Statistica Sinica **19** (2009), 391-408

REGRESSION ANALYSIS FOR CUMULATIVE INCIDENCE PROBABILITY UNDER COMPETING RISKS

Wei-Hwa Chang and Weijing Wang

Taiwan Suicide Prevention Center and National Chiao-Tung University

Abstract: The cumulative incidence function provides intuitive summary information about competing risks data. Via a mixture decomposition of this function, we study how covariates affect the cumulative incidence probability of a particular failure type at a chosen time point. Without specifying the corresponding failure time distribution, several inference methods are constructed based on imputation and weighting approaches. 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 well-known heart transplant data and compared with the analysis of Larson and Dinse (1985). In the on-line Supplement, we also apply our methods to analyze the Taiwan nationwide laboratory-confirmed severe acute respiratory syndrome (SARS) database.

Key words and phrases: Cause-specific hazard, cumulative incidence function, imputation, inverse probability of censoring, logistic regression, missing data, mixture model.

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 are usually summarized by the following two quantities. One is the cause-specific hazard function

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

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,

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

which measures the cumulative probability of developing type-j failure by time t. Note that both quantities make no assumption (such as independence) about the relationship between the competing risks events. There is a relationship between the two functions:

$$F_j(t) = \int_0^t S(u-)\lambda_j(u)du, \tag{1.1}$$

where $S(t) = \Pr(T > t) = \exp(-\int_0^t \sum_{j=1}^J \lambda_j(u)du)$ and $S(t-) = \Pr(T \geq t)$. Since $F_j(t)$ provides more direct information about the cumulative risk of the failure type of interest, it is more easily explained to clinicians.

This paper considers regression analysis of the cumulative incidence function. Different decompositions of $F_j(t)$ lead to different regression models. Equation (1.1) has been utilized by Cheng, Fine and Wei (1998) so that separate regression models for all causes of hazards are combined to make inference on $F_j(t)$. However, since the effect of a covariate on $\lambda_j(t)$ can be very different from its effect on $F_j(t)$, such an indirect approach can be misleading if the interest is in $F_j(t)$. Also, the parameters in the models for the cause-specific hazards may lack simple interpretation in terms of the cumulative incidence probabilities.

Another alternative is to model the whole 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) considered semi-parametric regression models of the form

$$g(F_1(t|\mathbf{z})) = h(t) + \mathbf{z}^T\boldsymbol{\theta}, \tag{1.2}$$

where $\mathbf{z}$ is a $p \times 1$ vector of covariates, $g(\cdot)$ is a known link function mapping from $(0, 1)$ to $(-\infty, \infty)$, and $h(t)$ is an unknown monotone function. Model (1.2) can also be explained in terms of a cure model based on the improper failure time

$$T_1 = T \cdot I(\tilde{B} = 1) + \infty \cdot I(\tilde{B} \neq 1).$$

Here a subject is said to be "cured" if one of the other competing events with $\tilde{B} \neq 1$ occurs earlier. For the choice of $g(\cdot)$, Fine and Gray (1999) considered the complementary log-log transformation with $g(u) = \log\{-\log(1 - u)\}$ that corresponds to the proportional hazards assumption on T_1. Fine (2001) suggested the logit transformation $g(u) = \log\{u/(1 - u)\}$ that corresponds to the proportional odds model on T_1. The model formulation in (1.2), however, seems somewhat restrictive. Specifically, when the model assumption holds, it follows that $g\{F_1(t|\mathbf{z}_1)\} - g\{F_1(t|\mathbf{z}_2)\} = (\mathbf{z}_1 - \mathbf{z}_2)^T\boldsymbol{\theta}$ for all t, which means that the cumulative incidence functions for subjects with different covariate values are "parallel" over the entire time span after the transformation of $g(\cdot)$. Although this drawback may be fixed by introducing time-varying covariates, the cure-model representation still lacks interpretability. Based on the improper failure time T_1, one can defined its "hazard" as

$$\tilde{\lambda}_1(t|\mathbf{z}) = -d\ln\{1 - F_1(t|\mathbf{z})\}/\,dt = \frac{dF_1(t|\mathbf{z})/dt}{1 - F_1(t - |\mathbf{z})},$$

where $F_1(t-|\mathbf{z}) = \Pr(T < t, \tilde{B} = 1|\mathbf{z})$. Notice that the denominator in the last identity, which indicates the at-risk probability for the first type of failure at time t, can be expressed as $\Pr(T \geq t, \tilde{B} = 1|\mathbf{z}) + \Pr(\tilde{B} \neq 1|\mathbf{z})$. That is, those subjects who have failed from other causes will always be treated as "at-risk" later on, which violates the common interpretation about the meaning of being "at-risk" for the event with $\tilde{B} = 1$.

The cumulative incidence function has also been analyzed based on the mixture decomposition

$$F_j(t) = \pi_j^*(1 - Q_j^*(t)) \quad (j = 1, \ldots, J), \tag{1.3}$$

where $\pi_j^* = \lim_{t\to\infty} F_j(t) = \Pr(\tilde{B} = j)$ 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 $\tilde{B} = j$. In the presence of covariates, both components of $F_j(t)$ on the right side of (1.3) can be modeled. For example, Larson and Dinse (1985) assumed a multinomial logit model for π_j^* and a parametric proportional hazard model for $Q_j^*(t)$. Kuk (1992) considered a similar mixture model, in which the latency component follows a semi-parametric proportional hazard model. However it was found that the two components in (1.3) could not be distinguished if the follow-up period is not long enough to completely recover the tail information of $F_j(t)$.

To remedy the problem of non-identifiability, Fine (1999) considered the representation

$$\begin{aligned} F_j(t \wedge \tau) &= \Pr(T \leq \tau, \tilde{B} = j) \Pr(T \leq t|T \leq \tau, \tilde{B} = j) \\ &= F_j(\tau) \cdot \{1 - Q_j(t|\tau)\}, \end{aligned}$$

where $U \wedge V = \min(U, V)$, and τ is a pre-determined time point located inside the support of the observed time variable. Statistical inference on $F_j(\tau)$ and $Q_j(t|\tau)$ are no longer subject to the potential problem of non-identifiability, as in (1.3), if τ is chosen properly to be located within the data range. In this article, we adopt this decomposition and consider the regression formulation

$$\begin{aligned} F_j(t \wedge \tau|\mathbf{Z}) &= \Pr(T \leq \tau, \tilde{B} = j|\mathbf{Z})\Pr(T \leq t|T \leq \tau, \tilde{B} = j, \mathbf{Z}) \\ &= \pi(\mathbf{Z}^T\boldsymbol{\beta}(\tau))(1 - Q_{j,\mathbf{Z}}(t|\tau)), \end{aligned} \tag{1.4}$$

where $\mathbf{Z} = [1, \mathbf{z}^T]^T$ is the $(p+1) \times 1$ vector of 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 lies within the data support. As in Fine (1999), we assume that $\pi(\mathbf{Z}^T\boldsymbol{\beta}(\tau))$ follows a binary regression model, such as the logistic regression model. In comparison, Fine (1999) imposed a transformation model on $Q_{j,\mathbf{Z}}(t|\tau)$, while we leave this latency distribution unspecified. Our main objective is to estimate $\boldsymbol{\beta}(\tau)$,

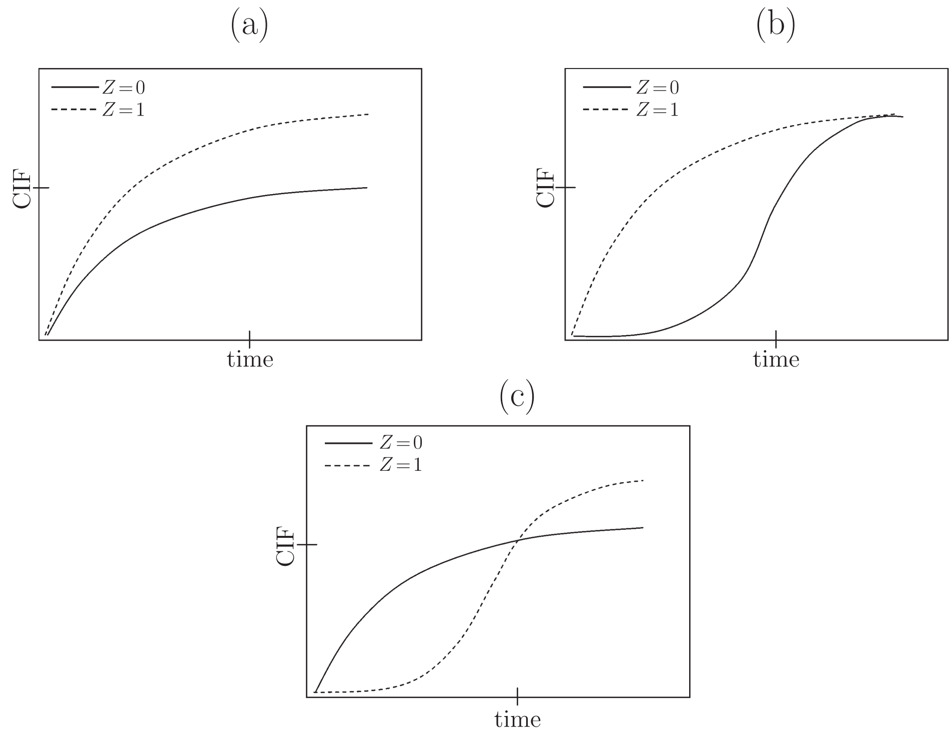

Figure 1. Illustration of possible Cumulative Incidence Functions (CIF) for a binary Z variable.

which measures the covariate effect on the cumulative probability of incidence by time τ.

Comparing the two regression formulations in (1.2) and (1.4), we find that setting $t = \tau$ and $g^{-1}(x) = \pi(x)$, the model at (1.2) coincides with the one at (1.4) and $\boldsymbol{\beta}(\tau) = [h(\tau),\ \boldsymbol{\theta}^T]^T$. In other words, (1.4) fits the data at a single time point τ while (1.2) considers modeling the entire time span. If (1.2) is appropriate, then the last p components of $\boldsymbol{\beta}(\tau)$ derived from (1.4) will be similar for different choices of τ. Therefore results obtained from (1.4) provide a way to verify the assumption of (1.2) or help choosing time-dependent covariates in that model. Figure 1 provides a graphical illustration to highlight the difference of the two models with a binary covariate. Figure 1a corresponds to the model at (1.2); in Figures 1b and 1c however, $F_1(t|0)$ and $F_1(t|1)$ have a crossing point, which violates model (1.2). Our model at (1.4) can include all three situations. It turns out that the dependency of $\boldsymbol{\beta}(\tau)$ on τ is not a subjective restriction, but provides the flexibility to detect possible change of covariate effect on the cumulative incidence probability at different time points.

The severe acute respiratory syndrome (SARS) provides an example that motivates model (1.4). SARS is an epidemic and life-threatening acute disease that resulted in a global outbreak in 2003. During the epidemic period, clinicians and the general public were more concerned with finding out the characteristics of a patient that would affect his/her probability of being discharged from the hospital and alive at a target time point. In the Taiwan nationwide laboratory-confirmed severe acute respiratory syndrome (SARS) database, an infected patient might be discharged and alive ($\tilde{B} = 1$) or have died in the hospital ($\tilde{B} = 2$). The original dataset contains complete information about the two outcomes and the corresponding failure time. There were 258 infected patients, of whom 58 were dead during the isolation period, and 200 were discharged from the hospital. In the Supplementary report, Figures S.1-S.5 provide the nonparametric estimators of $F_1(t)$ based on different levels of selected covariates. Notice that Figures S.4 and S.5, which show the estimated curves based on the covariates, Polymerase Chain Reaction (PCR) test and SARS viral load, respectively, match Figure 1c. The presence of crossing curves also suggests that the more general model at (1.4) is more suitable for the SARS data than the one at (1.2). Detailed analyses are presented in the on-line Supplement.

The major goal of this paper is to develop inference methods for estimating $\boldsymbol{\beta}(\tau)$ in the model at (1.4) in presence of censoring. In the SARS example, interim analysis based on incomplete data would provide timely information and hence would be helpful for decision making when the life-threatening disease was still epidemic. To estimate $\boldsymbol{\beta}(\tau)$, Fine (1999) proposed an estimating function, which is robust in the sense that model mis-specification of $Q_{j,\mathbf{Z}}(t|\tau)$ does not affect the estimation of $\boldsymbol{\beta}(\tau)$. In contrast, the likelihood approach often involves joint estimation and hence may lack of robustness if the other component of less interest is mis-specified (Larson and Dinse (1985)). The proposed methods also share the spirit of robustness as in Fine (1999), but we further consider efficiency improvement. In Section 2, we apply two principles for handling missing data to estimate $\boldsymbol{\beta}(\tau)$. The first approach utilizes the technique of weighting to adjust for the censoring bias, and can be considered as an extension of Fine's method (1999). The second uses imputation and extends the idea of Wang (2003) from a nonparametric setting to the current regression framework. Technical results are summarized in Section S4 of the on-line Supplement. 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, we consider only two causes of failures, namely $\tilde{B} = j$ $(j = 1, 2)$. Suppose that the first type of failure is of main interest. Let

$\{(T_i, \tilde{B}_i, \mathbf{Z}_i)\ (i = 1, \ldots, n)\}$ be a random sample of $(T, \tilde{B}, \mathbf{Z})$, with $\Delta_{ji} = I(T_i \leq \tau, \tilde{B}_i = j)$ $(j = 1, 2)$. We assume the model at (1.4) such that

$$E[\Delta_{1i}|\mathbf{Z}_i] = \pi(\mathbf{Z}_i^T\boldsymbol{\beta}) = \frac{\exp(\mathbf{Z}_i^T\boldsymbol{\beta})}{1 + \exp(\mathbf{Z}_i^T\boldsymbol{\beta})},$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}(\tau)$ is the parameter of interest. With the complete data, the likelihood function of $\boldsymbol{\beta}$ is given by

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

where $\bar{\pi}(t) = 1 - \pi(t)$, and the resulting score function is

$$\tilde{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{\Delta_{1i} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta})\right\} \frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})} \mathbf{Z}_i, \tag{2.1}$$

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

Given right censoring, let C be the censoring time. Observed variables are written as $\{(X_i, B_i, \mathbf{Z}_i)\ (i = 1, \ldots, n)\}$, *i.i.d.* replications of $(X, B, \mathbf{Z})$, where $X = T \wedge C$ and $B = \tilde{B} \cdot I(T \leq C)$. Note that the value of Δ_{1i} may be unknown due to censoring. It turns out that the likelihood function of $\boldsymbol{\beta}$ is very complicated and involves specification of several nuisance functions such as $Q_{j,\mathbf{Z}}(t|\tau)$ defined in (1.4), and $\Pr(T > t|\tilde{B} = j, T > \tau, \mathbf{Z})$ for $j = 1, 2$.

We modify the score function $\tilde{U}(\boldsymbol{\beta})$ by using two methods for handling missing data. The first approach utilizes observable proxies for Δ_{1i} by applying a weighting technique to adjust for their biases. The second approach imputes Δ_{1i} by an estimator of $E(\Delta_{1i}|X_i, B_i, \mathbf{Z}_i)$. We assume that, given $\mathbf{Z}$, C is independent of $(T, \tilde{B})$. To simplify the analysis, T and C are both continuous variables.

2.2. Inverse probability of weighting

Assume temporarily that the distribution of C does not depend on $\mathbf{Z}$. We will discuss possible modifications when this assumption does not hold. In presence of censoring, we can find observable proxies for Δ_{1i} and then apply the technique of inverse probability of censoring weighting (IPCW) to correct their biases. Specifically one has

$$\begin{aligned} E\left(\frac{I(X \leq \tau, B = 1)}{G(X-)}\middle|\mathbf{Z}\right) &= E\left[I(T \leq \tau, \tilde{B} = 1) E\left(\frac{I(T \leq C)}{G(T-)}\middle| T, \tilde{B}, \mathbf{Z}\right)\middle|\mathbf{Z}\right] \\ &= E\left[I(T \leq \tau, \tilde{B} = 1)\middle|\mathbf{Z}\right] = \pi(\mathbf{Z}^T\boldsymbol{\beta}), \end{aligned}$$

$$E\left[\frac{I(X>\tau)}{G(\tau)}+\frac{I(X\le\tau,B=2)}{G(X-)}\,\middle|\,\mathbf{Z}\right]=1-\pi(\mathbf{Z}^T\boldsymbol{\beta})=\bar\pi(\mathbf{Z}^T\boldsymbol{\beta}),$$

where $G(t)=\Pr(C>t)$ and $G(t-)=\Pr(C\ge t)$. These moment conditions can be utilized for construction of estimating functions of $\boldsymbol{\beta}$. Let

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

for $i=1,\ldots,n$. Replacing $G(t)$ with the Kaplan-Meier estimator

$$\hat G(t)=\prod_{u\le t}\left\{1-\frac{\sum_{k=1}^n I(X_k=u,B_k=0)}{\sum_{k=1}^n I(X_k\ge u)}\right\},\tag{2.2}$$

the resulting estimating functions are

$$U_{\mathrm{w}1}(\boldsymbol{\beta})=\sum_{i=1}^n\hat H_{1i}\frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{V_{1i}}\mathbf{Z}_i,\tag{2.3}$$
$$U_{\mathrm{w}2}(\boldsymbol{\beta})=\sum_{i=1}^n\hat H_{2i}\frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{V_{2i}}\mathbf{Z}_i,\tag{2.4}$$

where $\hat H_{ji}$ are H_{ji} ($j=1,2$) with G replaced by $\hat G$, and V_{ji} a weight function that measures the variation of $\hat H_{ji}$. A natural choice for V_{ji} is

$$\mathrm{Var}(H_{1i})=E\left(\frac{I(X_i\le\tau,B_i=1)}{G^2(X_i-)}\,\middle|\,\mathbf{Z}_i\right)-\pi^2(\mathbf{Z}_i^T\boldsymbol{\beta}),\tag{2.5}$$

but this involves unknown quantities and does not have an analytic expression. Based on a first-order Taylor expansion, the first term in (2.5) can be approximated by

$$E\left(\frac{1}{G(X_i-)}\,\middle|\,\mathbf{Z}_i\right)E\left(\frac{I(X_i\le\tau,B_i=1)}{G(X_i-)}\,\middle|\,\mathbf{Z}_i\right)\approx E\left(\frac{1}{G(X_i-)}\right)\pi(\mathbf{Z}_i^T\boldsymbol{\beta}).\tag{2.6}$$

Although $E\left(1/G(X-)\right)$ can be estimated by its moment estimator, this is sensitive to the tail behavior of $\hat G$ and may be unstable. Hence we suggest using a more robust quantity such as the sample median of $\{1/\hat G(X_i-):i=1,\ldots,n\}$, denoted as M_G. Accordingly we use $V_{1i}=\pi(\mathbf{Z}_i^T\boldsymbol{\beta})(M_G-\pi(\mathbf{Z}_i^T\boldsymbol{\beta}))$ and, by the same argument, we take $V_{2i}=\bar\pi(\mathbf{Z}_i^T\boldsymbol{\beta})(M_G-\bar\pi(\mathbf{Z}_i^T\boldsymbol{\beta}))$.

The estimating functions may be combined by constructing optimal estimating functions as discussed in Heyde (1997, Chap. 2). If $\mathbf{H}_i=[H_{1i},H_{2i}]^T$ for

$i = 1, \ldots, n$, the optimal estimating function of $\boldsymbol{\beta}$ based on $\mathbf{H} = [\mathbf{H}_1^T, \ldots, \mathbf{H}_n^T]^T$ is

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

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

$$\Sigma_{\mathbf{H}_i} = E(\mathbf{H}_i \mathbf{H}_i^T) = \begin{bmatrix} \mathrm{Var}\,(H_{1i}) & -\pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta}) \\ & \mathrm{Var}\,(H_{2i}) \end{bmatrix}.$$

Replacing $\mathrm{Var}\,(H_{ji})$ by V_{ji} $(j = 1, 2)$ whose forms have been suggested earlier, we obtain the estimating function

$$U_{\mathrm{w}^*}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[(V_{2i} - V_{3i})\hat{H}_{1i} - (V_{1i} - V_{3i})\hat{H}_{2i}\right] \frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{V_{1i}V_{2i} - V_{3i}^2}\mathbf{Z}_i, \tag{2.7}$$

where $V_{3i} = \pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})$.

The estimating functions in (2.3), (2.4) and (2.7) all reduce to $\tilde{U}(\beta)$ in the absence of censoring. With censored data, it is reasonable to suspect that $U_{\mathrm{w}^*}(\boldsymbol{\beta})$ is the most efficient one since it utilizes more information. It is interesting to note that the estimating function proposed by Fine (1999) actually has the form of $U_{\mathrm{w}1}(\beta)$ with a different weight $V_{1i} = \pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})$ that does not account for the effect of censoring. Via simulations, we will see how these different weight assignments affect the resulting estimators of $\boldsymbol{\beta}$.

Denote the solution to $U_{\mathrm{w}^*}(\boldsymbol{\beta}) = 0$ as $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$, and $\hat{\boldsymbol{\beta}}_{\mathrm{w}j}$ as the solution to $U_{\mathrm{w}j}(\boldsymbol{\beta}) = 0$ $(j = 1, 2)$. In Sections S4.1 and S4.2 of the on-line Supplement, we prove the asymptotic normality of $U_{\mathrm{w}^*}(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Note that

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0) = [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1}\, n^{-\frac{1}{2}} U_{\mathrm{w}^*}(\boldsymbol{\beta}_0) + o_p(1),$$

where $A_{\mathrm{w}^*}(\boldsymbol{\beta}_0) = -\lim_{n\to\infty} (1/n) \frac{\partial U_{\mathrm{w}^*}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$. Hence $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix

$$V_{\mathrm{w}^*} = [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1}\, \Gamma_{\mathrm{w}^*}\, [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1}, \tag{2.8}$$

where Γ_{w^*} is the asymptotic covariance matrix of $n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta}_0)$.

If the censoring variable C depends on discrete covariates, the Kaplan-Meier estimator $\hat{G}(t)$ can be evaluated for each covariate group. If the related covariate is continuous, we suggest two modification. In Section 2.3, we illustrate the use of the kernel method to estimate $\Pr(C > t|\mathbf{Z} = z)$. The other approach, which can avoid the curse of dimensionality, is to impose some parametric or semi-parametric models that describe the covariate effect on C.

2.3. Imputation by conditional mean

Alternatively, to handle possible incompleteness of $\Delta_1 = I(T \leq \tau, \tilde{B} = 1)$ due to censoring, one may impute its value by an estimate of the conditional mean given the data. Specifically $E[I(T \leq \tau, \tilde{B} = 1)|X, B, \mathbf{Z}]$ equals

$$I(X \leq \tau, B = 1) + I(X \leq \tau, B = 0)p_z(X),$$

where $p_z(x) = \Pr(T \leq \tau, \tilde{B} = 1|T > x, \mathbf{Z} = z)$. Two estimators of $p_z(x)$, denoted $\hat{p}_z^{(j)}(x)$ $(j = 1, 2)$, are proposed with specific forms given below. Replacing Δ_{1i} by

$$\hat{\Delta}_{1i}^{(j)} = I(X_i \leq \tau, B_i = 1) + I(X_i \leq \tau, B_i = 0)\hat{p}_z^{(j)}(X_i),$$

the score function (2.1) becomes

$$U_{Ij}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \hat{\Delta}_{1i}^{(j)} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta}) \right\} \frac{\pi_\phi(\mathbf{Z}_i^T\boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})}\mathbf{Z}_i \tag{2.9}$$

for $j = 1, 2$.

The proposed estimator $\hat{p}_z^{(1)}(x)$ is derived under a purely nonparametric setting that generalizes the nonparametric results in Wang (2003) and Satten and Datta (2001). Their ideas are roughly summarized in Section S4.3 of the on-line Supplement. With covariates, it follows that

$$p_z(x) = \Pr(T \leq \tau, \tilde{B} = 1|T > x, \mathbf{Z} = z) = \frac{\Pr(x < T \leq \tau, \tilde{B} = 1|\mathbf{Z} = z)}{S_z(x)}, \tag{2.10}$$

where $S_z(t) = \Pr(T > t|\mathbf{Z} = z)$ and $\Pr(x < T \leq \tau, \tilde{B} = 1|\mathbf{Z} = z)$ can be estimated nonparametrically. When $\mathbf{Z}$ takes only discrete values, a model-free estimator of $p_z(x)$ is given by

$$\hat{p}_z^{(1)}(x) = \frac{1}{\hat{S}_z(x)} \frac{\sum_{i=1}^{n} I(x < X_i \leq \tau, B_i = 1, \mathbf{Z}_i = z)/\hat{G}(X_i-)}{\sum_{i=1}^{n} I(\mathbf{Z}_i = z)}, \tag{2.11}$$

where $\hat{G}$ is obtained in (2.2) and

$$\hat{S}_z(t) = \prod_{u \leq t} \left\{ 1 - \frac{\sum_{i=1}^{n} I(X_i = u, B_i \neq 0, \mathbf{Z}_i = z)}{\sum_{i=1}^{n} I(X_i \geq u, \mathbf{Z}_i = z)} \right\}. \tag{2.12}$$

A nonparametric way of handling continuous $\mathbf{Z}$ is to apply some smoothing techniques. Using the idea of Dabrowska (1987), we obtain

$$\hat{p}_z^{(1)}(x)=\frac{1}{\sum_{i=1}^n\left[\frac{I(X_i>x,B_i\neq 0)}{\hat{G}(X_i-)}+\frac{I(X_i>X_{(m)})}{\hat{G}(X_{(m)})}\right]B_{n,i}(z)}\sum_{i=1}^n\frac{I(x<X_i\leq\tau,B_i=1)}{\hat{G}(X_i-)}B_{n,i}(z), \tag{2.13}$$

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

$$B_{n,i}(z)=\frac{K(a_n^{-1}(z-\mathbf{Z}_i))}{\sum_{\ell=1}^n K(a_n^{-1}(z-\mathbf{Z}_\ell))}, \tag{2.14}$$

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

The proposed estimator $\hat{p}_z^{(2)}(x)$ utilizes the model assumption in (1.4). Specifically equation (2.10) can be further expressed as $p_z(x;\boldsymbol{\beta},Q_{1,z}(\cdot|\tau),Q_{2,z}(\cdot|\tau),S_z(\tau))$ equal to

$$\frac{Q_{1,z}(x|\tau)\pi(z^T\boldsymbol{\beta})}{Q_{1,z}(x|\tau)\pi(z^T\boldsymbol{\beta})+Q_{2,z}(x|\tau)\{1-S_z(\tau)-\pi(z^T\boldsymbol{\beta})\}+S_z(\tau)}. \tag{2.15}$$

This still involves nuisance functions, namely $Q_{j,z}(x|\tau)$ $(j=1,2)$ and $S_z(\tau)$. Here we estimate these quantities in a nonparametric way. To simplify the presentation, we give the formula which includes both types of covariates by setting $B_{n,i}(z)=I(\mathbf{Z}_i=z)$ for discrete covariates, and use (2.14) for continuous covariates. The proposed estimator $\hat{Q}_{1,z}(t|\tau)$ can be written as

$$\prod_{u\leq t}\left\{1-\frac{\sum_{i=1}^n I(u=X_i\leq\tau,B_i=1)B_{n,i}(z)}{\sum_{i=1}^n\left[I(u\leq X_i\leq\tau,B_i=1)+I(u\leq X_i\leq\tau,B_i=0)\hat{p}_{\mathbf{Z}_i}^{(1)}(X_i)\right]B_{n,i}(z)}\right\}, \tag{2.16}$$

where the formula of $\hat{p}_{\mathbf{Z}}^{(1)}(x)$ is given in (2.11) or (2.13) for discrete and continuous $\mathbf{Z}$, respectively. The estimator of $Q_{2,z}(t|\tau)$, denoted as $\hat{Q}_{2,z}(t|\tau)$, has a similar expression as $\hat{Q}_{1,z}(t|\tau)$ with $B_i=1$ being replaced by $B_i=2$ as appropriate. The proposed estimator of $S_z(\tau)$ is

$$\prod_{u\leq\tau}\left\{1-\frac{\sum_{i=1}^n I(X_i=u,B_i\neq 0)B_{n,i}(z)}{\sum_{i=1}^n I(X_i\geq u)B_{n,i}(z)}\right\}. \tag{2.17}$$

The solution to $U_{Ij}(\boldsymbol{\beta})=0$ is denoted as $\hat{\boldsymbol{\beta}}_{Ij}$ $(j=1,2)$. These two estimating functions differ in the way they estimate $p_z(x)$. Via simulations, we examine whether the second proposal, which utilizes model information, has better performance. Since $U_{I2}(\boldsymbol{\beta})$ is a more complicated function of $\boldsymbol{\beta}$, to simplify the root-finding procedure we treat $\hat{\Delta}_{1i}^{(2)}$ 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 $m = 1, 2, \ldots$, etc. The modified equation is a simpler function of $\boldsymbol{\beta}$ and convergence can be achieved by only few steps of iterations.

In Section S4.4 of the on-line Supplement, we prove the asymptotic normality of $n^{-1/2}U_{I1}(\boldsymbol{\beta}_0)$ and that of $n^{1/2}(\hat{\boldsymbol{\beta}}_{I1} - \boldsymbol{\beta}_0)$, when $\mathbf{Z}$ is discrete. Similar arguments can be applied to establish asymptotic properties of $U_{I2}(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_{I2}$. For continuous covariates, asymptotic analysis is not provided since the method involves kernel smoothing, a technical issue and not our main focus. However, due to the complexity of the plugged-in nonparametric estimators for both types of covariates, we suggest applying the bootstrap re-sampling technique for variance estimation.

3. Numerical Studies

3.1. Simulation analysis

Finite-sample performance of the proposed estimators was evaluated via simulations. The covariate Z was generated first. Given Z, we generated $\Delta_1 \sim$ Bernoulli$(\pi(\beta_0 + \beta_1 Z))$. Then, given the value of Δ_1 we generated Δ_2, and then the failure time T given the values of (Δ_1, Δ_2). Finally, we generated the censoring time C which might censor the value of T and make (Δ_1, Δ_2) missing. The detailed data generation scheme is described in Section S3.1 of the on-line Supplement. Here we set the value of τ to be 2.5 and the sample size n to be 100 or 300. The parameters of interest are (β_0, β_1). Besides the three proposed estimators $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$, $\hat{\boldsymbol{\beta}}_{I1}$ and $\hat{\boldsymbol{\beta}}_{I2}$, for comparison we also evaluated the estimator proposed by Fine (1999), denoted as $\hat{\boldsymbol{\beta}}_F$, that solves $U_F(\boldsymbol{\beta}) = 0$. Recall that $U_F(\boldsymbol{\beta})$ has the form of $U_{\mathrm{w}1}(\boldsymbol{\beta})$ with $V_{1i} = \pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})$. Based on $1,000$ replications, we report 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 the other three estimators. For each case, we also evaluated the accuracy of the proposed variance estimators. The criteria include the average of the proposed standard deviation estimates (ASD), and the corresponding empirical coverage probabilities of nominal 95% confidence intervals for $\boldsymbol{\beta}$ (CP) based on $1,000$ replications. The standard deviation estimates of $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ and $\hat{\boldsymbol{\beta}}_F$ were computed using the formula given in (2.8), which can handle both discrete and continuous types of Z. Although (2.8) can be applied to estimate the standard deviation of $\hat{\boldsymbol{\beta}}_{I1}$, it is complicated and becomes intractable analytically when Z is continuous. Thus for $\hat{\boldsymbol{\beta}}_{Ij}$ $(j = 1, 2)$, we

Table 1. Finite-sample comparison for four estimators of $\beta_1 = -1.24$ when the covariate Z is binary. The label BS denotes the average bias, SD denotes the sample standard deviation, ASD denotes the average of the standard deviation estimates, CP denotes the empirical coverage probabilities of nominal 95% confidence intervals, MSE denotes the mean squared errors, and RE denotes the relative efficiency defined as the ratio of the MSE of $\hat{\boldsymbol{\beta}}_F$ to that of the others.

Sample size	% censored	Estimators	Comparison criteria					
			BS	SD	ASD	CP (%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_{w^*}$	-0.015	0.519	0.515	95.9	0.270	1.208
		$\hat{\boldsymbol{\beta}}_{I1}$	0.000	0.507	0.530	97.2	0.257	1.267
		$\hat{\boldsymbol{\beta}}_{I2}$	-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	$\hat{\boldsymbol{\beta}}_{w^*}$	-0.083	0.598	0.581	94.4	0.365	1.498
		$\hat{\boldsymbol{\beta}}_{I1}$	-0.061	0.579	0.613	96.7	0.339	1.610
		$\hat{\boldsymbol{\beta}}_{I2}$	-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	$\hat{\boldsymbol{\beta}}_{w^*}$	-0.011	0.306	0.297	95.1	0.094	1.208
		$\hat{\boldsymbol{\beta}}_{I1}$	-0.010	0.301	0.300	95.3	0.091	1.248
		$\hat{\boldsymbol{\beta}}_{I2}$	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	$\hat{\boldsymbol{\beta}}_{w^*}$	-0.028	0.338	0.338	94.8	0.115	1.411
		$\hat{\boldsymbol{\beta}}_{I1}$	-0.027	0.334	0.346	95.1	0.112	1.443
		$\hat{\boldsymbol{\beta}}_{I2}$	-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

used the bootstrap re-sampling method for variance estimation. The procedure is described in Section S3.2 of the Supplement.

Tables 1 lists the results when Z is binary. In Table S.4 and S.5 of the Supplement, we report the results when Z follows standard normal and uniform distributions, respectively. We only present the analysis for the estimation of β_1 since the results for β_0 are similar. The results show that all the proposed estimators were more efficient than $\hat{\boldsymbol{\beta}}_F$. In the Supplement, we can see that this phenomenon becomes more obvious when Z is continuous. We also observe a larger bias of $\hat{\boldsymbol{\beta}}_F$, especially when the sample size was small. We explain why $\hat{\boldsymbol{\beta}}_F$ is sometimes not stable in the Supplement. For all cases, the empirical coverage probabilities are close to the nominal level, and the values of ASD are close to those of SD. In Table S.6 of the Supplement, we investigate how robust the proposed methods are when C actually depends on Z.

3.2. Analysis of heart transplant data

The proposed inference procedures were applied to the Stanford Heart Transplant data (Crowley and Hu (1977, pp.28-29)). The main objective was 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 to other causes ($\tilde{B} = 2$). Among the 65 heart recipients, there were 29 rejected deaths ($B = 1$); 12 deaths were from other causes ($B = 2$), and 24 patients were censored ($B = 0$). The covariates included the waiting time from acceptance to surgery (w); the age at surgery (*age*), and a continuous mismatch score (m). Both m and *age* were 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.

The Cox proportional hazard model was fit for the censoring time C on each covariate separately, and all p-values are larger than 0.1. Hence we assume that the distribution of C does not depend on the covariates. The quantity of interest is $F_1(\tau) = \Pr(T \le \tau, \tilde{B} = 1)$, the cumulative incidence probability of rejection by time τ. We set $\tau = 250, 500, 900, 1,800$ (days). For each covariate, we ran simple logistic regression under the model

$$\log\left(\frac{F_1(\tau)}{1 - F_1(\tau)}\right) = \beta_0(\tau) + \beta_1(\tau)Z, \tag{3.1}$$

where Z is one of the covariates. The waiting time w was not significant at any value of τ. The effect of the mismatch score m was insignificant for small values of τ, and then became more obvious as τ increased. The covariate *age* is significant for all values of τ. Excluding w, we fit a multiple logistic regression model which contained *age* and m. In Table 2 we see that *age* still played an important role for all values of τ, but the effect of mismatch score was insignificant when it was considered jointly with *age*. We conclude that *age* was the determining factor of $F_1(\tau)$. That is, younger patients with transplant surgery were less likely to suffer transplant rejection.

Larson and Dinse (1985) analyzed the same dataset under the framework of model (1.3). They assumed that $\Pr(\tilde{B} = 1)$, the incidence rate of dying from transplant rejection, follows a logistic model and the latency distribution $1 - Q_j^*(t) = \Pr(T \le t | \tilde{B} = j)$ follows a proportional hazard model for $j = 1, 2$. Their analysis showed that no covariates had a significant effect on $\Pr(\tilde{B} = 1)$, but that both *age* and m were important for the latency distribution associated with

Table 2. Multiple regression analysis for the heart transplant data. In each cell, the estimated parameter and its standard error (in parenthesis) are given. Items with p-values < 0.05 are marked by ^{a}.

		Selected values of τ (days)			
		1800	900	500	250
U_{w^*}	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)
U_{I1}	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)
U_{I2}	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)
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)

transplant rejection. Our results agree with those of Larson and Dinse (1985) in that *age* plays an important role for $F_1(t)$. However, Larson and Dinse (1985) attributed the influence of *age* on $F_1(t)$ to the latency distribution $1 - Q_1^*(t)$; in contrast, our analysis showed that the effect of *age* on $F_1(\tau)$ persisted for all selected values of τ. It is reasonable to expect that such effect might carry on to affect $\Pr(\tilde{B} = 1)$.

To investigate how the two analyses contradicted each other, we plot the nonparametric estimators of $F_1(t)$ for different age groups in Figure 2. The sample was partitioned into three age levels such that group j represents the group with $age \leq 45$, $\in (45, 51)$, and > 51 for $j = 1, 2, 3$ respectively. The curves of the two older groups differ at the beginning, but then grow closer in time. The conclusion of Larson and Dinse (1985) that *age* affected the latency distribution seems to make sense if only the two older groups are considered. However, the youngest group had lower cumulative incidence probability of developing rejection throughout the entire study period. This supports our conclusion.

Figure 2 shows that the curve of the youngest group has no crossing with those of the older groups. We can formally verify the assumption of model (1.2) that *age* affected $F_1(t)$ homogeneously over time. Consider the hypothesis, $H_0 : \beta_1(\tau_1) = \beta_1(\tau_2)$ for any $\tau_1 \neq \tau_2$, based on the model in (3.1). The proposed test statistic is $[\hat{\beta}_{w^*,1}(\tau_1) - \hat{\beta}_{w^*,1}(\tau_2)]/\hat{V}$, where $\hat{\beta}_{w^*,1}(\tau_1)$ and $\hat{\beta}_{w^*,1}(\tau_2)$ are obtained by solving $U_{w^*}(\boldsymbol{\beta}(\tau)) = 0$ with τ evaluated at τ_1 and τ_2, respectively, and $\hat{V}$ is the estimated standard error of $\hat{\beta}_{w^*,1}(\tau_1) - \hat{\beta}_{w^*,1}(\tau_2)$. Related derivations are

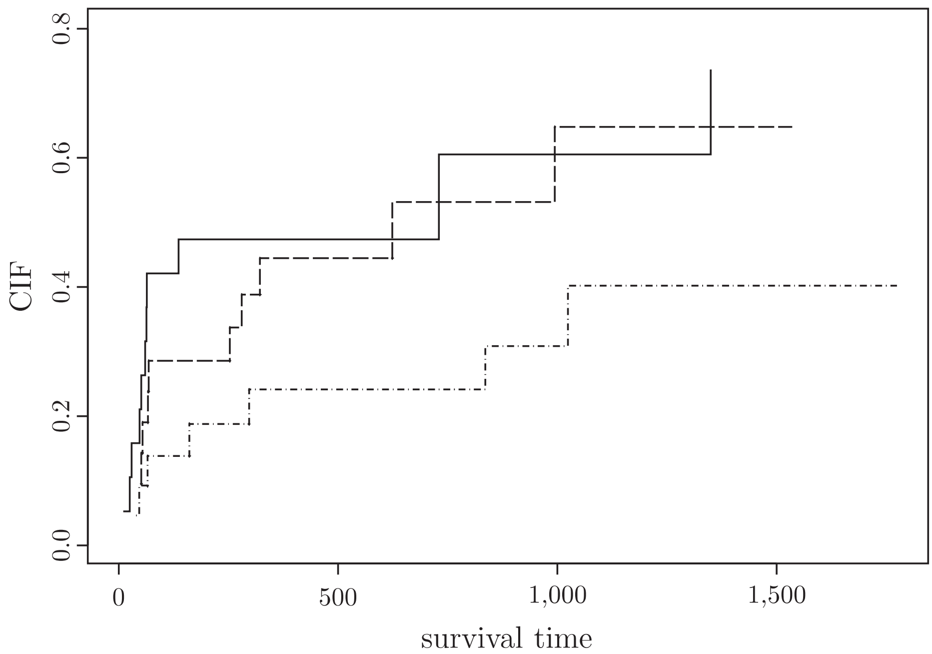

Figure 2. Estimated cumulative incidence functions of transplant rejection for three groups with $age \leq 45$ (−·−·−), $45 < age \leq 51$ (−−−) and $51 < age$ (——).

given in Section S4.5 of the on-line Supplement. Choosing $(\tau_1, \tau_2) = (1800, 500)$, the resulting p-value is 0.289, which implies that (1.2) is also a reasonable model.

To simplify the presentation, we assume a logistic model for all values of τ. The proposed method can be applied to different forms of $\pi(\cdot)$ for different values of τ. To determine an appropriate link function at a given time point, we choose the parametric link family $g(\mu;\rho) = \log\{[(1/(1-\mu))^{\rho} - 1]/\rho\}$, which includes the logistic model as the special case with $\rho = 1$. Testing the goodness-of-fit for the logistic model is equivalent to testing $\rho = 1$ under the more general family. Here we adopt the strategy proposed by Pregibon (1980) to implement model checking. His idea is based on applying a Taylor series expansion to $g(\mu;\rho)$ around $\rho = 1$, the hypothesized value. It follows that

$$\begin{aligned} g(\mu;\rho) &\simeq g(\mu;1) + (\rho - 1)\left\{\frac{\partial}{\partial\rho} g(u;\rho)\right\}_{\rho=1}, \\ &= \ln[\frac{\mu}{1-\mu}] - (\rho - 1)\frac{[\ln(1-\mu) + \mu]}{\mu}. \end{aligned}$$

The above derivations show that the correct link function $g(\mu;\rho) = \mathbf{Z}^T\boldsymbol{\beta}$ can be approximated by

$$g(\mu; 1) = \mathbf{Z}^T\boldsymbol{\beta} + (\rho - 1)\frac{[\ln(1-\mu) + \mu]}{\mu}. \tag{3.2}$$

Let $\hat{\mu}$ be the estimated value of μ evaluated under the null model with $\rho = 1$. Equation (3.2) can be viewed as a logistic model based on covariates $[\mathbf{Z}, [\ln(1-\hat{\mu}) + \hat{\mu}]/\hat{\mu}]$ with the corresponding parameters $\gamma^T = [\boldsymbol{\beta}^T, \rho - 1]$. To test the hypothesis, $H_0 : \rho = 1$, we consider the test statistic, $W = n^{-1}U_{\mathrm{w}^*}^T(\hat{\gamma})\hat{\Gamma}_{\mathrm{w}^*}^{-1}U_{\mathrm{w}^*}(\hat{\gamma})$, where $\hat{\gamma}$ is evaluated under the null hypothesis and $\hat{\Gamma}_{\mathrm{w}^*}$ is the estimated covariance matrix of $n^{-1/2}U_{\mathrm{w}^*}(\hat{\gamma})$. The result in Section S4.1 of the on-line Supplement can be applied to show that under H_0, W is asymptotically chi-square with one degree of freedom. For four selected time points 1,800, 900, 500 and 250, the resulting values of W are 7.341, 3.443, 2.025 and 0.737 respectively, suggesting that the logistic link becomes less suitable for larger τ. This also suggests that a different form of the link function should be considered for modeling the overall probability of dying due to transplant rejection, namely $\Pr(\tilde{B} = 1)$.

4. Concluding Remarks

In this article, we suggest inverse probability of censoring weighting and imputation for handling missing responses in the analysis of a logistic regression model. These approaches have been used by Jung (1996) and Subramanian (2001) for estimating the long-term survival rate without competing risks. The proposed estimating function $U_{\mathrm{w}^*}(\beta)$ further considers efficiency improvement by utilizing more data information in presence of censoring. The imputation approach had better performance in our simulations, but it also involved estimating more nuisance quantities. We have demonstrated that these nuisance functions can be handled nonparametrically by applying the results of Wang (2003) to the current regression setting which, however, may need to use smoothing techniques and hence is quite technically involved. Furthermore if the dimension of the continuous covariates is high, kernel smoothing may not work well unless the sample size is large. In such a case, one may try to reduce the dimension of $\mathbf{Z}$ based on a preliminary analysis, or impose additional model assumptions on the latency distributions.

Now we discuss the inverse probability of censoring weighting method further. We propose applying the weighting technique directly to two proxies of the response Δ_{1i}, namely $I(X_i \le \tau, B_i = j)$ $(j = 1, 2)$. Alternatively the weighting scheme can be applied to the mean-corrected response, $\Delta_{1i} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta})$. The latter method has been generalized by Robins and Rotnitzky (1992) for efficiency improvement. Specifically, they proposed the so-called augmented inverse probability of censoring weighting (AIPCW) method by adding an augmented term in the original weighted estimating function. Applying their idea to our problem, we obtain

$$U_A(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{I(B_i \neq 0)}{G(X_i-)}[\Delta_{1i} - \pi(\mathbf{Z}_i^T\boldsymbol{\beta})]\frac{\pi(\mathbf{Z}_i^T\boldsymbol{\beta})}{\pi(\mathbf{Z}_i^T\boldsymbol{\beta})\bar{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})}\mathbf{Z}_i + A_i(e, G), \tag{4.1}$$

where $A_i(e,G) = \int_0^{X_i} e(u,\mathbf{Z}_i)/G(u)dM_{C,i}(u)$ is the augmented term in which $e(u,\mathbf{Z}) = E(\Delta_1 - \pi(\mathbf{Z}^T\boldsymbol{\beta})|T \geq u, \mathbf{Z}) = p_{\mathbf{Z}}(u) - \pi(\mathbf{Z}^T\boldsymbol{\beta})$, with $p_{\mathbf{Z}}(u)$ defined in (2.10) and $dM_{C,i}(u) = I(X_i = u, B_i = 0) - I(X_i \geq u)\lambda_C(u)du$, with $\lambda_C(u)$ denoting the hazard function of the censoring variable C. Notice that the first term of $U_A(\boldsymbol{\beta})$ only involves uncensored observations. The second component $A_i(e,G)$ is a mean-zero augmented term contributed from subject i if censored. Robins and Rotnitzky (1992) have shown that the AIPCW method possesses attractive properties such as local efficiency and double robustness. The book by Tsiatis (2006) contains more detailed discussions about this method. Notice that $A_i(e,G)$ involves the conditional incidence probability $p_{\mathbf{Z}}(u)$ which also appears in the proposed imputation approach. This implies that we can view $U_A(\boldsymbol{\beta})$ as a way of combining the methods of weighting and imputation. Despite the aforementioned advantages, a crucial drawback of $U_A(\boldsymbol{\beta})$ is that its validity depends on the condition that $E(I(B \neq 0)/G(X-)|\mathbf{Z}) = 1$ for all $\mathbf{Z}$. It is easy to show that if the support of C is shorter than the support of T, $E(I(B \neq 0)/G(X-)|\mathbf{Z}) = \Pr(T \leq \tau_C|\mathbf{Z}) < 1$, where $\tau_C = \sup\{t : \Pr(C > t) > 0\}$. Due to this constraint, the application of the AIPCW method to our problem is limited.

For the imputation method, we suggest imputing Δ_{1i} by its conditional mean or the corresponding estimator. The so-called "multiple imputation" approach is another option. Specifically we can replace a missing value of Δ_{1i} by a random variable generated from the conditional distribution of Δ_{1i} given the observed data $(X, B, \mathbf{Z})$, which is a Bernoulli distribution with success probability equal to the estimated value of $p_Z(X)$ defined in (2.10). Subsequent analysis is carried out by treating the imputed sample as uncensored. The imputation procedure is repeated many times and the corresponding estimates of $\boldsymbol{\beta}$ are averaged to obtain the final estimate.

Acknowledgement

This paper was funded by the National Science Council of Taiwan (project number: 94-2118-M-009-002-). We thank Dr. Chen-Hsin Chen for insightful suggestions and thank Dr. Mei-Shiang Ho for kindly proving the SARS data.

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. Amer. Statist. Assoc.* **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. Roy. Statist. Soc. Ser. 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. Assoc.* **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.* Springer, New York.

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**, 169-180.

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

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Appl. Statist.* **29**, 15-24.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology - Methodological Issues* (Edited by N. Jewell, K. Dietz and V. Farewell), 297-331. Birkhauser, Boston.

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

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data.* Springer, New York.

Subramanian, S. (2001). Parameter estimation in regression for long-term survival rate from censored data. *J. Statist. Plann. Inference* **99**, 211-222.

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

Taiwan Suicide Prevention Center, 6F., No.63, Bo-ai Rd., Jhongjheng District, Taipei City 100, Taiwan(R.O.C.).

E-mail: songo.chang@gmail.com

Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan, R.O.C.

E-mail: wjwang@stat.nctu.edu.tw

(Received June 2006; accepted January 2008)