Ann Inst Stat Math (2016) 68:855–876
DOI 10.1007/s10463-015-0516-y

CrossMark

Competing risks data analysis under the accelerated failure time model with missing cause of failure

Ming Zheng · Renxin Lin · Wen Yu

Received: 13 August 2014 / Revised: 29 January 2015 / Published online: 12 March 2015

Abstract Competing risks data with missing cause of failure are analyzed under the accelerated failure time model which is a popular semiparametric linear model in survival analysis. The missing mechanism is assumed to be missing at random. The inverse probability weighted and double robust techniques are used to modify the rank-based estimating functions for competing risks data with complete observations on cause of failure. Proper optimization technique is utilized to obtain the desired estimators. The proposed algorithm overcomes the difficulty in solving the rank estimating equations with discontinuous estimating functions. The asymptotic properties of the proposed estimators are established. To implement the related inferences, a nonparametric bootstrap approach as well as a score test is developed. Simulation studies are carried out to assess the finite sample performance of the proposed method and validate the theoretical findings. The new estimating procedure is illustrated with the data from a bone marrow transplant study.

Keywords Bootstrap · Cause-specific hazard · Competing risks · Double robust · Inverse probability weighted · Missing at random · Rank estimator

1 Introduction

In survival analysis, individuals may fail from several different causes; such phenomena are referred to as competing risks. Suppose that every individual possesses a potential failure time for each competing risk, while the research interest lies only on one of them. People usually only observe the minimal failure time as well as the failure type. Also, in most cases the minimal failure times may subject to right

M. Zheng · R. Lin · W. Yu (✉)
Department of Statistics, School of Management, Fudan University, Shanghai 200433, China
e-mail: wenyu@fudan.edu.cn

censoring. For such competing risks data, cause-specific hazards have been widely used to relate the covariates with the interested failure time (Prentice and Kalbfleisch 1978; Cox and Oakes 1984). It is well known that by assuming conditional independence among different potential failure times, there is a one-to-one correspondence between the cause-specific hazard and the potential failure time's hazard. Under this assumption, imposing regression models on the hazard of the interested failure time is equivalent to introducing the model to its corresponding cause-specific hazard function.

It has been pointed out in literature that in some clinical trials and bioassay experiments, the cause of failure for some individuals might be unavailable to observe because the documentation needed for cause type identification is lost, or the cause type is difficult to determine, or the cause type detection is expensive to do for each subject, etc. Under these circumstances, the data on cause of failure are subject to missing. For analysis of competing risks data with missing cause of failure, assuming only one cause of failure is of interest, some authors have proposed statistical methods based on imposing various semiparametric regression models on the interested cause-specific hazard function. For example, Goetghebeur and Ryan (1995), Lu and Tsiatis (2001) and Hyun et al. (2012) used the most popular proportional hazards model. Gao and Tsiatis (2005) studied linear transformation model. Lu and Liang (2008) discussed additive hazard model. Some closely related research are on semiparametric regression analysis for right censored survival data with missing censoring indicators; c.f., McKeague and Subramanian (1998), Gijbels et al. (2007), Liu and Wang (2010) on Cox regression model, Zhou and Sun (2003), Song et al. (2010) on additive hazard models, Wang and Dinse (2011) on linear regression model, etc.

Although hazard-based semiparametric models play a central role when modeling survival data, linear regression models still provide valuable alternatives. The accelerated failure time (AFT) model is an attractive semiparametric linear model to practitioners because of its simple structure and ease of interpretation. For estimating the regression coefficients of the AFT model with usual right censored survival data, the rank-based estimating procedure has been widely studied by many authors, such as Prentice (1978), Tsiatis (1990), Ying (1993), Lin et al. (1998), among many others. The large sample properties, including the consistency and the asymptotic normality, are established for the rank estimators. However, it is well known that the rank estimating equations are difficult to solve numerically since the rank estimating functions are discontinuous in the regression parameters. Moreover, the limiting variance–covariance matrix of the rank estimators is also hard to estimate. These hamper the popularity of the AFT model. To overcome these difficulties, Jin et al. (2003) developed an algorithm to obtain the rank estimators via linear programming and an inference procedure based on the random weighting technique. Compared with right censored data, less attention has been paid to competing risks AFT model. With complete observations on cause of failure, Kalbfleisch and Prentice (2002) briefly mentioned imposing the AFT model on cause-specific hazards and applying the rank estimation procedure, with no discussion on theoretical properties. More recently, Lee and Lewbel (2009) explored the nonparametric identifiability of the competing risks AFT models and applied a sieve maximum likelihood procedure to estimate parameters.

To the best of our knowledge, although statistical analysis for competing risks data with missing cause of failure has been explored under various semiparametric survival models, no literature discussed the AFT model. In this paper, we propose an AFT model to relate the interested cause-specific hazard and a set of covariates. To deal with the missing cause of failure, modified rank estimating functions based on the inverse probability weighted and double robust techniques are developed under the missing at random (MAR) assumption. The two techniques, developed by Horvitz and Thompson (1952) and Robins et al. (1994) respectively, are commonly used in treating missing cause of failure or missing censoring information; c.f., Gao and Tsiatis (2005), Lu and Liang (2008), Song et al. (2010), Wang and Dinse (2011) and Hyun et al. (2012). Similar to the rank estimation procedure with ordinary right censored data, the proposed modified rank estimating function is neither continuous nor monotone in regression parameters. We develop an algorithm to transform the equation solving problem to an optimization problem. The optimization can be implemented reliably by some standard software packages. The asymptotic properties of the proposed estimators are studied. For the inferences about the regression parameters, since the proposed modified rank estimating function involves extra-estimators, the random weighting approach developed by Jin et al. (2003) cannot be extended directly. Therefore, we turn to a nonparametric bootstrap method for confidence intervals and propose a score test for testing problems.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the notation, and describe the AFT model for competing risks data and the MAR assumption for the missing cause of failure. In Sect. 3, we first develop a rank-based estimation procedure for the AFT model with completely observed cause of failure. Then, the modified rank estimating equations based on the inverse probability weighted and double robust techniques are constructed respectively to account for missing cause of failure. The asymptotic properties of the resulting estimators are discussed. A nonparametric bootstrap approach and a score test are developed to implement inferences. In Sect. 4, some simulation studies are conducted to assess the finite sample performances of the proposed estimators. Our method is applied to a real data example in Sect. 5. Section 6 concludes. All technique details are summarized in the "Appendix 7".

2 Notation and model specification

Without loss of generality, we assume that each subject in the study may experience only two causes of failure, which are labeled as 1 and 2. Use $\tilde{T}_1$ and $\tilde{T}_2$ to denote the potential failure times for causes 1 and 2, respectively. Let $\tilde{T} = \min\{\tilde{T}_1, \tilde{T}_2\}$ and Z be a $p \times 1$-dimensional covariate. Define the cause of failure indicator $\tilde{\delta} = 1$ if $\tilde{T} = \tilde{T}_1$ and $\tilde{\delta} = 2$ if $\tilde{T} = \tilde{T}_2$. For cause j, the cause-specific hazard function is defined as

$$\tilde{\lambda}_j^*(t|z) = \lim_{h\downarrow 0} h^{-1}\mathsf{P}\left(t \leqslant \tilde{T} < t+h, \tilde{\delta} = j \mid \tilde{T} \geqslant t, Z = z\right), \quad j = 1, 2.$$

We suppose that only the cause 1, i.e., $\tilde{T}_1$, is of interest. Then, the cause-specific hazard of interest is $\tilde{\lambda}_1^*(t|z)$.

An AFT model for the interested failure time $\tilde{T}_1$ relates itself with the covariate Z by assuming that

$$\log \tilde{T}_1 = \beta^\top Z + \varepsilon, \tag{1}$$

where β is a p-dimensional regression parameter and ε is a random error with an unspecified continuous distribution independent of Z. It is not difficult to derive that under the model (1), the conditional hazard function of $\tilde{T}_1$ given the covariate Z, i.e., $\lim_{h\downarrow 0} h^{-1}\mathsf{P}(t \leqslant \tilde{T}_1 < t+h \mid \tilde{T}_1 \geqslant t, Z=z)$, equals to $t^{-1}\lambda_\varepsilon\left(\log t - \beta^\top z\right)$, where $\lambda_\varepsilon(t)$ is the hazard function of ε. Here, instead of modeling $\tilde{T}_1$ directly, we impose the following model on the interested cause-specific hazard $\tilde{\lambda}_1^*(t|z)$:

$$\tilde{\lambda}_1^*(t|z) = \frac{1}{t}\lambda_0\left(\log t - \beta^\top z\right), \tag{2}$$

where β is a p-dimensional parameter of interest and $\lambda_0(t)$ is an unspecified hazard function. We treat the model (2) as an AFT model for the cause-specific hazard of interest.

When there exists right censoring, let C be the right censoring time. The observable failure time, denoted by T, is the minimum of $\tilde{T}$ and C, i.e., $T = \min\{\tilde{T}, C\}$. Define the observable cause of failure indicator $\delta = I\{\tilde{T} \leqslant C\}\tilde{\delta}$, where $I\{\cdot\}$ represents the indicator function. It means that $\delta = \tilde{\delta}$ if $\tilde{T}$ is observed, while $\delta = 0$ if $\tilde{T}$ is censored. We assume that given Z, C is conditionally independent of $(\tilde{T}, \tilde{\delta})$. For cause j, the observable cause-specific hazard function with censoring is defined to be $\lambda_j^*(t|z) = \lim_{h\downarrow 0} h^{-1}\mathsf{P}\left(t \leqslant T < t+h, \delta = j \mid T \geqslant t, Z=z\right)$, $j = 1, 2$. Under the conditional independence assumption of C and $(\tilde{T}, \tilde{\delta})$, $\lambda_j^*(t|z) = \tilde{\lambda}_j^*(t|z)$ for $j = 1, 2$.

As mentioned before, the cause of failure of a subject may be unknown and then the specific value for δ is missing. A binary variable R is introduced to identify the missingness of the failure cause. We set $R = 1$ if the value of δ is known and 0 if the value of δ is missing. Similar to some existing literature, we assume that missing cause does not occur for the censored subject, that is, R always takes 1 when $\delta = 0$. To this point, the observation can be summarized as $\{R, T, Z, A, R\delta\}$, where A represents some auxiliary covariates that are not used in the model (2). For a sample consisting of n subjects, the observed data $\{R_i, T_i, Z_i, A_i, R_i\delta_i\}$, $i = 1, \ldots, n$, are treated as independent and identically distributed (i.i.d.) copies of $\{R, T, Z, A, R\delta\}$. For the missing mechanism, we follow the assumption that the cause of failure is MAR, that is,

$$\mathsf{P}\left(R = 1 \mid \delta, \delta > 0, X\right) = \mathsf{P}\left(R = 1 \mid \delta > 0, X\right) = \pi(X), \tag{3}$$

where $X = (T, Z, A)$. Consequently, it is not difficult to see that $\mathsf{P}(R = 1 \mid \delta, X) = \pi(X)I\{\delta > 0\} + I\{\delta = 0\}$.

Finally we define some counting processes for further use. Let $e(\beta) = \log T - \beta^\top Z$, $N(\beta, t) = I\{e(\beta) \leqslant t, \delta = 1\}$ and $Y(\beta, t) = I\{e(\beta) \geqslant t\}$. For $i = 1, \ldots, n$, let $e_i(\beta)$, $N_i(\beta, t)$ and $Y_i(\beta, t)$ be the ith copy of $e(\beta)$, $N(\beta, t)$ and $Y(\beta, t)$, respectively.

3 Estimation and inference procedures

3.1 Rank-based estimating equations for competing risks data

Let β_0 be the true value of the regression coefficients β. We first consider the situation where all the δ_i's are observable. It can be shown that under the proposed competing risks AFT model (2) and the conditional independence of C and $(\tilde{T}, \tilde{\delta})$ given Z, $N_i(\beta_0, t) - \int_{-\infty}^{t} Y_i(\beta_0, u)\lambda_0(u)\mathrm{d}u$ is a zero-mean martingale process with respect to certain appropriate filtration. Motivated by this fact, we propose the following joint rank-based estimating equations for β_0 and $\Lambda_0(t) = \int_{-\infty}^{t} \lambda_0(u)\mathrm{d}u$:

$$\sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta, t)\, Z_i\, (\mathrm{d}N_i(\beta, t) - Y_i(\beta, t)\mathrm{d}\Lambda(t)) = 0, \tag{4}$$

$$\sum_{i=1}^{n} (\mathrm{d}N_i(\beta, t) - Y_i(\beta, t)\mathrm{d}\Lambda(t)) = 0, \quad t \in \mathbb{R}, \tag{5}$$

where $\phi(\beta, t)$ is a possibly data-dependent weight function. Given β fixed, by solving (5) for $d\Lambda(t)$, we obtain a solution $\mathrm{d}\hat{\Lambda}_\beta(t) = \sum_{i=1}^{n} \mathrm{d}N_i(\beta, t) / \sum_{i=1}^{n} Y_i(\beta, t)$. Replacing $\mathrm{d}\Lambda(t)$ in (4) by $\mathrm{d}\hat{\Lambda}_\beta(t)$, we obtain the following estimating equation for β_0

$$S_\phi(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta, t) \left(Z_i - \bar{Z}(\beta, t)\right) \mathrm{d}N_i(\beta, t) = 0,$$

where $\bar{Z}(\beta, t) = \sum_{i=1}^{n} Z_i Y_i(\beta, t) / \sum_{i=1}^{n} Y_i(\beta, t)$. The solution, denoted by $\hat{\beta}_\phi$, is treated as a rank-based estimator for the regression parameter.

The idea of obtaining $\hat{\beta}_\phi$ is similar to that of the weighted log-rank estimator for usual right censored survival data (see, for example, Tsiatis 1990; Ying 1993). Using arguments similar to those in Ying (1993), it can be proved that under suitable regularity conditions, $\hat{\beta}_\phi$ is consistent and $\sqrt{n}(\hat{\beta}_\phi - \beta_0)$ converges in distribution to a zero-mean normal random vector. However, since $S_\phi(\beta)$ is not continuous or monotone in β, the estimating equation $S_\phi(\beta) = 0$ is difficult to solve. Moreover, the variance–covariance matrix of the limiting distribution of $\sqrt{n}(\hat{\beta}_\phi - \beta_0)$ depends on the derivative of $\lambda_0(t)$ and is not easy to estimate. Jin et al. (2003) found that the rank estimator with the Gehan-type weight function can be obtained by minimizing a convex objective function through a standard linear programming technique. The rank estimators with other weight functions are then obtained iteratively by solving a sequence of approximated monotone estimating equations. Each iteration can be executed via linear programming. For inferences, instead of estimating the limiting variance–covariance matrix directly, Jin et al. (2003) proposed a random weighting technique. The methods developed by Jin et al. (2003) can be applied here directly to obtain $\hat{\beta}_\phi$ and make inferences with competing risks data when the cause of failure is observed.

3.2 Inverse probability weighted estimating equations

When the cause of failure is missing, $S_\phi(\beta)$ is no longer available. We apply the inverse probability weighted estimation method which is commonly used in missing data problems. An estimator for the probability of a complete case, $\mathsf{P}(R = 1 \mid \delta, X) = \pi(X)I\{\delta > 0\} + I\{\delta = 0\}$, is necessary. Here, we assume that $\pi(X)$ is decided by a parametric model $\pi(X, \gamma)$, where γ is an r-dimensional unknown parameter. Based on the MAR assumption, γ can be estimated by the maximum likelihood estimator $\hat{\gamma}$ which maximizes the likelihood function

$$\prod_{i=1}^{n} \pi(X_i, \gamma)^{R_i I\{\delta_i > 0\}} (1 - \pi(X_i, \gamma))^{1-R_i}$$

with respect to γ. Let $\tilde{\pi}(\delta, X, \gamma) = \pi(X, \gamma)I\{\delta > 0\} + I\{\delta = 0\}$. We propose the following joint inverse probability weighted rank estimating equations

$$\sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta, t) Z_i \left(\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \mathrm{d}N_i(\beta, t) - Y_i(\beta, t)\mathrm{d}\Lambda(t) \right) = 0, \quad (6)$$

$$\sum_{i=1}^{n} \left(\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \mathrm{d}N_i(\beta, t) - Y_i(\beta, t)\mathrm{d}\Lambda(t) \right) = 0, \quad t \in \mathbb{R}. \quad (7)$$

The justification of the proposed joint estimating equations lies in the fact that $\mathsf{E}[R_i N_i(\beta_0, t)/\tilde{\pi}(\delta_i, X_i, \gamma) - \int_{-\infty}^{t} Y_i(\beta_0, u)\lambda_0(u)\mathrm{d}u] = 0$ for all $t \in \mathbb{R}$ when γ takes the true value. Again, by solving (7) for $\mathrm{d}\Lambda(t)$ with fixed β, we obtain a solution

$$\mathrm{d}\hat{\Lambda}_{\beta}^{\mathrm{IPW}}(t) = \frac{\sum_{i=1}^{n} \frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \mathrm{d}N_i(\beta, t)}{\sum_{i=1}^{n} Y_i(\beta, t)}.$$

Replacing $\mathrm{d}\Lambda(t)$ in (6) by $\mathrm{d}\hat{\Lambda}_{\beta}^{\mathrm{IPW}}(t)$, we get the following inverse probability weighted rank estimating equation for β_0

$$S_{\phi}^{\mathrm{IPW}}(\beta) = \sum_{i=1}^{n} \frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \int_{-\infty}^{\infty} \phi(\beta, t) \left(Z_i - \bar{Z}(\beta, t) \right) \mathrm{d}N_i(\beta, t) = 0. \quad (8)$$

We propose to use the root of (8), denoted by $\hat{\beta}_{\phi}^{\mathrm{IPW}}$, to be the estimator of β_0. We call it weighted log-rank inverse probability weighted estimator.

Similar to $S_\phi(\beta)$, $S_{\phi}^{\mathrm{IPW}}(\beta)$ is not componentwise continuous in β. We develop an algorithm mimicking that of Jin et al. (2003) to get the proposed estimator. We start with Gehan weight which is given by $\phi_G(\beta, t) = n^{-1} \sum_{i=1}^{n} Y_i(\beta, t)$. By some algebra, it can be shown that Gehan inverse probability weighted estimating function $S_{\phi_G}^{\mathrm{IPW}}(\beta)$ equals to

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} I\{\delta_i = 1\}\left(Z_i - Z_j\right) I\{e_i(\beta) \leqslant e_j(\beta)\}.$$

It is not difficult to see that $S_{\phi_G}^{\text{IPW}}(\beta)$ is the gradient in β of the following convex function

$$L_{\phi_G}^{\text{IPW}}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} I\{\delta_i = 1\}\left(e_i(\beta) - e_j(\beta)\right)^{-},$$

where a^{-} stands for $|a|I\{a < 0\}$. Consequently, the proposed $\hat{\beta}_{\phi_G}^{\text{IPW}}$ can be obtained by minimizing $L_{\phi_G}^{\text{IPW}}(\beta)$ with respect to β. The minimization can be implemented by linear programming or some optimization packages in softwares without much difficultly.

For a weight function different from $\phi_G(\beta, t)$, say $\phi(\beta, t)$, define $\psi(\beta, t) = \phi(\beta, t)/\phi_G(\beta, t)$. Instead of solving (8) directly, we consider the modified estimating equation

$$\tilde{S}_{\phi}^{\text{IPW}}(\beta) = \sum_{i=1}^{n}\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \times \int_{-\infty}^{\infty}\psi\left(\hat{\beta}, t+(\beta-\hat{\beta})^{\top}Z_i\right)\phi_G(\beta, t)\left(Z_i-\bar{Z}(\beta, t)\right)\mathrm{d}N_i(\beta, t)=0, \tag{9}$$

where $\hat{\beta}$ is a consistent estimator of β_0. It is not difficult to see that solving (9) is equivalent to minimizing the convex function

$$\tilde{L}_{\phi}^{\text{IPW}}(\beta, \hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})}\psi\left(\hat{\beta}, e_i(\hat{\beta})\right) I\{\delta_i = 1\}\left(e_i(\beta) - e_j(\beta)\right)^{-}.$$

Based on this procedure, an iterative algorithm is developed as follows. First define $\hat{\beta}_{(0)} = \hat{\beta}_{\phi_G}^{\text{IPW}}$. Then for each $k \geqslant 1$, let $\hat{\beta}_{(k)} = \text{argmin}_{\beta}\tilde{L}_{\phi}^{\text{IPW}}(\beta, \hat{\beta}_{(k-1)})$. If the sequence $\hat{\beta}_{(k)}$ converges to a limit as $k \to \infty$, we will use $\hat{\beta}_{(k)}$ as the proposed estimator $\hat{\beta}_{\phi}^{\text{IPW}}$ when certain convergence criterion is met. As pointed by Jin et al. (2003), if there exists a limit for the sequence of $\hat{\beta}_{(k)}$, the limit must satisfy (8). Based on the experience from our simulation study and real data analysis, we find that $\hat{\beta}_{(k)}$ converges as k increases in almost all the cases.

3.3 Double robust estimating equations

Similar to Gao and Tsiatis (2005) and Lu and Liang (2008), a so-called double robust technique developed by Robins et al. (1994) can be applied here to gain robustness against the model misspecification on $\pi(X)$. Specifically, define $\rho(X) = \mathsf{P}(\delta = 1 \mid \delta > 0, X)$ and posit a parametric model for this conditional probability, that is, assume $\rho(X) = \rho(X, \eta)$ where η is an s-dimensional unknown parameter. Under the MAR assumption (3), we have that

$$\rho(X) = \mathsf{P}(\delta = 1 \mid \delta > 0, X) = \mathsf{P}(\delta = 1 \mid R = 1, \delta > 0, X).$$

This implies that $\rho(X)$ can be estimated from the complete cases with $R_i = 1$ and $\delta_i > 0$. Based on the imposed parametric model, a maximum likelihood estimator $\hat{\eta}$ can be obtained by maximizing the likelihood

$$\prod_{i=1}^{n} \rho(X_i, \eta)^{I\{\delta_i=1, R_i=1\}} (1 - \rho(X_i, \eta))^{I\{\delta_i=2, R_i=1\}}$$

with respect to η. Let $\tilde{N}(\beta, t) = I\{e(\beta) \leqslant t, \delta > 0\}$ and $\tilde{N}_i(\beta, t) = I\{e_i(\beta) \leqslant t, \delta_i > 0\}$ for $i = 1, \ldots, n$. We consider the following joint augmented inverse probability weighted rank estimating equations

$$\sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta, t) Z_i \times \left(\frac{R_i \mathrm{d}N_i(\beta, t)}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} - \frac{R_i - \tilde{\pi}(\delta_i, X_i, \hat{\gamma})}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \rho(X_i, \hat{\eta}) \mathrm{d}\tilde{N}_i(\beta, t) - Y_i(\beta, t) \mathrm{d}\Lambda(t) \right) = 0,$$

$$\sum_{i=1}^{n} \left(\frac{R_i \mathrm{d}N_i(\beta, t)}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} - \frac{R_i - \tilde{\pi}(\delta_i, X_i, \hat{\gamma})}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \rho(X_i, \hat{\eta}) \mathrm{d}\tilde{N}_i(\beta, t) - Y_i(\beta, t) \mathrm{d}\Lambda(t) \right) = 0, \ \ t \in \mathbb{R}.$$

The justification of the proposed joint estimating equations comes from Proposition 1 presented in Sect. 3.4. Similar calculation brings the augmented inverse probability weighted rank estimating equation

$$S_{\phi}^{\mathrm{DR}}(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta, t) \left(Z_i - \bar{Z}(\beta, t) \right) \times \left(\frac{R_i dN_i(\beta, t)}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} - \frac{R_i - \tilde{\pi}(\delta_i, X_i, \hat{\gamma})}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})} \rho(X_i, \hat{\eta}) \mathrm{d}\tilde{N}_i(\beta, t) \right) = 0. \quad (10)$$

Compared with $S_{\phi}^{\mathrm{IPW}}(\beta)$, the additional term involving $\rho(X_i, \hat{\eta})$ in $S_{\phi}^{\mathrm{DR}}(\beta)$ is the augmentation part which could bring robustness against misspecified $\pi(X_i, \gamma)$. Denote the root of (10) by $\hat{\beta}_{\phi}^{\mathrm{DR}}$ and we call it weighted log-rank double robust estimator. The meaning of double robustness is discussed in more detail in Sect. 3.4.

Solving (10) is not easy since $S_{\phi}^{\mathrm{DR}}(\beta)$ again is not continuous in β. We still transform this equation solving problem to an optimization problem and start with the Gehan weight function ϕ_G. By some careful calculation, we find that $S_{\phi_G}^{\text{DR}}(\beta)$ is the gradient in β of the following function

$$L_{\phi_G}^{\text{DR}}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{R_i I\{\delta_i=1\}}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})}-\frac{R_i-\tilde{\pi}(\delta_i, X_i, \hat{\gamma})}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})}\rho(X_i, \hat{\eta})\right)\left(e_i(\beta)-e_j(\beta)\right)^{-}.$$

The Gehan weight double robust estimator can be obtained by minimizing $L_{\phi_G}^{\text{DR}}(\beta)$ with respect to β. It should be pointed out that although $L_{\phi_G}^{\text{DR}}(\beta)$ has a similar form as $L_{\phi_G}^{\text{IPW}}(\beta)$, the minimization of the former one could not be transformed to a linear programming directly. Some well-designed optimization packages in various softwares may be useful to minimize the function. In our numeric studies, we use the optimization package 'fminsearch' in Matlab to implement the minimization. It works reasonably well for moderate sample sizes. Finally, for arbitrary weight function ϕ, an iterative algorithm similar to that developed in Sect. 3.2 can be obtained to get $\hat{\beta}_{\phi}^{\text{DR}}$.

3.4 Large sample properties

In this subsection, we concentrate on discussing the asymptotic properties of $\hat{\beta}_{\phi}^{\text{DR}}$. The large sample properties of $\hat{\beta}_{\phi}^{\text{IPW}}$ can be obtained similarly. In the following discussion, we assume the regularity conditions listed in the "Appendix 7" hold. Some more notation are also needed. Define

$$\begin{aligned}\tilde{M}(t, \beta, \gamma, \eta, \Lambda) &= \frac{RN(\beta, t)}{\tilde{\pi}(\delta, X, \gamma)} - \frac{R-\tilde{\pi}(\delta, X, \gamma)}{\tilde{\pi}(\delta, X, \gamma)}\rho(X, \eta)\tilde{N}(\beta, t)\\ &\quad - \int_{-\infty}^{t} Y(\beta, u)\mathrm{d}\Lambda(u),\end{aligned}$$

and $A_{\phi} = \mathsf{E}[\int_{-\infty}^{\infty}\overline{\phi}(t)Y(\beta_0, t)(Z-\bar{z}(t))^{\otimes 2}\dot{\lambda}_0(t)\mathrm{d}t]$, where $\overline{\phi}(t)$ is the limit in probability of $\phi(\beta_0, t)$, $\bar{z}(t) = \mathsf{E}[ZY(\beta_0, t)]/\mathsf{E}[Y(\beta_0, t)]$, $a^{\otimes 2}$ means $aa^{\top}$ for any column vector a, and $\dot{\lambda}_0$ stands for the derivative of λ_0. By introducing suitable regularity conditions on the parametric models $\pi(X, \gamma)$ and $\rho(X, \eta)$, one may show that $\hat{\gamma}$ and $\hat{\eta}$ have limits no matter if the models hold for the data or not. Let γ^* and η^* denote the two limits. When the parametric models are correctly specified, the limits become the true values of the corresponding parameters. Moreover, let I_{γ} and S_{γ} denote the information matrix and the score vector for $\hat{\gamma}$ evaluated at γ^*, and I_{η} and S_{η} denote the counterparts for $\hat{\eta}$ evaluated at η^*, respectively.

The double robustness of $\hat{\beta}_{\phi}^{\text{DR}}$ means that if either $\pi(X, \gamma)$ or $\rho(X, \eta)$ is correctly specified, $\hat{\beta}_{\phi}^{\text{DR}}$ is weakly consistent. This can be proved based on the following proposition.

Proposition 1 *If either the parametric model $\pi(X, \gamma)$ or $\rho(X, \eta)$ is correctly specified, we have that $\mathsf{E}[\tilde{M}(t, \beta_0, \gamma^*, \eta^*, \Lambda_0)] = 0$ for all $t \in \mathbb{R}$.*

The proof of Proposition 1 is given in the "Appendix 7". By combining the result of Proposition 1 and the arguments in Ying (1993) and Gao and Tsiatis (2005), one can establish the weak consistency of the proposed estimator. We omit the details here.

The following proposition which is proved in the "Appendix 7" formally presents the asymptotic distribution of the proposed estimator.

Proposition 2 *Under the conditions listed in the "Appendix 7", if either the parametric model $\pi(X,\gamma)$ or $\rho(X,\eta)$ is correctly specified, we have that $\sqrt{n}(\hat{\beta}_{\phi}^{\mathrm{DR}}-\beta_0)$ is asymptotically normally distributed with mean zero and variance–covariance matrix*

$$A_{\phi}^{-1}\mathsf{E}\left[\left(\int_{-\infty}^{\infty}\overline{\phi}(t)\,(Z-\bar{z}(t))\,d\tilde{M}(t,\beta_0,\gamma^*,\eta^*,\Lambda_0)-B_{\gamma}I_{\gamma}^{-1}S_{\gamma}-B_{\eta}I_{\eta}^{-1}S_{\eta}\right)^{\otimes 2}\right]A_{\phi}^{-1},$$

where B_γ and B_η are quantities given in the "Appendix 7".

The double robustness is also reflected by Proposition 2 in the sense that $\hat{\beta}_{\phi}^{\mathrm{DR}}$ is asymptotically normal when either $\pi(X,\gamma)$ or $\rho(X,\eta)$ is the right model.

Finally, the asymptotic distribution of $\hat{\beta}_{\phi}^{\mathrm{IPW}}$ is given by the following proposition.

Proposition 3 *Under the conditions listed in the "Appendix 7", if the parametric model $\pi(X,\gamma)$ is correctly specified, we have that $\sqrt{n}(\hat{\beta}_{\phi}^{\mathrm{IPW}}-\beta_0)$ is asymptotically normally distributed with mean zero and variance–covariance matrix*

$$A_{\phi}^{-1}\mathsf{E}\left[\left(\int_{-\infty}^{\infty}\overline{\phi}(t)\,(Z-\bar{z}(t))\,\mathrm{d}M(t)-C_{\gamma}I_{\gamma}^{-1}S_{\gamma}\right)^{\otimes 2}\right]A_{\phi}^{-1},$$

where

$$M(t)=\frac{RN(\beta_0,t)}{\tilde{\pi}(\delta,X,\gamma^*)}-\int_{-\infty}^{t}Y(\beta_0,u)\lambda_0(u)\mathrm{d}u,$$

and C_γ is a quantity given in the "Appendix 7".

Note that the consistency and asymptotic normality of $\hat{\beta}_{\phi}^{\mathrm{IPW}}$ depend on the correct model specification for $\pi(X)$.

3.5 Inference procedures

Based on the derived large properties, inferences can be made about the regression coefficients. However, similar to the situation without missingness, the matrix A_ϕ in the limiting variance–covariance matrix is not easy to estimate. As we have mentioned, when there is no missing failure indicators, Jin et al. (2003) proposed a random weighting technique in which the objective function was perturbed by a positive random variable with mean 1 and variance 1 for many times to get the perturbed estimates. The perturbed estimates were shown to have the same asymptotic distributions as the weighted log-rank estimators given the observed data. As a result the inference could be implemented with the perturbed estimates.

The random weighting procedure is easy to follow. However, the technique cannot be applied in our case directly. One reason is that the proposed estimating function $S_{\phi}^{\mathrm{DR}}(\beta)$ in (10) contains estimators $\hat{\gamma}$ and $\hat{\eta}$. Thus, here we turn to a nonparametric bootstrap method. Specifically, we sample data with sample size n from the empirical distribution of the original sample $\{R_i, T_i, Z_i, A_i, R_i\delta_i\}$, $i = 1, \ldots, n$. For each resampled data, we calculate the weighted log-rank inverse probability estimator or the weighted log-rank double robust estimator by our proposed algorithm. The resampling procedure could be replicated for a large number (say, M) of times, and M estimates for the regression parameter would be obtained. We use the empirical distribution of the M resampled estimates to calculate the desired standard errors of the proposed estimators and consequently carry out the inferences.

Naturally, the nonparametric bootstrap requires intensive computation. Thus, we could consider turning to a score test especially for the global inference problem $\mathsf{H}_0 : \beta_0 = b$ v.s. $\mathsf{H}_1 : \beta_0 \neq b$, where $b \in \mathbb{R}^p$ is a p-dimensional constant vector to test for. To develop the test statistic, for each $i = 1, \ldots, n$, define

$$g_{\phi,i}^{\mathrm{DR}}(\beta) = \int_{-\infty}^{\infty} \phi(\beta, t)\left(Z_i - \bar{Z}(\beta, t)\right) \mathrm{d}\tilde{M}_i(t, \beta, \hat{\gamma}, \hat{\eta}, \hat{\Lambda}_{\beta}^{\mathrm{DR}}) - \hat{B}_{\gamma}(\beta)\hat{I}_{\gamma}^{-1}\hat{S}_{\gamma,i} - \hat{B}_{\eta}(\beta)\hat{I}_{\eta}^{-1}\hat{S}_{\eta,i},$$

where

$$\tilde{M}_i(t, \beta, \gamma, \eta, \Lambda) = \frac{R_i N_i(\beta, t)}{\tilde{\pi}(\delta_i, X_i, \gamma)} - \frac{R_i - \tilde{\pi}(\delta_i, X_i, \gamma)}{\tilde{\pi}(\delta_i, X_i, \gamma)}\rho(X_i, \eta)\tilde{N}_i(\beta, t) - \int_{-\infty}^{t} Y_i(\beta, u)\mathrm{d}\Lambda(u),$$

$$\hat{\Lambda}_{\beta}^{\mathrm{DR}}(t) = \int_{-\infty}^{t} \frac{\sum_{i=1}^{n}\left(\frac{R_i}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})}\mathrm{d}N_i(\beta, u) - \frac{R_i - \tilde{\pi}(\delta_i, X_i, \hat{\gamma})}{\tilde{\pi}(\delta_i, X_i, \hat{\gamma})}\rho(X_i, \hat{\eta})\mathrm{d}\tilde{N}_i(\beta, u)\right)}{\sum_{i=1}^{n} Y_i(\beta, u)},$$

$\hat{B}_{\gamma}(\beta)$ and $\hat{B}_{\eta}(\beta)$ are the estimators obtained by substituting sample average for expectation and $(\hat{\gamma}, \hat{\eta}, \hat{\Lambda}_{\beta}^{\mathrm{DR}})$ for $(\gamma^*, \eta^*, \Lambda_0)$ in B_{γ} and B_{η}, $\hat{I}_{\gamma}$ and $\hat{I}_{\eta}$ are the sample information matrices, and $\hat{S}_{\gamma,i}$ and $\hat{S}_{\eta,i}$ are the sample score vectors for the ith observation evaluated at $(\hat{\gamma}, \hat{\eta})$. Define

$$\hat{V}_{\phi}^{\mathrm{DR}}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_{\phi,i}^{\mathrm{DR}}(\beta)^{\otimes 2}.$$

It can be shown that $\hat{V}_{\phi}^{\mathrm{DR}}(\beta_0)$ is a consistent estimator of

$$\mathsf{E}\left[\left(\int_{-\infty}^{\infty} \overline{\phi}(t)\left(Z - \bar{z}(t)\right)\mathrm{d}\tilde{M}(t, \beta_0, \gamma^*, \eta^*, \Lambda_0) - B_{\gamma}I_{\gamma}^{-1}S_{\gamma} - B_{\eta}I_{\eta}^{-1}S_{\eta}\right)^{\otimes 2}\right]$$

which is the variance–covariance matrix of $n^{-1/2} S_\phi^{\mathrm{DR}}(\beta_0)$. Define the score test statistic $n^{-1} S_\phi^{\mathrm{DR}}(b)^\top \hat{V}_\phi^{\mathrm{DR}}(b)^{-1} S_\phi^{\mathrm{DR}}(b)$. It is not difficult to see that under the assumptions of Proposition 2, when $\mathsf{H}_0 : \beta_0 = b$ holds, the proposed test statistic converges in distribution to a Chi-squared random variable with p degrees of freedom. The result can be used to do the test as well as construct a score-based confidence region for β_0. Using the same idea, it is also easy to develop score test based on $S_\phi^{\mathrm{IPW}}(\beta)$.

4 Simulation studies

Two sets of simulation studies are carried out to assess the finite sample performance of the proposed estimators. In the first set of simulation, we consider two random number generating schemes:

SCHEME 1: We choose an AFT model with two covariates for $\tilde{T}_1$, i.e., $\log \tilde{T}_1 = \alpha_1 Z_1 + \alpha_2 Z_2 + \varepsilon$, where Z_1 is generated from a Bernoulli distribution with success probability 0.5, Z_2 is generated from a uniform distribution on (0, 1), ε follows the standard extreme value distribution and all the variables are independent with each other. α_1 and α_2 are set to be 1 and -1, respectively. Given covariates, the failure time of competing risk $\tilde{T}_2$ is generated independently of $\tilde{T}_1$, following a Gompertz distribution with conditional hazard, denoted by $\lambda_2(t|Z_1, Z_2)$, equals to $\lambda \exp(\xi_0 + \xi_1 t)$ with $\lambda = 1$, $\xi_0 = -1.5$ and $\xi_1 = 0.5$. Note that when $\tilde{T}_1$ and $\tilde{T}_2$ are conditionally independent, model (1) implies that (2) holds with $\beta_1 = \alpha_1$ and $\beta_2 = \alpha_2$. The censoring time C follows a uniform distribution on $(0.5, 4)$. The missing indicator R for cause of failure is generated from a logistic model given by $\log[\pi(X)/(1 - \pi(X))] = \gamma_0 + \gamma_1 T + \gamma_2 Z_1 + \gamma_3 Z_2$, with $\gamma_0 = 1.5$, $\gamma_1 = -1$, $\gamma_2 = -1$ and $\gamma_3 = -1$. Under the above settings, $\rho(X)$ follows a logistic model of the form $\log[\rho(X)/(1-\rho(X))] = -\xi_0 - \log\lambda - \xi_1 T - \alpha_1 Z_1 - \alpha_2 Z_2$. One would expect that about 66 % failures are from the cause of interest, 22 % are from the cause of competing risk and the remaining 12 % are censored observations. Furthermore, among the failures, one would expect 52 % missing cause of failure.

SCHEME 2: Let $\tilde{T} = \min\{\tilde{T}_1, \tilde{T}_2\}$ follows an AFT model with two covariates, i.e., $\log \tilde{T} = \alpha_1 Z_1 + \alpha_2 Z_2 + \varepsilon$, where Z_1 and Z_2 follow the same distribution as those in scheme 1, and ε follows the standard logistic distribution. We set $\alpha_1 = -0.5$ and $\alpha_2 = 1$. Let $\mathsf{P}(\tilde{\delta} = 1 \mid \tilde{T}, Z_1, Z_2) = 0.7$. The scheme implies that model (2) holds with $\beta_1 = \alpha_1$ and $\beta_2 = \alpha_2$. Also, under this scheme, $\tilde{T}_1$ and $\tilde{T}_2$ can be correlated with each other given the covariates. The censoring time C and the missing indicator R follow the same distribution as those in scheme 1. The above settings result in $\rho(X) = 0.7$. One would expect about 54 % failures from the cause of interest, 23 % from the cause of competing risk and 23 % censoring. Among the observed failures, there are about 56 % missing cause of failure.

The sample size n is chosen to be 200. For the weight function ϕ in $S_\phi^{\mathrm{IPW}}(\beta)$ and $S_\phi^{\mathrm{DR}}(\beta)$, we choose the Gehan weight and log-rank weight $\phi_L(\beta, t) = 1$. Thus, we have four pairs of rank estimators $(\hat{\beta}_{\phi_G,1}^{\mathrm{IPW}}, \hat{\beta}_{\phi_G,2}^{\mathrm{IPW}})$, $(\hat{\beta}_{\phi_G,1}^{\mathrm{DR}}, \hat{\beta}_{\phi_G,2}^{\mathrm{DR}})$, $(\hat{\beta}_{\phi_L,1}^{\mathrm{IPW}}, \hat{\beta}_{\phi_L,2}^{\mathrm{IPW}})$

Table 1 Simulation results of complete-case, inverse probability weighted and double robust rank-based estimators with correctly specified parametric models

Method	Weight (ϕ)	Parameter	Scheme 1				Scheme 2			
			Bias	SD	BSE	CP	Bias	SD	BSE	CP
C-C	Gehan	β_1	0.260	0.365	0.358	0.909	−0.005	0.421	0.430	0.952
		β_2	−0.064	0.571	0.550	0.942	0.801	0.734	0.757	0.847
	Log-rank	β_1	0.374	0.344	0.336	0.789	−0.023	0.422	0.422	0.947
		β_2	−0.062	0.506	0.494	0.938	1.056	0.695	0.724	0.691
IPW	Gehan	β_1	0.029	0.264	0.265	0.953	−0.005	0.288	0.299	0.955
		β_2	−0.036	0.454	0.455	0.953	0.061	0.507	0.550	0.964
	Log-rank	β_1	0.050	0.272	0.272	0.952	−0.006	0.293	0.299	0.947
		β_2	−0.018	0.407	0.409	0.947	0.085	0.534	0.558	0.959
DR	Gehan	β_1	0.034	0.255	0.259	0.952	−0.006	0.274	0.289	0.960
		β_2	−0.032	0.426	0.431	0.958	0.043	0.487	0.539	0.969
	Log-rank	β_1	0.040	0.270	0.270	0.942	−0.008	0.268	0.275	0.959
		β_2	−0.014	0.402	0.403	0.939	0.032	0.480	0.509	0.963

C-C complete-case estimate, *IPW* inverse probability weighted estimate, *DR* double robust estimate, *Bias simulated bias of the estimates*, *SD* simulated standard deviation of the estimates, *BSE* average of estimated standard error using the bootstrap method, *CP* empirical coverage probabilities of 95 % Wald-type confidence intervals based on bootstrap standard errors

and $(\hat{\beta}^{\text{DR}}_{\phi_L,1}, \hat{\beta}^{\text{DR}}_{\phi_L,2})$. Besides the proposed rank estimators, we also calculate the naive complete-case estimators which are only based on the observations with complete cause of failure. In this set of simulation, we use the correct parametric models for $\pi(X)$ and $\rho(X)$ under both schemes. One thousand replications of the random numbers are generated. For each replication, we resample from the generated data for 250 times (i.e., $M = 250$) to obtain the bootstrap estimates for the standard errors. In Table 1, for each estimator, we report the simulated biases, the simulated standard errors, the averages of the bootstrap estimates for standard errors and the empirical coverage probabilities of the 95 % Wald-type confidence interval.

Under both schemes, since the missing mechanism is MAR, the complete-case estimators are biased in some cases. By contrast, the proposed rank estimators are essentially unbiased in almost all the cases. The averages of the bootstrap estimates for standard error are in general close to the simulated standard errors of the rank estimators. The Wald-type confidence intervals based on the normal approximation give out appropriate coverage rates. Moreover, the double robust estimators are slightly more efficient than the inverse probability weighted ones.

We use the proposed score test statistic based on $S^{\text{DR}}_{\phi}(\beta)$ for global testing of the regression parameter vector. The Gehan weight and log-rank weight are considered. It is easy to see that the test statistics should be calibrated by the quantiles of the Chi-squared distribution with degrees of freedom 2. The nominal level is set to be 0.05. Five thousands replications of the random numbers are generated and the score test statistic is quick to calculate. Under scheme 1, to test for $\mathsf{H}_0 : \beta_1 = 1, \beta_2 = -1$,

Table 2 Simulation results of inverse probability weighted and double robust rank-based estimators with misspecified parametric models

	Weight (ϕ)	Method	IPW				DR			
		Parameter	Bias	SD	BSE	CP	Bias	SD	BSE	CP
Case 1	Gehan	β_1	0.693	0.376	0.364	0.540	0.032	0.258	0.254	0.942
		β_2	0.312	0.579	0.455	0.905	−0.024	0.417	0.409	0.950
	Log-rank	β_1	0.811	0.355	0.348	0.337	0.033	0.251	0.247	0.937
		β_2	0.294	0.532	0.517	0.899	−0.058	0.380	0.367	0.946
Case 2	Gehan	β_1	0.029	0.264	0.265	0.953	0.029	0.257	0.272	0.955
		β_2	−0.036	0.454	0.455	0.953	−0.035	0.442	0.457	0.956
	Log-rank	β_1	0.050	0.272	0.272	0.952	0.041	0.282	0.265	0.949
		β_2	−0.018	0.407	0.409	0.947	−0.010	0.438	0.402	0.949
Case 3	Gehan	β_1	0.693	0.376	0.364	0.540	0.247	0.261	0.254	0.861
		β_2	0.312	0.579	0.455	0.905	0.112	0.437	0.428	0.923
	Log-rank	β_1	0.811	0.355	0.348	0.337	0.239	0.230	0.228	0.838
		β_2	0.294	0.532	0.517	0.899	0.125	0.392	0.379	0.926

IPW inverse probability weighted estimate, *DR* double robust estimate, *Bias* simulated bias of the estimates, *SD* simulated standard deviation of the estimates, *BSE* average of estimated standard error using the bootstrap method. *CP* empirical coverage probabilities of 95 % Wald-type confidence intervals based on bootstrap standard errors

the empirical size of the score test statistic with Gehan weight function is 0.057, and 0.08 with log-rank weight. Under scheme 2, to test for $\mathsf{H}_0 : \beta_1 = -0.5, \beta_2 = 1$, the empirical size with Gehan weight is 0.061, and 0.075 with log-rank weight.

In the second set of simulation, we consider three cases with misspecified parametric models, using the random number generating scheme 1. In case 1, we use a constant to estimate $\pi(X)$, which means that the parametric model for $\pi(X)$ is incorrectly specified. Meanwhile, the parametric model for $\rho(X)$ is kept correctly specified. In case 2, a constant is used to estimated $\rho(X)$ and the model for $\pi(X)$ is correctly specified. In case 3 both $\pi(X)$ and $\rho(X)$ are estimated by constants, implying neither model is correctly specified. The sample size is 200. One thousand replications of the random numbers are generated. For each replication, we resample from the generated sample for 250 times to obtain the bootstrap estimates for the standard errors. We consider the inverse probability weighted and double robust rank estimators with the Gehan and log-rank wight, respectively. The simulation results are summarized in Table 2.

In cases 1 and 2 where one of the parametric models is correctly specified, we find that the double robust rank estimators are unbiased and the bootstrap approach gives out adequate coverage probabilities. However, in case 1 where $\pi(X)$ is incorrectly specified, the inverse probability weighted rank estimators are obviously biased. In case 3, when both the models are misspecified, it is expectable that the two types of the proposed rank estimators are biased. For our simulation settings, the bias for the double robust rank estimators are relatively smaller. This observation is quite similar to that mentioned in Gao and Tsiatis (2005).

5 A real example

Here, we apply the proposed approach to a data set from a bone marrow transplant study described by Sierra et al. (2002). The study involved 452 primary myelodysplasia patients who received transplants from HLA-identical siblings and were registered with the International Bone Marrow Transplant Registry. There were two competing risks in this study. One was treatment-related death defined as death in complete remission and the other was relapse defined as recurrence of myelodysplasia. The former was the cause of interest. The age and the platelet before transplantation of each patient were regarded as covariates. In our analysis, 408 patients with complete covariate information obtained from the 'timereg' package for R are considered. Among these patients, 161 died in complete remission, 87 relapsed and the rest were censored. The covariate age is centered by its sample mean and the platelet is categorized into two levels with 1 for more than 100×10^9 and 0 for less. Let T be the observed failure time, δ the cause of failure (with 1 for death in remission, 2 for relapse and 0 for censoring), and Z_a and Z_p the centered age and the categorized platelet, respectively.

The causes of failure of all patients in the original data are known for each patient. To illustrate our method, we artificially delete some causes of failure among those $\delta = 1$ or 2. Three missing mechanisms are considered. The first one is missing completely at random (MCAR) with missing probability 23 %. The second one is MAR with the missing probability following the logistic model $\log[\pi(X)/(1-\pi(X))] = 0.5 + T - Z_a$, where $X = (T, Z_a, Z_p)$. The third one is non-ignorable missing (NMAR) with the logistic model $\log[\pi(X)/(1-\pi(X))] = 0.5 + T - Z_a - I\{\delta = 1\}$. The missing probability is around 23 % under MAR and around 26 % under NMAR. We use the logistic models $\log[\pi(X,\gamma)/(1-\pi(X,\gamma))] = \gamma_0 + \gamma_1 T + \gamma_2 Z_a + \gamma_3 Z_p$ and $\log[\rho(X,\eta)/(1-\rho(X,\eta))] = \eta_0 + \eta_1 T + \eta_2 Z_a + \eta_3 Z_p$ to estimate $\pi(X)$ and $\rho(X)$, respectively. We fit the data under the AFT model (2) with two covariates Z_a and Z_p. Under each mechanism, we consider the complete-case, inverse probability weighted and double robust estimators with Gehan and log-rank weights. We also calculate the Gehan and log-rank estimators for the two regression coefficients with full data for comparison. The results are summarized in Table 3.

From the results of the full data estimation, we find that both covariates have significant effect on the interested cause-specific hazard of death in remission, with the age increasing the hazard while the higher platelet level decreasing the hazard, under both weight functions. Under MCAR, all the three methods adjusting for missing cause of failure have comparable performances. Under MAR, the complete-case estimators are obviously biased with smaller absolute estimated values. By comparison, the proposed inverse probability weighted and double robust estimators are quite close to their full data counterparts. The estimated standard errors are a little larger than those of the full data estimators because of missingness. Under NMAR, although the assumption of our approach is violated, the proposed estimators still give out quite robust performances in the sense that they are closer to the corresponding full data estimators than the complete-case ones. Finally, it should be mentioned that Hyun et al. (2012) analyzed the same data by assuming proportional hazards model for the interested cause-specific hazard. We use the same missing mechanisms as theirs, and our findings are similar to theirs.

Table 3 Analysis results for the bone marrow transplant data

Missing			Age			Platelet		
Mechanism	Weight (ϕ)	Method	EST	SE	p Value	EST	SE	p-Value
Full	Gehan		−0.883	0.174	<0.001	1.480	0.412	<0.001
	Log-rank		−1.082	0.202	<0.001	1.620	0.502	0.001
MCAR	Gehan	C-C	−1.087	0.201	<0.001	1.521	0.440	<0.001
		IPW	−0.928	0.194	<0.001	1.495	0.417	<0.001
		DR	−0.926	0.188	<0.001	1.446	0.411	<0.001
	Log-rank	C-C	−1.285	0.224	<0.001	1.562	0.584	0.007
		IPW	−1.127	0.223	<0.001	1.567	0.528	0.003
		DR	−1.124	0.214	<0.001	1.548	0.523	0.003
MAR	Gehan	C-C	−0.774	0.194	<0.001	1.276	0.386	<0.001
		IPW	−0.896	0.199	<0.001	1.452	0.390	<0.001
		DR	−0.968	0.201	<0.001	1.521	0.412	<0.001
	Log-rank	C-C	−0.939	0.222	<0.001	1.368	0.486	0.005
		IPW	−1.098	0.223	<0.001	1.609	0.467	<0.001
		DR	−1.150	0.214	<0.001	1.666	0.499	<0.001
NMAR	Gehan	C-C	−0.661	0.186	<0.001	1.292	0.422	0.002
		IPW	−0.851	0.205	<0.001	1.590	0.413	<0.001
		DR	−0.940	0.199	<0.001	1.647	0.414	<0.001
	Log-rank	C-C	−0.802	0.233	<0.001	1.361	0.522	0.009
		IPW	−1.027	0.231	<0.001	1.730	0.525	<0.001
		DR	−1.109	0.226	<0.001	1.779	0.468	<0.001

Full Full data estimation, *EST* estimate of regression coefficient, *SE* estimate of standard error, *C-C* complete-case estimate, *IPW* inverse probability weighted estimate, *DR* double robust estimate

6 Concluding remarks

The AFT model provides an attractive alternative to hazard-based semiparametric models for analyzing competing risks data. When the cause of failure is subject to missing and the missing mechanism is MAR, we propose two modified rank-based estimating equations with the inverse probability weighted and double robust technique to get estimators for the regression coefficients. Because the proposed estimating functions are not continuous in the regression parameters, the problem of solving the proposed estimating equations is transformed into proper optimization problem. The desired estimators can be obtained by minimizing carefully designed objective functions. Under mild conditions, the proposed estimators are shown to be consistent and asymptotically normal. Especially, the double robust estimator may be still consistent and asymptotically normal even when the missing probability is incorrectly modeled. When developing inference procedures, to overcome the difficulty in directly estimating the limiting variance–covariance matrix, a nonparametric bootstrap method is devised to estimate the standard errors. We also develop a score test for global inference about the regression parameters. Moreover, all the techniques developed can be extended to deal with the competing risks free right censored survival data with missing censoring indicators under the MAR assumption.

For the missing data analysis problem, it is sometimes of interest to derive the semiparametric efficient influence function for the parameter of interest via double robust approach. It is well known that for the case of complete cause of failure, the efficient influence function for the regression parameters under the AFT model can be derived and depends on the derivative of $\lambda_0(t)$. When there is missing cause of failure, deriving the efficient influence function needs further investigation.

7 Appendix

We first give out the formulas of B_γ, B_η and C_γ:

$$B_\gamma = \mathsf{E}\left[\int_{-\infty}^{\infty} \overline{\phi}(t)\,(Z-\overline{z}(t))\,\frac{R\dot{\pi}_\gamma(X,\gamma^*)^\top}{\pi(X,\gamma^*)^2}\left(\mathrm{d}N(\beta_0,t)-\rho(X,\eta^*)\mathrm{d}\tilde{N}(\beta_0,t)\right)\right],$$
$$B_\eta = \mathsf{E}\left[\int_{-\infty}^{\infty} \overline{\phi}(t)\,(Z-\overline{z}(t))\,\frac{R-\pi(X,\gamma^*)}{\pi(X,\gamma^*)}\dot{\rho}_\eta(X,\eta^*)^\top\mathrm{d}\tilde{N}(\beta_0,t)\right]$$

and

$$C_\gamma = \mathsf{E}\left[\int_{-\infty}^{\infty} \overline{\phi}(t)\,(Z-\overline{z}(t))\,\frac{R\dot{\pi}_\gamma(X,\gamma^*)^\top}{\pi(X,\gamma^*)^2}\mathrm{d}N(\beta_0,t)\right],$$

where $\dot{\pi}_\gamma(X,t)=\partial\pi(X,\gamma)/\partial\gamma$ and $\dot{\rho}_\eta(X,t)=\partial\rho(X,\eta)/\partial\eta$.

We assume the following conditions to prove the properties in Sect. 3.4:

C1. The covariates Z have bounded support.
C2. The density function corresponding to $\lambda_0(t)$, denoted by $f(t)$, and its derivative $\dot{f}(t)$ are bounded and $\int(\dot{f}(t)/f(t))^2 f(t)\mathrm{d}t<\infty$.
C3. The failure times $\tilde{T}_1$, $\tilde{T}_2$ and the censoring time C have uniformly bounded densities.
C4. The weight function ϕ satisfies Condition 5 in Ying (1993).
C5. The maximum likelihood estimators $\hat{\gamma}$ and $\hat{\eta}$ satisfy $\sqrt{n}\|\hat{\gamma}-\gamma^*\|=O_p(1)$ and $\sqrt{n}\|\hat{\eta}-\eta^*\|=O_p(1)$, where $\|\cdot\|$ stands for the Euclidean norm.
C6. $\pi(X,\gamma)$ is uniformly bounded away from 0.

Conditions *C1–C4* are similar to Condition 1–Condition 5 introduced by Ying (1993). These conditions are sufficient to guarantee the consistency and asymptotic normality for the weighted log-rank estimators without missing cause of failure. Condition *C5* will be satisfied if one imposes some regularity conditions on the parametric models $\pi(X,\gamma)$ and $\rho(X,\eta)$. The final condition is usually imposed to ensure that there is no 0 quantity in the denominator of the proposed estimating functions.

Proof of Proposition 1 Some algebra calculation yields that

$$\tilde{M}(t, \beta_0, \gamma^*, \eta^*, \Lambda_0) = \left(N(t, \beta_0) - \int_{-\infty}^{t} Y(u, \beta_0) \mathrm{d}\Lambda_0(u)\right) + \left(\frac{R - \pi(X, \gamma^*)}{\pi(X, \gamma^*)}\right) \left(I\{\delta = 1\} - \rho(X, \eta^*)\right) \tilde{N}(\beta_0, t). \quad (11)$$

It is easy to see that $N(t, \beta_0) - \int_{-\infty}^{t} Y(u, \beta_0) \mathrm{d}\Lambda_0(u)$ is a zero mean martingale process. When $\pi(X, \gamma)$ is correctly specified, we have that $\mathsf{E}[R \mid \delta, X] = \pi(X, \gamma^*)$ under MAR. Thus,

$$\mathsf{E}\left[\left(\frac{R - \pi(X, \gamma^*)}{\pi(X, \gamma^*)}\right) \left(I\{\delta = 1\} - \rho(X, \eta^*)\right) \tilde{N}(\beta_0, t) \,\middle|\, \delta, X\right] = \frac{\mathsf{E}[R|\delta, X] - \pi(X, \gamma^*)}{\pi(X, \gamma^*)} \left(I\{\delta = 1\} - \rho(X, \eta^*)\right) \tilde{N}(\beta_0, t) = 0,$$

resulting in that the second term of the right-hand side in (11) has zero mean. On the other hand, if $\rho(X_i, \eta)$ is correctly specified, we have that $\mathsf{E}[I\{\delta = 1\}|\delta > 0, R_1, X_1] = \rho(X, \eta^*)$ under MAR. Thus,

$$\mathsf{E}\left[\left(\frac{R - \pi(X, \gamma^*)}{\pi(X, \gamma^*)}\right) \left(I\{\delta = 1\} - \rho(X, \eta^*)\right) \tilde{N}(\beta_0, t) \,\middle|\, R, X\right] = \left(\frac{R - \pi(X, \gamma^*)}{\pi(X, \gamma^*)}\right) \tilde{N}(\beta_0, t) \mathsf{E}\left[(\mathsf{E}[I\{\delta = 1\}|\delta > 0, R, X] - \rho(X, \eta^*)) I\{\delta = 1\}|R, X\right] = 0,$$

resulting in that the second term of the right-hand side in (11) has zero mean. Consequently, if either $\pi(X, \gamma)$ or $\rho(X, \eta)$ is correctly specified, the desired conclusion follows. □

Proof of Proposition 2 The proof mainly consists of two parts. The first part is to prove the asymptotic normality of $n^{-1/2} S_\phi^{\mathrm{DR}}(\beta_0)$. The second part is to show the asymptotic linearity, that is, for any sequence $\mathrm{d}_n \to 0$,

$$\sup_{\|\beta - \beta_0\| < \mathrm{d}_n} \left\{\|S_\phi^{\mathrm{DR}}(\beta) - S_\phi^{\mathrm{DR}}(\beta_0) - A_\phi(\beta - \beta_0)\| / (\sqrt{n} + n\|\beta - \beta_0\|)\right\} = o_p(1). \quad (12)$$

We first prove the first part. Define

$$\tilde{S}_{\phi}^{\mathrm{DR}}(\beta,\gamma,\eta)=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\phi(\beta,t)\left(Z_i-\bar{Z}(\beta,t)\right)$$
$$\times\left(\frac{R_i\mathrm{d}N_i(\beta,t)}{\tilde{\pi}(\delta_i,X_i,\gamma)}-\frac{R_i-\tilde{\pi}(\delta_i,X_i,\gamma)}{\tilde{\pi}(\delta_i,X_i,\gamma)}\rho(X_i,\eta)\mathrm{d}\tilde{N}_i(\beta,t)\right).$$

Using arguments similar to those in the proof of Theorem 1 in Lin et al. (1998), we can show that

$$\frac{1}{\sqrt{n}}\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\gamma^*,\eta^*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\overline{\phi}(t)\,(Z_i-\overline{z}(t))\,\mathrm{d}\tilde{M}_i(t,\beta_0,\gamma^*,\eta^*,\Lambda_0)+o_p(1). \tag{13}$$

Taking partial derivative of $\tilde{S}_{\phi}^{\mathrm{DR}}(\beta,\gamma,\eta)$ with respective to γ and η, we obtain that

$$\frac{\partial\tilde{S}_{\phi}^{\mathrm{DR}}(\beta,\gamma,\eta)}{\partial\gamma}=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\phi(\beta,t)\left(Z_i-\overline{Z}(\beta,t)\right)$$
$$\times\frac{R_i\dot{\pi}_{\gamma}(X_i,\gamma)^{\top}}{\pi(X_i,\gamma)^2}\left(\mathrm{d}N_i(\beta,t)-\rho(X_i,\eta)\mathrm{d}\tilde{N}_i(\beta,t)\right),$$

and

$$\frac{\partial\tilde{S}_{\phi}^{\mathrm{DR}}(\beta,\gamma,\eta)}{\partial\eta}=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\phi(\beta,t)\left(Z_i-\overline{Z}(\beta,t)\right)\frac{R_i-\pi(X_i,\gamma)}{\pi(X_i,\gamma)}\dot{\rho}_{\eta}(X_i,\eta)^{\top}\mathrm{d}\tilde{N}_i(\beta,t).$$

By the Law of Large Number, it is easy to see that $n^{-1}\partial\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\gamma^*,\eta^*)/\partial\gamma$ and $n^{-1}\partial\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\gamma^*,\eta^*)/\partial\eta$ converge in probability to B_γ and B_η as $n\to\infty$, respectively. By Taylor expansion of $\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\hat{\gamma},\hat{\eta})$ around γ^* and η^*, and under the condition *C5*, we have that

$$\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\hat{\gamma},\hat{\eta})=\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\gamma^*,\eta^*)-\sum_{i=1}^{n}B_\gamma I_\gamma^{-1}S_{\gamma,i}-\sum_{i=1}^{n}B_\eta I_\eta^{-1}S_{\eta,i}+o_p(\sqrt{n}),$$

where $S_{\gamma,i}$ and $S_{\eta,i}$ are the ith score vector for $\hat{\gamma}$ and $\hat{\eta}$, respectively. By the definition of $\tilde{S}_{\phi}^{\mathrm{DR}}(\beta,\gamma,\eta)$ and (13), we can obtain that

$$\frac{1}{\sqrt{n}}S_{\phi}^{\mathrm{DR}}(\beta_0)=\frac{1}{\sqrt{n}}\left(\tilde{S}_{\phi}^{\mathrm{DR}}(\beta_0,\gamma^*,\eta^*)-\sum_{i=1}^{n}B_\gamma I_\gamma^{-1}S_{\gamma,i}-\sum_{i=1}^{n}B_\eta I_\eta^{-1}S_{\eta,i}\right)+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int_{-\infty}^{\infty}\overline{\phi}(t)\,(Z_i-\overline{z}(t))\,\mathrm{d}\tilde{M}_i(t,\beta_0,\gamma^*,\eta^*,\Lambda_0)\right.$$
$$\left.-B_\gamma I_\gamma^{-1}S_{\gamma,i}-B_\eta I_\eta^{-1}S_{\eta,i}\right)+o_p(1). \tag{14}$$

Note that when $\pi(X_i,\gamma)$ is correctly specified, $B_\eta = 0$ and $\mathsf{E}(S_{\gamma,i}) = 0$, and when $\rho(X_i,\eta)$ is correctly specified, $B_\gamma = 0$ and $\mathsf{E}(S_{\eta,i}) = 0$. Thus, if either $\pi(X_i,\gamma)$ or $\rho(X_i,\eta)$ is correctly specified, $\int_{-\infty}^{\infty}\overline{\phi}(t)(Z_i - \overline{z}(t))\mathrm{d}\tilde{M}_i(t,\beta_0,\gamma^*,\eta^*,\Lambda_0) - B_\gamma I_\gamma^{-1}S_{\gamma,i} - B_\eta I_\eta^{-1}S_{\eta,i}$ has mean zero and variance–covariance matrix

$$\mathsf{E}\left[\left(\int_{-\infty}^{\infty}\overline{\phi}(t)\,(Z-\bar{z}(t))\,\mathrm{d}\tilde{M}(t,\beta_0,\gamma^*,\eta^*,\Lambda_0) - B_\gamma I_\gamma^{-1}S_\gamma - B_\eta I_\eta^{-1}S_\eta\right)^{\otimes 2}\right]. \tag{15}$$

Thus, by the multivariate central limit theorem, we can show that $n^{-1/2}S_\phi^{\mathrm{DR}}(\beta_0)$ is asymptotically normally distributed with mean zero and variance–covariance matrix (15).

Next, for the asymptotic linearity (12), we can use arguments similar to those in the proof of Theorem 1 in Ying (1993) and Theorem 2 in Lin et al. (1998). Specifically, write $S_\phi^{\mathrm{DR}}(\beta) - S_\phi^{\mathrm{DR}}(\beta_0)$ as $U(\beta) - U(\beta_0) + D(\beta,\beta_0,\hat{\gamma},\hat{\eta})$, where $U(\beta) = \sum_{i=1}^n \int_{-\infty}^{\infty}\phi(\beta,t)(Z_i - \bar{Z}(\beta,t))\mathrm{d}N_i(\beta,t)$ and

$$\begin{aligned}D(\beta,\beta_0,\hat{\gamma},\hat{\eta}) = \sum_{i=1}^n &\left[\left(\frac{R_i - \pi(X_i,\hat{\gamma})}{\pi(X_i,\hat{\gamma})}\right)\left(I\{\delta_i = 1\} - \rho(X_i,\hat{\eta})\right)\right.\\ &\times\left(\int_{-\infty}^{\infty}\phi(\beta,t)\left(Z_i - \bar{Z}(\beta,t)\right)\mathrm{d}\tilde{N}_i(\beta,t)\right.\\ &\left.\left.-\int_{-\infty}^{\infty}\phi(\beta_0,t)\left(Z_i - \bar{Z}(\beta_0,t)\right)\mathrm{d}\tilde{N}_i(\beta_0,t)\right)\right].\end{aligned}$$

Note that $U(\beta)$ here has the same form as the $U(\beta)$ defined in Lin et al. (1998), and their results show that $\sup_{\|\beta-\beta_0\|<\mathrm{d}_n}\{\|U(\beta) - U(\beta_0) - A_\phi(\beta-\beta_0)\|/(\sqrt{n}+n\|\beta-\beta_0\|)\} = o_p(1)$. Meanwhile, under the condition *C5*, it is easy to see that

$$\left\|\sum_{i=1}^n\left(\frac{R_i - \pi(X_i,\hat{\gamma})}{\pi(X_i,\hat{\gamma})}\right)\left(I\{\delta_i = 1\} - \rho(X_i,\hat{\eta})\right)\right\| = O_p(\sqrt{n}).$$

Thus, using arguments similar to those in the proof of Theorem 1 in Ying (1993), one can show that $\sup_{\|\beta-\beta_0\|<\mathrm{d}_n}\{\|D(\beta,\beta_0,\hat{\gamma},\hat{\eta})\|\}/(\sqrt{n}+n\|\beta-\beta_0\|)\} = o_p(1)$. The asymptotic linearity (12) then follows by applying the triangle inequality.

Based on (12) and the consistency of $\hat{\beta}_\phi^{\mathrm{DR}}$, we can derive that

$$\sqrt{n}\left(\hat{\beta}_\phi^{\mathrm{DR}} - \beta_0\right) = -A_\phi^{-1}\frac{1}{\sqrt{n}}S_\phi^{\mathrm{DR}}(\beta_0) + o_p(1).$$

Then combining (14) and the Slutsky theorem, the desired conclusion follows immediately. □

Proof of Proposition 3 Using almost the same arguments as those in the proof of Proposition 2, we can derive that

$$-A_\phi \sqrt{n}\left(\hat{\beta}_\phi^{\mathrm{IPW}} - \beta_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\int_{-\infty}^{\infty} \overline{\phi}(t)\,(Z_i - \bar{z}(t))\,\mathrm{d}M_i(t) - C_\gamma I_\gamma^{-1} S_{\gamma,i} \right) + o_p(1),$$

where

$$M_i(t) = \frac{R_i N_i(\beta_0, t)}{\tilde{\pi}(\delta_i, X_i, \gamma^*)} - \int_{-\infty}^{t} Y_i(\beta_0, u)\lambda_0(u)\mathrm{d}u.$$

Thus, the conclusion follows immediately. □

Acknowledgments The authors thank the Associate Editor and two referees whose insightful comments and suggestions have resulted in a great deal of improvement. The research of Ming Zheng was supported by the National Natural Science Foundation of China (11271081). The research of Wen Yu was supported by the National Natural Science Foundation of China (11101091).

References

Cox, D. R., Oakes, D. (1984). *Analysis of survival data*. New York: Chapman and Hall.

Gao, G., Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*, *92*, 875–891.

Gijbels, I., Lin, D. Y., Ying, Z. (2007). Non- and semi-parametric analysis of failure time data with missing failure indicators. *IMS Lecture Notes—Monograph Series, Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, *54*, 203–223.

Goetghebeur, E., Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika*, *82*, 821–834.

Horvitz, D. G., Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*, 663–685.

Hyun, S., Lee, J., Sun, Y. (2012). Proportional hazards model for competing risks data with missing cause of failure. *Journal of Statistical Planning and Inference*, *142*, 1767–1779.

Jin, Z., Lin, D. Y., Wei, L. J., Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, *90*, 341–353.

Kalbfleisch, J. D., Prentice, R. D. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Lee, S., Lewbel, A. (2009). Nonparametric identification of accelerated failure time competing risks models. Working paper, University College London and Boston College.

Lin, D. Y., Wei, L. J., Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, *85*, 605–618.

Liu, C., Wang, Q. (2010). Semiparametric estimation for regression coefficients in the Cox model with failure indicators missing at random. *Statistica Sinica*, *20*, 1125–1142.

Lu, K., Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*, *57*, 1191–1197.

Lu, W., Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica*, *18*, 219–234.

McKeague, I. W., Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information. *Scandinavian Journal of Statistics*, *25*, 589–601.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, *65*, 167–179.

Prentice, R. L., Kalbfleisch, J. D. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, *34*, 541–554.

Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*, 846–866.
Sierra, J., Perez, W., Rozman, W., Carreras, C., Klein, J., Rizzo, J., et al. (2002). Bone marrow transplantation from HLA-identical siblings as treatment for myelodysplasia. *Blood*, *100*, 1997–2004.
Song, X., Sun, L., Mu, X., Dinse, G. E. (2010). Additive hazards regression with censoring indicators missing at random. *The Canadian Journal of Statistics*, *38*, 333–351.
Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, *18*, 354–372.
Wang, Q., Dinse, G. E. (2011). Linear regression analysis of survival data with missing censoring indicators. *Lifetime Data Analysis*, *17*, 256–279.
Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, *21*, 76–99.
Zhou, X., Sun, L. (2003). Additive hazards regression with missing censoring information. *Statistica Sinica*, *13*, 1237–1257.