Specifically

$$v_{1i} = E\left(\frac{I(X_i \le \tau, \delta_{1i} = 1)}{G^2_{L_i}(X_i)} \bigg| \mathbf{Z}_i \right) - \pi^2(\mathbf{Z}_i^T \boldsymbol{\beta}). \tag{11}$$

Based on the first-order Taylor expansion, the first term of the variance (11) can be approximated by

$$E\left(\frac{1}{G_{L_i}(X_i)} \bigg| \mathbf{Z}_i \right) E\left(\frac{I(X_i \le \tau, \delta_{1i} = 1)}{G_{L_i}(X_i)} \bigg| \mathbf{Z}_i \right) = E\left(\frac{1}{G_{L_i}(X_i)} \bigg| L_i \right) \pi(\mathbf{Z}_i^T \boldsymbol{\beta}).$$

Although $E\left(1/G_L(X) \middle| L\right)$ can be estimated by its moment estimate, this quantity is too sensitive to the tail behavior of $\hat{G}_L$. We suggest to use a related but more robust quantity instead such as the sample median of $\{1/\hat{G}_{L_i}(X_i) : L_i = L, i = 1, \dots, n\}$, denoted as M_L. Accordingly a reasonable proxy of v_{1i} is $\hat{v}_{1i} = \pi(\mathbf{Z}_i^T\boldsymbol{\beta})(M_{L_i} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta}))$, and by the same argument, we have $\hat{v}_{2i} = \bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})(M_{L_i} - \bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta}))$. Using $\hat{v}_{1i}$, $\hat{v}_{2i}$ and $\hat{G}_{L_i}(t)$ in replacement of the corresponding quantities in (10), the resulting estimating function becomes

$$U_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[(\hat{v}_{2i} - v_{3i})\hat{h}_{1i} - (\hat{v}_{1i} - v_{3i})\hat{h}_{2i} \right] \frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{\hat{v}_{1i}\hat{v}_{2i} - v^2_{3i}} \mathbf{Z}_i, \tag{12}$$

where

$$\hat{h}_{1i} = \frac{I(X_i \le \tau, \delta_{1i} = 1)}{\hat{G}_{L_i}(X_i)} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta}),\ \hat{h}_{2i} = \frac{I(X_i > \tau)}{\hat{G}_{L_i}(\tau+)} + \frac{I(X_i \le \tau, \delta_{2i} = 1)}{\hat{G}_{L_i}(X_i)} - \bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta}),$$

and $v_{3i} = \pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})$. Since the estimating function $U_1(\boldsymbol{\beta})$ is motivated by the idea of constructing the optimal estimating function, hopefully it is more efficient than $U_F(\boldsymbol{\beta})$ which only utilizes partial data. Finite sample comparison will be evaluated via simulations. Let $\boldsymbol{\beta}_0$ be the true value of $\boldsymbol{\beta}$. The solution to $U_1(\boldsymbol{\beta}) = 0$ is denoted as $\hat{\boldsymbol{\beta}}_1$. In Appendix A and B, we prove asymptotic normality of $U_1(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_1$. An *iid* expression of $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)$ is given in (A.4) which will be used in variance estimation described in Appendix B.

If C depends on continuous covariates L, $\hat{G}_L(t)$ can be estimated by the nonparametric smoothing techniques summarized in Section 2.3. Sometimes to avoid the curse of dimensionality, one may impose parametric or a semi-parametric models to describe the covariate effect on C.

2.3 Imputation by Conditional Mean

Alternatively one can impute Δ_{1i} by an estimate of its conditional mean given the data. Specifically $E(\Delta_{1i}|X_i, \delta_{1i}, \delta_{2i}, \mathbf{Z}_i)$ equals

$$I(X_i \le \tau, \delta_{1i} = 1) + I(X_i \le \tau, \delta_i = 0) p_{\mathbf{Z}_i}(X_i), \tag{13}$$

where $p_z(x) = E(\Delta_{1i}|\delta_i = 0, X_i = x; \mathbf{Z}_i = z)$. The first proposed estimator of $p_z(x)$ is derived under a purely nonparametric setting which generalizes the nonparametric results in Wang (2003) and Satten and Datta (2001). Their ideas are roughly summarized in Appendix C. With covariates, it follows that

$$p_z(x) = \Pr(T \le \tau, \tilde{B} = 1 | T > x, \mathbf{Z} = z) = \frac{\Pr(x < T \le \tau, \tilde{B} = 1 | \mathbf{Z} = z)}{S_z(x)}, \tag{14}$$

where $S_z(t) = \Pr(T > t | \mathbf{Z} = z)$. When $\mathbf{Z}$ takes discrete values only, we can partition the sample into several subsets according to the covariate values and then obtain

$$\hat{S}_z(t) = \prod_{u \le t} \left\{ 1 - \frac{\sum_{i=1}^n I(X_i = u, \delta_i = 1, \mathbf{Z}_i = z)}{\sum_{i=1}^n I(X_i \ge u, \mathbf{Z}_i = z)} \right\}. \tag{15}$$

Accordingly a model-free estimator of $p_z(x)$, denoted by $\hat{p}_z(x)$, is given by

$$\hat{p}_z(x) = \frac{1}{\hat{S}_z(x)} \frac{\sum_{i=1}^n I(x < X_i \le \tau, \delta_{1i} = 1, \mathbf{Z}_i = z) / \hat{G}_{L_i}(X_i)}{\sum_{i=1}^n I(\mathbf{Z}_i = z)}. \tag{16}$$

where $\hat{G}_{L_i}$ is obtained in (8).

For continuous $\mathbf{Z}$, nonparametric estimators of $\Pr(x < T \le \tau, \tilde{B} = 1 | \mathbf{Z} = z)$ and $S_z(x)$ can be derived using smoothing techniques. By applying the techniques of non-parametric regression as illustrated in Dabrowska (1987), the formula of $\hat{p}_z(x)$ becomes

$$\hat{p}_z(x) = \frac{1}{\sum_{i=1}^n \left[\frac{I(X_i > x, \delta_i = 1)}{\hat{G}_{L_i}(X_i)} + \frac{I(X_i > X_{(m)})}{\hat{G}_{L_i}(X_{(m)}+)} \right] B_{n,i}(z)} \sum_{i=1}^n \frac{I(x < X_i \le \tau, \delta_{1i} = 1)}{\hat{G}_{L_i}(X_i)} B_{n,i}(z), \tag{17}$$

where $X_{(m)}$ is the largest observed failure time and $B_{n,i}(z)$ is a random set of non-negative weights. Candidates of the weight function include kernel type weights, nearest neighbors or local linear weights. For example one can use the kernel type weight

$$B_{n,i}(z) = \frac{K(a_n^{-1}(z - \mathbf{Z}_i))}{\sum_{j=1}^n K(a_n^{-1}(z - \mathbf{Z}_j))}, \tag{18}$$

where $K(\cdot)$ is an appropriate kernel function and a_n is a sequence of bandwidth.

Based on (13) one can construct the following estimating function

$$U_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \hat{\Delta}_{1i} - \pi(\mathbf{Z}_i^T \boldsymbol{\beta}) \right\} \frac{\pi_\phi(\mathbf{Z}_i^T \boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T \boldsymbol{\beta}) \bar{\pi}(\mathbf{Z}_i^T \boldsymbol{\beta})} \mathbf{Z}_i, \tag{19}$$

where $\hat{\Delta}_{1i} = I(X_i \leq \tau, \delta_{1i} = 1) + I(X_i \leq \tau, \delta_i = 0)\hat{p}_{\mathbf{Z}_i}(X_i)$ and $\hat{p}_{\mathbf{Z}_i}(X_i)$ is given in (16) or (17) depending on the type of $\mathbf{Z}$.

The second proposal of estimating $p_z(x)$ utilizes the model assumption (3) in which (14) can be further expressed as $p_z(x; \boldsymbol{\beta}, Q_{1,z}(\cdot|\tau), Q_{2,z}(\cdot|\tau), S_z(\tau))$ which equals

$$\frac{Q_{1,z}(x|\tau)\pi(z^T\boldsymbol{\beta})}{Q_{1,z}(x|\tau)\pi(z^T\boldsymbol{\beta}) + Q_{2,z}(x|\tau)\{1 - S_z(\tau) - \pi(z^T\boldsymbol{\beta})\} + S_z(\tau)}. \tag{20}$$

We can estimate $S_z(\tau)$ by

$$\prod_{u \leq \tau} \left\{ 1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1) B_{n,i}(z)}{\sum_{i=1}^{n} I(X_i \geq u) B_{n,i}(z)} \right\}, \tag{21}$$

and $Q_{j,z}(t|\tau)$ $(j = 1, 2)$ by

$$\prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(u = X_i \leq \tau, \delta_{ji} = 1) B_{n,i}(z)}{\sum_{i=1}^{n} \left[I(u \leq X_i \leq \tau, \delta_{ji} = 1) + I(u \leq X_i \leq \tau, \delta_i = 0) w_{j,\mathbf{Z}_i}(X_i) \right] B_{n,i}(z)} \right\}, \tag{22}$$

where $w_{1,z}(x)$ is $\hat{p}_z(x)$ given in (16) or (17) depending on the type of $\mathbf{Z}$ and $w_{2,z}(x)$ is obtained by replacing δ_1 in $\hat{p}_z(x)$ with δ_2. For continuous covariates, the weight function $B_{n,i}(z)$ is given in (18) and for discrete covariates $B_{n,i}(z) = I(\mathbf{Z}_i = z)$. By plugging in appropriate estimators of $S_z(\tau)$ and $Q_{j,z}(t|\tau)$ $(j = 1, 2)$ in (20), the second proposed estimating function based on (13) is given by

$$U_3(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \tilde{\Delta}_{1i} - \pi(\mathbf{Z}_i^T \boldsymbol{\beta}) \right\} \frac{\pi_\phi(\mathbf{Z}_i^T \boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T \boldsymbol{\beta}) \bar{\pi}(\mathbf{Z}_i^T \boldsymbol{\beta})} \mathbf{Z}_i, \tag{23}$$

where $\tilde{\Delta}_{1i} = I(X_i \leq \tau, \delta_{1i} = 1) + I(X_i \leq \tau, \delta_i = 0) p_{\mathbf{Z}_i}(X_i; \boldsymbol{\beta}, \hat{Q}_{1,\mathbf{Z}_i}(\cdot|\tau), \hat{Q}_{2,\mathbf{Z}_i}(\cdot|\tau), \hat{S}_{\mathbf{Z}_i}(\tau))$, and $\hat{S}_{\mathbf{Z}_i}(\tau)$ and $\hat{Q}_{j,\mathbf{Z}_i}(\cdot|\tau)$ $(j = 1, 2)$ are given in (21) and (22).

The solution to $U_j(\boldsymbol{\beta}) = 0$ is denoted as $\hat{\boldsymbol{\beta}}_j$ $(j = 2, 3)$. The two estimating functions $U_2(\boldsymbol{\beta})$ and $U_3(\boldsymbol{\beta})$ differ in the ways of estimating $p_z(x)$. Since $U_3(\boldsymbol{\beta})$ estimates $p_z(x)$

using the model information, we expect that it may perform better than $U_2(\boldsymbol{\beta})$. Their finite-sample performances will be compared via simulations. Since $U_3(\boldsymbol{\beta})$ is a more complicated function of $\boldsymbol{\beta}$ than $U_2(\boldsymbol{\beta})$, numerical computation of $\hat{\boldsymbol{\beta}}_3$ becomes more difficult. To simplify the root-finding procedure, one may treat $\tilde{\Delta}_{1i}$ as a fixed number in the mth iteration by using $p_z(x;\hat{\boldsymbol{\beta}}^{(m-1)},\hat{Q}_{1,z}(\cdot|\tau),\hat{Q}_{2,z}(\cdot|\tau),\hat{S}_z(\tau))$ instead, where $\hat{\boldsymbol{\beta}}^{(m-1)}$ is the solution in the previous step. The final solution is obtained via an iterative procedure with $m = 1, 2, \ldots$. The modified estimating function is a simpler function of $\boldsymbol{\beta}$ and thus convergence can be achieved by only few steps of iterations.

When $\mathbf{Z}$ is discrete, we prove asymptotic normality of $n^{-1/2}U_2(\boldsymbol{\beta}_0)$ and that of $n^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0)$ in Appendix D. Similar arguments can be applied to establish asymptotic properties of $U_3(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_3$. When there are continuous covariates, the imputed estimators $\hat{p}_z(x)$ and $\hat{Q}_{j,z}(x)$ involve a smoothing weight function $B_{n,i}(z)$, which makes the analytic analysis more complicated. Since this technical issue is not our main focus, we suggest to apply the bootstrap re-sampling technique to approximate the distributions of $U_j(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_j$ $(j = 2, 3)$ for further inference. Its empirical performance is also examined in our simulations and data analysis.

3 Numerical Analysis

3.1 Simulation Studies

Finite sample performances of the proposed estimators were assessed via simulations. The covariate Z was generated from three distributions. For the discrete case, Z was generated from a *Bernoulli* random variable with $\Pr(Z = 0) = \Pr(Z = 1) = 1/2$. For the continuous case, Z was generated from $\text{Normal}(0, 1)$ or $\text{Unif}(-3, 3)$. Given Z, we generated $\Delta_1 = I(T \le \tau, \tilde{B} = 1)$ and $\Delta_2 = I(T \le \tau, \tilde{B} = 2)$ as follows. First Δ_1 was generated from a *Bernoulli* distribution with mean $\pi(\beta_0 + \beta_1 Z)$, where

$$\pi(\beta_0 + \beta_1 Z) = \frac{\exp(\beta_0 + \beta_1 Z)}{1 + \exp(\beta_0 + \beta_1 Z)}.$$

If $\Delta_1 = 1$, set $\Delta_2 = 0$ and, if $\Delta_1 = 0$, then generate Δ_2 from a *Bernoulli* distribution with mean $\pi(\alpha_0 + \alpha_1 Z)$. The parameters of interest are β_0 and β_1. Given (Δ_1, Δ_2), the failure time T was generated from a distribution with the density function f_T which can be expressed as

$$f_T(t) = \begin{cases} f_1(t|\tau, Z) & \text{if} \quad (\Delta_1, \Delta_2) = (1, 0) \\ f_2(t|\tau, Z) & \text{if} \quad (\Delta_1, \Delta_2) = (0, 1) \\ f_3(t|\tau, Z) & \text{if} \quad (\Delta_1, \Delta_2) = (0, 0), \end{cases}$$

where $f_j(t|\tau, Z)$ $(j = 1, 2)$ are density functions with supports no greater than τ and $f_3(t|\tau, Z)$ is a density function whose value exceeds τ. In the simulations, we set

$$f_j(t|\tau, Z) = \frac{f_{Y_j}(t)}{1 - S_{Y_j}(\tau)} I(t \leq \tau) \quad (j = 1, 2) \quad \text{and} \quad f_3(t|\tau, Z) = \frac{f_{Y_3}(t)}{S_{Y_3}(\tau)} I(t > \tau),$$

where $f_{Y_j}(t)$ and $S_{Y_j}(t)$ is the density function and the survival function of Y_j which follows the accelerated failure-time model of the form

$$\ln Y_j = \gamma_{0,j} + \gamma_{1,j} Z + \sigma_j \cdot W_j, \tag{24}$$

where $\gamma_{0,j}$, $\gamma_{1,j}$ and σ_j are (nuisance) parameters and W_j is the error distribution. The censoring variable was generated from $\text{Unif}(c_0, c_0 + c_1)$, where c_0 is pre-specified and the value of c_1 is determined to achieve the targeted censoring proportion (i.e. 30% or 40%). After generating a random sample of $(\Delta_1, \Delta_2, T, Z, C)$, denoted as $\{(\Delta_{1i}, \Delta_{2i}, T_i, Z_i, C_i)$ $(i = 1, \ldots, n)\}$, the observed data is $\{(X_i, \delta_{1i}, \delta_{2i}, Z_i)$ $(i = 1, \ldots, n)\}$, where $X_i = T_i \wedge C_i$, $\delta_{1i} = \Delta_{1i} I(T_i \leq C_i)$ and $\delta_{2i} = \Delta_{2i} I(T_i \leq C_i)$. The value of τ is set to be 2.5. The sample size n was set to be 100 or 300.

Four estimators of $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ were under evaluation. Besides the three proposed estimators $\hat{\boldsymbol{\beta}}_j$ $(j = 1, 2, 3)$, for comparison, we also evaluated the estimator proposed by Fine (1999), denote it by $\hat{\boldsymbol{\beta}}_F$ which solves $U_F(\boldsymbol{\beta}) = 0$. The criteria of comparison include the average bias (BS), the sample standard deviation (SD), the mean squared errors (MSE) and the relative efficiency (RE) which is defined as the ratio of the mean square errors of $\hat{\boldsymbol{\beta}}_F$ to that of $\hat{\boldsymbol{\beta}}_j$ $(j = 1, 2, 3)$.

Tables 1, 2 and 3 list the results for three types of the covariate Z respectively. For illustration, only the results for the estimator of β_1 are presented; the results for the estimator of β_0 are similar and hence omitted. The proposed estimators $\hat{\boldsymbol{\beta}}_j$ $(j = 1, 2, 3)$ are more efficient than $\hat{\boldsymbol{\beta}}_F$ especially in the case of continuous covariate. Furthermore, $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\beta}}_3$, obtained based on the imputation approach, perform better than $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_F$ which utilize the weighting approach. The estimator $\hat{\boldsymbol{\beta}}_3$ with model-based imputation slightly performs better than $\hat{\boldsymbol{\beta}}_2$ obtained by nonparametric imputation. The standard deviation estimates of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_F$ were computed using the formula given in (A.4) for three types of Z.

The standard deviation estimate of $\hat{\boldsymbol{\beta}}_2$ can also be obtained by applying the formula (A.4) if Z is discrete but becomes intractable analytically if Z is continuous. The same problem occurs to standard deviation estimation of $\hat{\boldsymbol{\beta}}_3$. In such a situation, we used the bootstrap re-sampling method. Specifically 1000 sub-samples were drawn with replacement from the original sample, and for ith sub-sample, we get $\hat{\boldsymbol{\beta}}_k^{(i)}$ by solving $U_k(\boldsymbol{\beta}) = 0$ $(k = 2, 3)$. Then the standard deviation estimate of $\hat{\boldsymbol{\beta}}_k$ can be obtained as the sample standard deviation of $\hat{\boldsymbol{\beta}}_k^{(i)}$ $(i = 1, \ldots, 1000)$. In Tables 1 $\sim$ 3, we listed the average of the proposed standard deviation estimates (ASD) and the corresponding empirical coverage probabilities of nominal 95% confidence intervals for $\boldsymbol{\beta}$ (CP). For all the cases, the empirical coverage probabilities are close to the nominal level and the values of ASD are close to those of SD, the sample standard deviation.

In Table 4, we investigated the situation when C actually depends on Z such that $\ln C = \gamma_{0,c} + \gamma_{1,c} Z + \sigma_c W_c$. For the case of discrete Z, the data generation scheme is same as that described for Table 1 and, for the case of continuous Z, the setup is same as in Table 2. In computation of the proposed estimators, we evaluated two estimators of $G(t) = \Pr(C \geq t)$. One is the Kaplan-Meier estimator and the other is a kernel-type estimator which simply replaces $\delta_i = 1$ by $\delta_i = 0$ in formula (21). Note that the former is based on the wrong assumption that C does not depend on Z. It turns out that the results based on the Kaplan-Meier estimator of $G(t)$ are biased while the kernel approach yields less biased estimators. Generally speaking, it seems that the misspecification of $\hat{G}(t)$ has more influence on the bias term (BS) and less on the standard deviation (SD). Our proposed estimators $\hat{\boldsymbol{\beta}}_j$ $(j = 1, 2, 3)$ relatively are more robust than $\hat{\boldsymbol{\beta}}_F$ under the mis-specification.

3.2 Data Analysis

The proposed inference procedures are applied to the Stanford Heart Transplant data (Crowley and Hu, 1977, p. 28-29). The main objective is to explore the relationship between certain covariates and the cause of death due to transplant rejection. This dataset was analyzed by Larson and Dinse (1985) in the context of a mixture model. Deaths were attributed to transplant rejection ($\tilde{B} = 1$) or other causes ($\tilde{B} = 2$). Among the 65 heart recipients, there were 29 rejected death ($B = 1$); 12 deaths were from other causes ($B = 2$) and 24 patients were censored ($B = 0$). The covariates include the waiting time from acceptance to surgery (w); the age at surgery (*age*) and a continuous mismatch score (m). Both m and *age* are transformed to have zero mean and unit variance, and w was recorded as a binary variable according to whether or not the waiting time exceeded 31 days. The survival time T (in days) was measured from the date of transplant surgery.

To determine if the censoring time C is related to the covariates, a Cox proportional hazard model was fitted for the censoring time C on each covariate respectively. Since that all p-values are greater than 0.1, the result indicates that all covariates have no significant effects on C. Hence we assume the independence between C and Z. Let $\pi(\tau) = \Pr(T \leq \tau, \tilde{B} = 1)$. We first investigated the marginal effect of each covariate for four selected values of $\tau = 250, 500, 900, 1800$. Under the logistic model

$$\log \left(\frac{\pi(\tau)}{1 - \pi(\tau)} \right) = b_0(\tau) + b_1(\tau) Z,$$

all proposed methods show that the effect of waiting time w is not significant and hence can be removed. For other covariates, *age* is significant for all values of τ and the effect of the mismatch score m is insignificant for small values of τ and then becomes more obvious as τ increases. We further consider multiple logistic regression by including *age* and m jointly in the model. In Table 5 we see that *age* still plays an important role for all values of τ; on the other hand the effect of mismatch score can be neglected when it is considered jointly with *age*. Thus we conclude that *age* is the determining factor of $\pi(\tau)$, that is, a patient with younger age at transplant surgery will have a lower chance to develop transplant rejection by time τ.

Larson and Dinse (1985) investigated the same dataset by setting $\tau = \tau_T$. It was assumed that the incidence rate of transplant rejection, $\Pr(\tilde{B} = 1)$, follows a logistic model and the latency distribution $1 - Q_j(t) = \Pr(T \leq t|\tilde{B} = j)$ follows a proportional hazard model with

$$\Pr(T > t|\tilde{B} = j, \mathbf{Z}) = \exp\left[-\int_0^t h_j(u)\exp(\mathbf{Z}^T\gamma_j)du\right] \tag{25}$$

where $h_j(t)$ is specified as a piece-wise exponential function for $j = 1, 2$. Their analysis showed that no covariates have significant effect on $\Pr(\tilde{B} = 1)$ but both *age* and m were important for the latency distribution associated with transplant rejection. Specifically, for patients destined to die from transplant rejection, one with younger age at transplant surgery and with better tissue match would have better survival prognosis.

Although our analysis took $\tau < \tau_T$, Table 5 shows that the effect of *age* on $\pi(\tau)$ persists throughout all selected values of τ. It is reasonable to expect that the effect of *age* will continue to $\pi(\tau_T) = \Pr(\tilde{B} = 1)$ as τ approaches to τ_T. From this aspect, we thought that the age at surgery plays an important role on a patient's chance of dying from transplant rejection. Our conclusion conflicts with that of Larson and Dinse (1985) which contributes the *age* effect only to the latency distribution $1 - Q_1(t)$.

To clarify the contradiction between the two analyses, we checked whether the *age* effect on the hazard of $T|\tilde{B} = 1$ is proportional. We partitioned the sample to three groups with $age \leq 45$ (*group* 1), $45 < age \leq 51$ (*group* 2) and $51 < age$ (*group* 3). Let $\hat{Q}_{1,j}(t)$ denote the estimator of $Q_1(t)$ for *group* j, where $\hat{Q}_{1,j}(t)$ $(j = 1, 2, 3)$ were calculated using the formula proposed by Wang (2003) described in Appendix C. Under the proportional hazard assumption, plots of $\ln(-\ln(\hat{Q}_{1,j}))$ should be approximately parallel for $j = 1, 2, 3$. From Figure 1, we found that the curves are close to parallel for $j = 1, 2$. However $\ln(-\ln(\hat{Q}_{1,3}))$ intersects with $\ln(-\ln(\hat{Q}_{1,2}))$ first and then crosses $\ln(-\ln(\hat{Q}_{1,1}))$ at later time. The latter phenomenon is an indication that the proportional assumption for *age* on Q_1 may not be adequate. Since Larson and Dinse (1985) adopted a parametric approach by estimating the incidence rates and latency distributions jointly, we doubt that, if the assumption on the latency part is not accurate, the corresponding result for $\Pr(\tilde{B} = 1)$ may not be valid. To confirm the above conjecture, we divided the data into two groups with $age \leq 45$ and $age > 45$ and the results are similar.

4 Concluding Remarks

The proposed methods can be directly compared with the result of Fine (1999) in which the weighting approach is used to estimate the regression parameters for the marginal probability of a particular cause of failure. Here we consider two methods, namely imputation and weighting. For the weighting approach, we also make concrete suggestions to improve efficiency. The imputation approach tends to work better in simulations but the corresponding analytic derivations are more complicated and may involve smoothing techniques if the covariate is continuous. The two inference techniques are popular in handling missing information and have been used by Jung (1996) and Subramanian (2001) in estimating the long term survival rate without competing risks.

Comparing our model (3) with the model (1) studied in Fine and Gray (1999) and Fine (2001), we find that setting $t = \tau$ and $g^{-1}(x) = \pi(x)$, model (1) coincides with model (3) and $\boldsymbol{\beta}(\tau) = [h(\tau),\ \boldsymbol{\theta}^T]^T$. In other words, model (3) fits the data at a single time point τ while model (1) considers the entire time span. If model (1) is appropriate, then the last p components of $\boldsymbol{\beta}(\tau)$ derived from model (3) will be similar for different choices of τ. Therefore results obtained for model (3) provides a way to verify the assumption of model (1) or help choosing time-dependent covariates in that model. Formally the assumption of model (1) can be verified by testing $H_0 : \tilde{\boldsymbol{\beta}}_0(\tau_1) = \tilde{\boldsymbol{\beta}}_0(\tau_2)$ for any $\tau_1 \neq \tau_2$, where $\tilde{\boldsymbol{\beta}}_0(\tau)$ is the true value of the last p components of $\boldsymbol{\beta}_0(\tau)$ evaluated at truncation time τ. Specifically let

$$R = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{p\times(p+1)}.$$

Under H_0, $n^{1/2}R\cdot\left\{\hat{\boldsymbol\beta}_1(\tau_2)-\hat{\boldsymbol\beta}_1(\tau_1)\right\}$ can be approximated by

$$n^{-1/2}R\cdot\left\{A_1^{-1}(\mathbf{Z}^T\boldsymbol\beta_0(\tau_2))\cdot U_1(\boldsymbol\beta_0(\tau_2))-A_1^{-1}(\mathbf{Z}^T\boldsymbol\beta_0(\tau_1))\cdot U_1(\boldsymbol\beta_0(\tau_1))\right\}$$

which converges to a mean-zero p dimensional normal random variable. The analytic expression of the above function given in (A.4) can be used in performing the test.

Appendix

Appendix A: *Asymptotic properties of* $U_1(\boldsymbol\beta)$

Recall the notations defined for the estimating function $U_1(\boldsymbol\beta)$ in Section 2.2, we have

$$U_1(\boldsymbol\beta)=\sum_{i=1}^n\left[(\hat v_{2i}-v_{3i})h_{1i}-(\hat v_{1i}-v_{3i})h_{2i}\right]\frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol\beta)}{\hat v_{1i}\hat v_{2i}-v_{3i}^2}\mathbf{Z}_i+B_{2n}(\boldsymbol\beta),$$

where

$$\begin{aligned} B_{2n}(\boldsymbol\beta) &= \sum_{i=1}^n\left\{\left[\frac{I(X_i\le\tau,\delta_{1i}=1)}{G_{L_i}(X_i)}\frac{\hat v_{2i}-v_{3i}}{\hat v_{1i}\hat v_{2i}-v_{3i}^2}\pi_\phi(\mathbf{Z}_i^T\boldsymbol\beta)\mathbf{Z}_i\right]\frac{G_{L_i}(X_i)-\hat G_{L_i}(X_i)}{\hat G_{L_i}(X_i)}\right. \\ &\quad -\left[\frac{I(X_i\le\tau,\delta_{2i}=1)}{G_{L_i}(X_i)}\frac{\hat v_{1i}-v_{3i}}{\hat v_{1i}\hat v_{2i}-v_{3i}^2}\pi_\phi(\mathbf{Z}_i^T\boldsymbol\beta)\mathbf{Z}_i\right]\frac{G_{L_i}(X_i)-\hat G_{L_i}(X_i)}{\hat G_{L_i}(X_i)} \\ &\quad \left.-\left[\frac{I(X_i>\tau)}{G_{L_i}(\tau+)}\frac{\hat v_{1i}-v_{3i}}{\hat v_{1i}\hat v_{2i}-v_{3i}^2}\pi_\phi(\mathbf{Z}_i^T\boldsymbol\beta)\mathbf{Z}_i\right]\frac{G_{L_i}(\tau+)-\hat G_{L_i}(\tau+)}{\hat G_{L_i}(\tau+)}\right\}. \end{aligned}$$

Suppose that the true value $\boldsymbol\beta_0$ is located in the interior of the parameter space, which is a bounded convex region and $\pi_\phi(\cdot)$ is bounded. To derive the asymptotic distribution of $n^{-1/2}U_1(\boldsymbol\beta_0)$, we first express the Kaplan-Meier estimator $\hat G_L(t)$ in the integral form,

$$\frac{G_L(t)-\hat G_L(t)}{G_L(t)}=\sum_{i=1}^n\int_0^t\frac{\hat G_L(u-)}{G_L(u)}\frac{I(L_i=L)dM_{C,i}(u)}{\bar Y_L(u)},$$

where

$$M_{C,i}(u) = I(X_i \le u, \delta_i = 0) - \int_0^u I(X_i \ge s) d\Lambda_{C,L_i}(s),$$

$\bar{Y}_L(u) = \sum_{i=1}^n I(X_i \ge u, L_i = L)$ and $\Lambda_{C,L}(s)$ is the cumulative hazard function of C given L. By the uniform convergence of the Kaplan-Meier estimator, we can write $n^{-1/2} B_{2n}(\boldsymbol{\beta}_0)$ as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\infty [q_{1,L_i}(t;\boldsymbol{\beta}_0) - q_{2,L_i}(t;\boldsymbol{\beta}_0) - q_{3,L_i}(t;\boldsymbol{\beta}_0)] \left(\frac{\bar{Y}_{L_i}(t)}{n}\right)^{-1} dM_{C,i}(t) + o_p(1),$$

where

$$q_{1,L_i}(t;\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{k=1}^n I(X_k \ge t, L_k = L_i) \left[\frac{I(X_k \le \tau, \delta_{1k} = 1)}{G_{L_k}(X_k)}\right] \frac{\mathrm{v}_{2k} - \mathrm{v}_{3k}}{\mathrm{v}_{1k}\mathrm{v}_{2k} - \mathrm{v}_{3k}^2} \pi_\phi(\mathbf{Z}_k^T \boldsymbol{\beta}_0) \mathbf{Z}_k, \tag{A.1}$$

$$q_{2,L_i}(t;\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{k=1}^n I(X_k \ge t, L_k = L_i) \left[\frac{I(X_k \le \tau, \delta_{2k} = 1)}{G_{L_k}(X_k)}\right] \frac{\mathrm{v}_{1k} - \mathrm{v}_{3k}}{\mathrm{v}_{1k}\mathrm{v}_{2k} - \mathrm{v}_{3k}^2} \pi_\phi(\mathbf{Z}_k^T \boldsymbol{\beta}_0) \mathbf{Z}_k, \tag{A.2}$$

$$q_{3,L_i}(t;\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{k=1}^n I(\tau \ge t, L_k = L_i) \left[\frac{I(X_k > \tau)}{G_{L_k}(\tau+)}\right] \frac{\mathrm{v}_{1k} - \mathrm{v}_{3k}}{\mathrm{v}_{1k}\mathrm{v}_{2k} - \mathrm{v}_{3k}^2} \pi_\phi(\mathbf{Z}_k^T \boldsymbol{\beta}_0) \mathbf{Z}_k, \tag{A.3}$$

$\mathrm{v}_{1k} = \pi(\mathbf{Z}_k^T\boldsymbol{\beta}_0)(\tilde{M}_{L_k} - \pi(\mathbf{Z}_k^T\boldsymbol{\beta}_0))$, $\mathrm{v}_{2k} = \bar{\pi}(\mathbf{Z}_k^T\boldsymbol{\beta}_0)(\tilde{M}_{L_k} - \bar{\pi}(\mathbf{Z}_k^T\boldsymbol{\beta}_0))$, $\mathrm{v}_{3k} = \bar{\pi}(\mathbf{Z}_k^T\boldsymbol{\beta}_0)\pi(\mathbf{Z}_k^T\boldsymbol{\beta}_0)$ and $\tilde{M}_L$ is the median of the random variable $1/G_L(X)$.

Therefore $n^{-1/2} U_1(\boldsymbol{\beta}_0)$ can be expressed as $n^{-1/2} \sum_{i=1}^n \xi_i + o_p(1)$, where

$$\begin{aligned} \xi_i \;=\; & \left\{ \left[\frac{I(X_i \le \tau, \delta_{1i} = 1)}{G_{L_i}(X_i)} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta}_0)\right] (\mathrm{v}_{2i} - \mathrm{v}_{3i}) \right. \\ - & \left. \left[\frac{I(X_i > \tau)}{G_{L_i}(\tau+)} + \frac{I(X_i \le \tau, \delta_{2i} = 1)}{G_{L_i}(X_i)} - \bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta}_0)\right] (\mathrm{v}_{1i} - \mathrm{v}_{3i}) \right\} \frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta}_0)}{\mathrm{v}_{1i}\mathrm{v}_{2i} - \mathrm{v}_{3i}^2} \mathbf{Z}_i \\ + & \int_0^\infty \frac{q_{L_i}(t;\boldsymbol{\beta}_0)}{y_{L_i}(t)} dM_{C,i}(t), \end{aligned}$$

$y_{L_i}(t) = \lim_{n\to\infty} \bar{Y}_{L_i}(t)/n$ and $q_{L_i}(t;\boldsymbol{\beta}_0) = \lim_{n\to\infty}[q_{1,L_i}(t;\boldsymbol{\beta}_0) - q_{2,L_i}(t;\boldsymbol{\beta}_0) - q_{3,L_i}(t;\boldsymbol{\beta}_0)]$. Since $\{\xi_i \ (i = 1, ..., n)\}$ are zero-mean independent random variables, by the multivariate central limit theorem, $n^{-1/2} U_1(\boldsymbol{\beta}_0)$ has an asymptotic normal distribution with mean 0 and covariance matrix $\Gamma_1 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \xi_i \xi_i^T$.

Appendix B: *Asymptotic properties of* $\hat{\boldsymbol{\beta}}_1$

Let $\hat{\boldsymbol{\beta}}_1$ be the solution to $U_1(\boldsymbol{\beta}) = 0$. Since $U_1(\boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\beta}$ and has a bounded derivative, consistency of $\hat{\boldsymbol{\beta}}_1$ follows. By a Taylor expansion of $n^{-1/2}U_1(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}_0$, we can write

$$0 = n^{-1/2}U_1(\hat{\boldsymbol{\beta}}_1) = n^{-1/2}U_1(\boldsymbol{\beta}_0) - A_1(\boldsymbol{\beta}_0)\, n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) + o_p(1),$$

where

$$A_1(\boldsymbol{\beta}_0) = - \lim_{n\to\infty} \frac{1}{n} \frac{\partial U_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

It follows that

$$n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) = [A_1(\boldsymbol{\beta}_0)]^{-1}\, n^{-1/2}U_1(\boldsymbol{\beta}_0) + o_p(1). \tag{A.4}$$

Hence $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $V_1 = [A_1(\boldsymbol{\beta}_0)]^{-1}\,\Gamma_1\,[A_1(\boldsymbol{\beta}_0)]^{-1}$.

Replacing $\boldsymbol{\beta}_0$, G_L, $y_L(t)$ and $d\Lambda_{C,L}(t)$ with $\hat{\boldsymbol{\beta}}_1$, $\hat{G}_L$, $\bar{Y}_L(t)/n$ and $dN_{C,L}(t)/\bar{Y}_L(t)$, where $N_{C,L}(t) = \sum_k I(X_k \le t, \delta_k = 0, L_k = L)$, respectively, $\hat{\xi}_i$ equals

$$\left\{ \left[\frac{I(X_i \le \tau, \delta_{1i} = 1)}{\hat{G}_{L_i}(X_i)} - \pi(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1) \right] (\hat{\mathrm{v}}_{2i} - \hat{\mathrm{v}}_{3i}) \right.$$

$$\left. - \left[\frac{I(X_i > \tau)}{\hat{G}_{L_i}(\tau+)} + \frac{I(X_i \le \tau, \delta_{2i} = 1)}{\hat{G}_{L_i}(X_i)} - \bar{\pi}(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1) \right] (\hat{\mathrm{v}}_{1i} - \hat{\mathrm{v}}_{3i}) \right\} \frac{\pi_\phi(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1)}{\hat{\mathrm{v}}_{1i}\hat{\mathrm{v}}_{2i} - \hat{\mathrm{v}}_{3i}^2}\mathbf{Z}_i$$

$$+ \frac{nI(\delta_i = 0)\hat{q}_{L_i}(X_i; \hat{\boldsymbol{\beta}}_1)}{\sum_{k=1}^n I(X_k \ge X_i, L_k = L_i)} - \sum_{j=1}^n \frac{nI(\delta_j = 0, X_i \ge X_j, L_j = L_i)\hat{q}_{L_i}(X_j; \hat{\boldsymbol{\beta}}_1)}{\left(\sum_{k=1}^n I(X_k \ge X_j, L_k = L_i)\right)^2},$$

where

$\hat{\mathrm{v}}_{1i} = \pi(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1)(M_{L_i} - \pi(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1))$, $\hat{\mathrm{v}}_{2i} = \bar{\pi}(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1)(M_{L_i} - \bar{\pi}(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1))$, $\hat{\mathrm{v}}_{3i} = \bar{\pi}(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1)\pi(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1)$,

$$\hat{q}_{L_i}(t; \hat{\boldsymbol{\beta}}_1) = \hat{q}_{1,L_i}(t; \hat{\boldsymbol{\beta}}_1) - \hat{q}_{2,L_i}(t; \hat{\boldsymbol{\beta}}_1) - \hat{q}_{3,L_i}(t; \hat{\boldsymbol{\beta}}_1),$$

and $\hat{q}_{j,L_i}(t;\hat{\boldsymbol{\beta}}_1)$ $(j=1,2,3)$ are obtained by using $\hat{\boldsymbol{\beta}}_1$, $\hat{G}_{L_k}$ and $(\hat{\mathrm{v}}_{1k},\hat{\mathrm{v}}_{2k},\hat{\mathrm{v}}_{3k})$ instead of $\boldsymbol{\beta}_0$, G_{L_k} and $(\mathrm{v}_{1k},\mathrm{v}_{2k},\mathrm{v}_{3k})$ in (A.1)−(A.3). It follows that the covariance matrix Γ_1 can be estimated by $\hat{\Gamma}_1 = n^{-1}\sum_{i=1}^{n}\hat{\xi}_i\hat{\xi}_i^T$ and then

$$\hat{V}_1 = \left[\hat{A}_1(\hat{\boldsymbol{\beta}}_1)\right]^{-1}\hat{\Gamma}_1\left[\hat{A}_1(\hat{\boldsymbol{\beta}}_1)\right]^{-1}$$

where

$$\hat{A}_1(\hat{\boldsymbol{\beta}}_1) = \sum_{i=1}^{n}\frac{1}{n}\left[\frac{\hat{\mathrm{v}}_{1i}+\hat{\mathrm{v}}_{2i}-2\hat{\mathrm{v}}_{3i}}{\hat{\mathrm{v}}_{1i}\hat{\mathrm{v}}_{2i}-\hat{\mathrm{v}}_{3i}^2}\pi_\phi^2(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}_1)\mathbf{Z}_i\mathbf{Z}_i^T\right].$$

Appendix C: *Previous nonparametric results of Wang (2003)*

Wang (2003) estimated the conditional path probability given $T > x$. Modifying her idea, we can estimate $p_j(x) = \Pr(T \le \tau, \tilde{B} = j | T > x)$ by

$$\hat{p}_j(x) = \frac{1}{n\hat{S}(x)}\sum_{i=1}^{n}\frac{I(x < X_i \le \tau, \delta_{ji}=1)}{\hat{G}(X_i)},$$

where $\hat{S}(x)$ is the Kaplan-Meier estimator of $S(x)$ which, according to (Satten and Datta, 2001), can be re-expressed as an average of inverse probability of censoring given by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(X_i > x, \delta_i = 1)}{\hat{G}(X_i)} + \frac{I(X_i > X_{(m)})}{\hat{G}(X_{(m)}+)}\right],$$

where $X_{(m)}$ denotes the largest observed failure time. Based on Wang's idea, $Q_j(t|\tau)$ can be estimated by

$$\prod_{u\le t}\left\{1 - \frac{\sum_{i=1}^{n} I(u = X_i \le \tau, \delta_{ji} = 1)}{\sum_{i=1}^{n}\left[I(u \le X_i \le \tau, \delta_{ji}=1) + I(u \le X_i \le \tau, \delta_{ji} = 0)\hat{p}_j(X_i)\right]}\right\}.$$

Appendix D: *Asymptotic properties of* $U_2(\boldsymbol{\beta})$

Suppose that $\mathbf{Z}$ takes K distinct values, $z_1,\ldots,z_K$. Original data are partitioned into K mutually exclusive subsets, $\left\{\left(\Delta_{1k}^j, X_k^j, \delta_{1k}^j, \delta_{2k}^j\right)\ (k=1,\ldots,n_j)\right\}$, which corresponds to the set of $\{i : (\Delta_{1i}, X_i, \delta_{1i}, \delta_{2i}, \mathbf{Z}_i = z_j)\ (i=1,\ldots,n)\}$ and $n_j = \sum_{i=1}^{n} I(\mathbf{Z}_i = z_j)$. Let

$\delta_k^j = \delta_{1k}^j + \delta_{2k}^j$ and $\delta_{3k}^j = 1 - \delta_k^j$, we have $p_{z_j}(X_k^j) = E(\Delta_{1k}^j \big| \delta_k^j = 0, X_k^j, \mathbf{Z} = z_j)$, which can be estimated by

$$\hat{p}_{z_j}(X_k^j) = \frac{1}{n_j \hat{S}_{z_j}(X_k^j)} \sum_{h=1}^{n_j} \frac{I(X_k^j < X_h^j \leq \tau, \delta_{1h}^j = 1)}{\hat{G}_{z_j}(X_h^j)},$$

where $\hat{S}_{z_j}(t)$ and $\hat{G}_{z_j}(t)$ are Kaplan-Meier estimators of $S_{z_j}(t) = \Pr(T > t|\mathbf{Z} = z_j)$ and $G_{z_j}(t) = \Pr(C \geq t|\mathbf{Z} = z_j)$. The estimating equation $U_2(\boldsymbol{\beta})$ can be re-expressed as

$$U_2(\boldsymbol{\beta}) = \sum_{j=1}^{K} \left\{ \sum_{k=1}^{n_j} \left[\hat{\Delta}_{1k,j} - \pi(\boldsymbol{\beta}^T z_j) \right] \frac{\pi_\phi(\boldsymbol{\beta}^T z_j)}{\pi(\boldsymbol{\beta}^T z_j)\bar{\pi}(\boldsymbol{\beta}^T z_j)} z_j \right\},$$

where $\hat{\Delta}_{1k,j} = I(\delta_{1k}^j = 1, X_k^j \leq \tau) + I(\delta_k^j = 0, X_k^j \leq \tau)\hat{p}_{z_j}(X_k^j)$.

To derive asymptotic distribution of $n^{-1/2}U_2(\boldsymbol{\beta}_0)$, we first express it as sum of the following two terms,

$$\begin{aligned} \frac{1}{\sqrt{n}} U_2(\boldsymbol{\beta}_0) \;&=\; \sum_{j=1}^{K} \sqrt{\frac{n_j}{n}} \left\{ \frac{1}{\sqrt{n_j}} \sum_{k=1}^{n_j} \left[E_k^j - \pi(\boldsymbol{\beta}_0^T z_j) \right] \Psi_{z_j}(\boldsymbol{\beta}_0) \right\} \\ &+\; \sum_{j=1}^{K} \sqrt{\frac{n_j}{n}} \left\{ \frac{1}{\sqrt{n_j}} \sum_{k=1}^{n_j} \delta_{3k}^j \left[\hat{p}_{z_j}(X_k^j) - p_{z_j}(X_k^j) \right] \Psi_{z_j}(\boldsymbol{\beta}_0) \right\} \qquad \text{(A.5)} \end{aligned}$$

where

$$E_k^j = E\left(\Delta_{1k}^j \big| X_k^j, \delta_{1k}^j, \delta_{2k}^j; \mathbf{Z} = z_j \right) = I(\delta_{1k}^j = 1, X_k^j \leq \tau) + I(\delta_k^j = 0, X_k^j \leq \tau) p_{z_j}(X_k^j)$$

and

$$\Psi_{z_j}(\boldsymbol{\beta}_0) = \frac{\pi_\phi(\boldsymbol{\beta}_0^T z_j)}{\pi(\boldsymbol{\beta}_0^T z_j)\bar{\pi}(\boldsymbol{\beta}_0^T z_j)} z_j.$$

Denote the last part of (A.5) by $C_2(\boldsymbol{\beta}_0)$, by the strong consistency of Kaplan-Meier estimators, we have

$$C_2(\boldsymbol{\beta}_0) \;=\; \sum_{j=1}^{K} \left\{ \sqrt{\frac{n_j}{n}} \Psi_{z_j}(\boldsymbol{\beta}_0) \left[C_{2.1}^j + C_{2.2}^j \right] \right\} + o_p(1),$$

where

$$C^j_{2.1} = \frac{1}{\sqrt{n_j}} \sum_{k=1}^{n_j} \left[\frac{\delta^j_{3k}}{n_j S_{z_j}(X^j_k)} \sum_{h=1}^{n_j} I(X^j_k < X^j_h \le \tau, \delta^j_{1h} = 1) \left(\frac{1}{\hat{G}_{z_j}(X^j_h)} - \frac{1}{G_{z_j}(X^j_h)} \right) \right],$$

$$C^j_{2.2} = \frac{1}{\sqrt{n_j}} \sum_{k=1}^{n_j} \left[\delta^j_{3k} \left(\frac{1}{\hat{S}_{z_j}(X^j_k)} - \frac{1}{S_{z_j}(X^j_k)} \right) \frac{1}{n_j} \sum_{h=1}^{n_j} \left(\frac{I(X^j_k < X^j_h \le \tau, \delta^j_{1h} = 1)}{G_{z_j}(X^j_h)} \right) \right].$$

Interchanging the summations in $C^j_{2.1}$, we get

$$C^j_{2.1} = \frac{1}{\sqrt{n_j}} \sum_{h=1}^{n_j} \left[B(X^j_h) \frac{I(X^j_h \le \tau, \delta^j_{1h} = 1)}{G_{z_j}(X^j_h)} \frac{\hat{G}_{z_j}(X^j_h) - G_{z_j}(X^j_h)}{G_{z_j}(X^j_h)} \right] + o_p(1)$$

where

$$B(X^j_h) = \lim_{n_j \to \infty} \frac{1}{n_j} \sum_{k=1}^{n_j} \frac{\delta^j_{3k} I(X^j_k < X^j_h)}{S_{z_j}(X^j_k)}.$$

One can write

$$\frac{\hat{G}_{z_j}(t) - G_{z_j}(t)}{G_{z_j}(t)} = \sum_{l=1}^{n_j} \int_0^t \frac{\hat{G}_{z_j}(u-)}{G_{z_j}(u)} \frac{dM^j_{C,l}(u)}{\bar{Y}^j(u)}$$

where

$$\bar{Y}^j(u) = \sum_{i=1}^{n_j} I(X^j_i \ge u), \quad M^j_{C,l}(u) = I(X^j_l \le u, \delta^j_{3l} = 1) - \int_0^u I(X^j_l \ge s) d\Lambda^j_C(s),$$

and $\Lambda^j_C(s)$ is the cumulative hazard function of C given $\mathbf{Z} = z_j$. It follows that

$$C^j_{2.1} = \frac{1}{\sqrt{n_j}} \sum_{l=1}^{n_j} \int_0^\infty \frac{q^j(u)}{p^j(u)} dM^j_{C,l}(u) + o_p(1),$$

where

$$q^j(u) = \lim_{n_j \to \infty} \frac{1}{n_j} \sum_{h=1}^{n_j} B(X^j_h) \frac{I(u \le X^j_h \le \tau, \delta^j_{1h} = 1)}{G_{z_j}(X^j_h)} \quad \text{and} \quad p^j(u) = \lim_{n_j \to \infty} \frac{\bar{Y}^j(u)}{n_j}.$$

Similarly, one can write

$$C^j_{2.2} = \frac{1}{\sqrt{n_j}} \sum_{l=1}^{n_j} \int_0^\infty \frac{r^j(u)}{p^j(u)} \, dM^j_{T,l}(u) + o_p(1),$$

where

$$r^j(u) = \lim_{n_j\to\infty} \frac{1}{n_j}\sum_{k=1}^{n_j} \frac{\delta_{3k}^j I(X_k^j \geq u) P_{z_j}(X_k^j)}{S_{z_j}(X_k^j)},\ \ P_{z_j}(X_k^j) = \lim_{n_j\to\infty} \frac{1}{n_j}\sum_{h=1}^{n_j} \frac{\delta_{1h}^j I(X_k^j < X_h^j \leq \tau)}{G_{z_j}(X_h^j)},$$

and

$$M_{T,l}^j(u) = I(X_l^j \leq u, \delta_{3l}^j = 0) - \int_0^u I(X_l^j \geq s) d\Lambda_T^j(s),$$

$\Lambda_T^j(s)$ is the cumulative hazard function of T given $\mathbf{Z} = z_j$.

In summary, we have

$$\frac{1}{\sqrt{n}} U_2(\boldsymbol{\beta}_0) = \sum_{j=1}^K \sqrt{\frac{n_j}{n}} \left(\frac{1}{\sqrt{n_j}} \sum_{k=1}^{n_j} \zeta_k^j \right) \Psi_{z_j}(\boldsymbol{\beta}_0) + o_p(1)$$

where

$$\zeta_k^j = E_k^j - \pi(\boldsymbol{\beta}_0^T z_j) + \int_0^\infty \frac{q^j(u)}{p^j(u)} dM_{C,k}^j(u) + \int_0^\infty \frac{r^j(u)}{p^j(u)} dM_{T,k}^j(u).$$

Notice that $(\zeta_1^j, \ldots, \zeta_{n_j}^j)$ are zero-mean independent random variables for each j where $j = 1, \ldots, K$. By the multivariate central limit theorem, $\frac{1}{\sqrt{n}} U_2(\boldsymbol{\beta}_0)$ has an asymptotical normal distribution with mean 0 and covariance matrix

$$\Gamma_2 = \lim_{n\to\infty} n^{-1} \sum_{j=1}^K \left(\sum_{k=1}^{n_j} {\zeta_k^j}^2 \right) \Psi_{z_j}(\boldsymbol{\beta}_0) \Psi_{z_j}^T(\boldsymbol{\beta}_0).$$

Let $\hat{\boldsymbol{\beta}}_2$ be the solution of $U_2(\boldsymbol{\beta}) = 0$. Asymptotic properties of $\hat{\boldsymbol{\beta}}_2$ can be obtained as of $\hat{\boldsymbol{\beta}}_1$ stated in Appendix B. According to (A.4), $n^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $V_2 = [A_2(\boldsymbol{\beta}_0)]^{-1}\Gamma_2[A_2(\boldsymbol{\beta}_0)]^{-1}$ where

$$A_2(\boldsymbol{\beta}_0) = E\left[\frac{\pi_\phi^2(\mathbf{Z}^T\boldsymbol{\beta}_0)}{\pi(\mathbf{Z}^T\boldsymbol{\beta}_0)\bar{\pi}(\mathbf{Z}^T\boldsymbol{\beta}_0)} \mathbf{Z}\mathbf{Z}^T \right].$$

References

Cheng, S. C., Fine, J. P. and Wei, L. J. (1998) Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, **54**, 219–228.

Crowley, J. and Hu, M. (1977) Covariance analysis of heart transplant survival data. *J. Am. Statist. Ass.*, **72**, 27–36.

Dabrowska, D. M. (1987) Non-parametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181–197.

Fine, J. P. (1999) Analyzing competing risks data with transformation models. *J. R. Statist. Soc.* B, **61**, 817–830.

Fine, J. P. and Gray, R. J. (1999) A proportional hazards model for the subdistribution of a competing risk. *J. Am. Statist. Ass.*, **94**, 496–509.

Fine, J. P. (2001) Regression modeling of competing crude failure probabilities. *Biostatistics*, **2**, 85–97.

Heyde, C. C. (1997) *Quasi-likelihood and its application : a general approach to optimal parameter estimation.* New York : Springer.

Jung, S-H (1996) Regression analysis for long-term survival rate. *Biometrika*, **83**, 227-232 .

Kuk, A. Y. C. (1992) A semiparametric mixture model for the analysis of competing risks data. *Austral. J. Statist.*, **34**(2), 169–180.

Larson, M. G. and Dinse, G. E. (1985) A mixture model for the regression analysis of competing risks data. *Applied Statistics*, **34**, 201–211.

Satten, G. A. and Datta, S. (2001) The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician*, **55**, 207-210.

Subramanian, S. (2001) Parameter estimation in regression for long-term survival rate from censored data. *Journal of Statistical Planning and Inference*, **99**, 211-222.

Wang, W. (2003) Nonparametric estimation of the sojourn time distributions for a multipath model. *J. R. Statist. Soc.* B, **65**, 921–935.

Sample size	% censored	Estimators	Criteria for estimators of β_1 BS	SD	ASD	CP (%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_1$	-0.015	0.519	0.515	95.9	0.270	1.208
		$\hat{\boldsymbol{\beta}}_2$	0.000	0.507	0.530	97.2	0.257	1.267
		$\hat{\boldsymbol{\beta}}_3$	-0.001	0.507	0.523	97.0	0.257	1.267
		$\hat{\boldsymbol{\beta}}_F$	0.028	0.570	0.580	96.8	0.326	1
100	40	$\hat{\boldsymbol{\beta}}_1$	-0.083	0.598	0.581	94.4	0.365	1.498
		$\hat{\boldsymbol{\beta}}_2$	-0.061	0.579	0.613	96.7	0.339	1.610
		$\hat{\boldsymbol{\beta}}_3$	-0.060	0.579	0.630	96.3	0.339	1.613
		$\hat{\boldsymbol{\beta}}_F$	-0.111	0.731	0.739	96.9	0.546	1
300	30	$\hat{\boldsymbol{\beta}}_1$	-0.011	0.306	0.297	95.1	0.094	1.208
		$\hat{\boldsymbol{\beta}}_2$	-0.010	0.301	0.300	95.3	0.091	1.248
		$\hat{\boldsymbol{\beta}}_3$	0.001	0.297	0.296	95.2	0.088	1.283
		$\hat{\boldsymbol{\beta}}_F$	-0.013	0.336	0.334	96.0	0.113	1
300	40	$\hat{\boldsymbol{\beta}}_1$	-0.028	0.338	0.338	94.8	0.115	1.411
		$\hat{\boldsymbol{\beta}}_2$	-0.027	0.334	0.346	95.1	0.112	1.443
		$\hat{\boldsymbol{\beta}}_3$	-0.026	0.333	0.340	95.3	0.112	1.447
		$\hat{\boldsymbol{\beta}}_F$	-0.033	0.401	0.411	96.4	0.162	1

Table 1: *Finite-sample performance of different estimators of* $\beta_1 = -1.24$ *based on* 1000 *replications when the covariate* Z *is binary and* W_1 *follows the standard logistic distribution with* $(\gamma_{0,1}, \gamma_{1,1}, \sigma_1) = (0.26, -0.5, 0.5)$*,* W_2 *follows the standard logistic distribution with* $(\gamma_{0,2}, \gamma_{1,2}, \sigma_2) = (-0.23, -0.14, 0.33)$*,* W_3 *follows the extreme value distribution with* $(\gamma_{0,3}, \gamma_{1,3}, \sigma_3) = (-0.14, -0.2, 1)$ *and* $(\alpha_0, \alpha_1) = (0, 1.64)$*.*

Sample size	% censored	Estimators	Criteria for estimators of β_1					
			BS	SD	ASD	CP (%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_1$	0.134	0.498	0.446	94.1	0.266	2.317
		$\hat{\boldsymbol{\beta}}_2$	-0.070	0.455	0.474	94.0	0.212	2.909
		$\hat{\boldsymbol{\beta}}_3$	-0.073	0.455	0.444	93.4	0.212	2.906
		$\hat{\boldsymbol{\beta}}_F$	0.134	0.774	0.841	93.4	0.617	1
100	40	$\hat{\boldsymbol{\beta}}_1$	0.138	0.541	0.479	94.3	0.312	3.742
		$\hat{\boldsymbol{\beta}}_2$	-0.079	0.475	0.476	95.3	0.231	5.044
		$\hat{\boldsymbol{\beta}}_3$	-0.077	0.474	0.481	95.7	0.231	5.062
		$\hat{\boldsymbol{\beta}}_F$	0.157	1.069	1.549	96.1	1.167	1
300	30	$\hat{\boldsymbol{\beta}}_1$	0.038	0.257	0.252	94.4	0.067	4.689
		$\hat{\boldsymbol{\beta}}_2$	-0.025	0.249	0.247	95.7	0.063	5.035
		$\hat{\boldsymbol{\beta}}_3$	-0.025	0.248	0.251	96.0	0.062	5.098
		$\hat{\boldsymbol{\beta}}_F$	0.097	0.553	0.566	93.8	0.316	1
300	40	$\hat{\boldsymbol{\beta}}_1$	0.048	0.280	0.270	95.1	0.081	6.940
		$\hat{\boldsymbol{\beta}}_2$	-0.065	0.251	0.252	93.5	0.067	8.322
		$\hat{\boldsymbol{\beta}}_3$	-0.064	0.253	0.258	94.0	0.068	8.212
		$\hat{\boldsymbol{\beta}}_F$	0.101	0.742	0.792	96.2	0.560	1

Table 2: *Finite-sample performance of different estimators of* $\beta_1 = 1.8$ *based on* 1000 *replications when the covariate follows a standard normal distribution and* W_1 *follows the extreme value distribution with* $(\gamma_{0,1}, \gamma_{1,1}, \sigma_1) = (1, -0.6, 0.4)$*,* W_2 *follows the extreme value distribution with* $(\gamma_{0,2}, \gamma_{1,2}, \sigma_2) = (0.24, -0.5, 1)$*,* W_3 *follows the standard logistic distribution with* $(\gamma_{0,3}, \gamma_{1,3}, \sigma_3) = (0.25, -0.25, 0.25)$ *and* $(\alpha_0, \alpha_1) = (1.2, 1)$*.*

Sample size	% censored	Estimators	Criteria for estimators of β_1					
			BS	SD	ASD	CP (%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_1$	0.041	0.291	0.261	93.3	0.086	3.330
		$\hat{\boldsymbol{\beta}}_2$	0.037	0.259	0.264	97.0	0.068	4.211
		$\hat{\boldsymbol{\beta}}_3$	0.032	0.258	0.270	97.2	0.068	4.259
		$\hat{\boldsymbol{\beta}}_F$	0.090	0.529	0.607	93.2	0.288	1
100	40	$\hat{\boldsymbol{\beta}}_1$	0.073	0.347	0.308	93.5	0.126	3.568
		$\hat{\boldsymbol{\beta}}_2$	0.049	0.311	0.317	96.4	0.099	4.521
		$\hat{\boldsymbol{\beta}}_3$	0.044	0.308	0.313	96.0	0.097	4.626
		$\hat{\boldsymbol{\beta}}_F$	0.127	0.657	0.891	93.3	0.448	1
300	30	$\hat{\boldsymbol{\beta}}_1$	0.019	0.143	0.146	95.1	0.021	4.303
		$\hat{\boldsymbol{\beta}}_2$	-0.028	0.124	0.140	96.5	0.016	5.543
		$\hat{\boldsymbol{\beta}}_3$	-0.030	0.122	0.131	96.1	0.016	5.664
		$\hat{\boldsymbol{\beta}}_F$	0.059	0.293	0.292	94.2	0.089	1
300	40	$\hat{\boldsymbol{\beta}}_1$	0.028	0.172	0.163	94.2	0.030	6.892
		$\hat{\boldsymbol{\beta}}_2$	0.015	0.167	0.168	95.6	0.028	7.443
		$\hat{\boldsymbol{\beta}}_3$	0.015	0.163	0.170	95.5	0.027	7.777
		$\hat{\boldsymbol{\beta}}_F$	0.099	0.447	0.495	94.6	0.209	1

Table 3: *Finite-sample performance of different estimators of* $\beta_1 = 0.86$ *based on* 1000 *replications when the covariate* Z *follows a uniform distribution and* W_1 *follows the extreme value distribution with* $(\gamma_{0,1}, \gamma_{1,1}, \sigma_1) = (-0.6, 1, 0.8)$*,* W_2 *follows the extreme value distribution with* $(\gamma_{0,2}, \gamma_{1,2}, \sigma_2) = (0.24, 0.4, 1)$*,* W_3 *follows the standard logistic distribution with* $(\gamma_{0,3}, \gamma_{1,3}, \sigma_3) = (0.25, -0.25, 0.25)$ *and* $(\alpha_0, \alpha_1) = (1.2, 1)$.

Sample size = 300, % censored = 30

Covariate Z	$\hat{G}$	Criteria	Estimators of β_1 $\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_2$	$\hat{\boldsymbol{\beta}}_3$	$\hat{\boldsymbol{\beta}}_F$
Binary	Kernel-type estimator	BS	-0.029	-0.015	-0.001	-0.037
		SD	0.319	0.316	0.315	0.323
		MSE	0.103	0.100	0.099	0.105
	Kaplan-Meier estimator	BS	0.092	0.083	0.081	-0.989
		SD	0.326	0.318	0.318	0.471
		MSE	0.114	0.108	0.108	1.199
Standard Normal	Kernel-type estimator	BS	-0.073	-0.066	-0.064	0.098
		SD	0.254	0.245	0.244	0.260
		MSE	0.070	0.064	0.064	0.077
	Kaplan-Meier estimator	BS	-0.191	-0.106	-0.103	2.775
		SD	0.258	0.248	0.246	1.058
		MSE	0.103	0.073	0.071	8.823

Table 4: *Robustness analysis when the censoring variable depends on Z. For binary Z, $\beta_1 = -1.24$ and $(\gamma_{0,c}, \gamma_{1,c}, \sigma_c) = (1.5, -1, 0.8)$; and for $Z \sim N(0,1)$, $\beta_1 = 1.8$ and $(\gamma_{0,c}, \gamma_{1,c}, \sigma_c) = (1.2, 1, 1)$. For both types of Z, W_c follows the standard logistic distribution.*

	τ	1800	900	500	250
U_1	int	0.545 (0.463)	-0.037 (0.374)	-0.653 (0.311)	-1.016 (0.333)
	age	1.561 (0.542)^a	1.279 (0.382)^a	0.970 (0.310)^a	1.070 (0.351)^a
	m	0.727 (0.549)	0.786 (0.496)	0.691 (0.392)	0.672 (0.386)

	τ	1800	900	500	250
U_2	int	0.139 (0.470)	-0.136 (0.410)	-0.775 (0.336)	-1.087 (0.375)
	age	1.357 (0.569)^a	1.208 (0.518)^a	0.927 (0.370)^a	1.052 (0.442)^a
	m	0.665 (0.629)	0.790 (0.654)	0.563 (0.432)	0.601 (0.452)

	τ	1800	900	500	250
U_3	int	0.137 (0.464)	-0.152 (0.410)	-0.760 (0.333)	-1.076 (0.378)
	age	1.329 (0.527)^a	1.197 (0.458)^a	0.921 (0.380)^a	1.047 (0.438)^a
	m	0.598 (0.580)	0.696 (0.553)	0.543 (0.410)	0.588 (0.451)

	τ	1800	900	500	250
U_F	int	0.420 (0.484)	-0.061 (0.370)	-0.657 (0.308)	-1.002 (0.330)
	age	1.624 (0.748)^a	1.265 (0.412)^a	0.949 (0.307)^a	1.080 (0.356)^a
	m	0.416 (0.603)	0.570 (0.502)	0.613 (0.395)	0.634 (0.394)

Table 5: *Multiple Regression analysis for Heart Transplant data. In each cell, the estimated parameter and its standard error (in parenthesis) are given. The item with p-value* < 0.05 *is marked by* ^{a}.

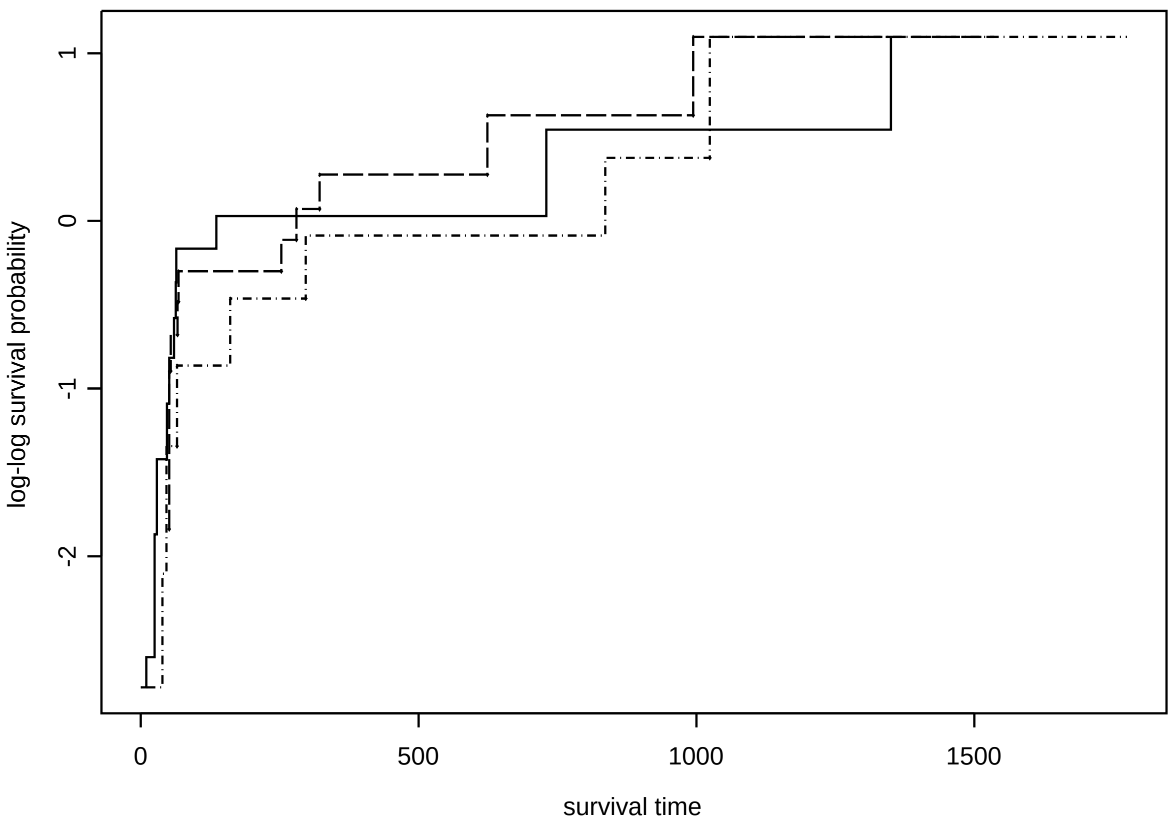

Figure 1: Plot of the log-log conditional survival function versus survival time for three groups with $age \leq 45$ (–·–·–), $45 < age \leq 51$ (– – –) and $51 < age$ (——).