

Published in final edited form as:

J Stat Plan Inference. 2012 July ; 142(7): 1767–1779. doi:10.1016/j.jspi.2012.02.037.

Proportional hazards model for competing risks data with missing cause of failure

Seunggeun Hyun^a, Jimin Lee^b, and Yanqing Sun^c

Seunggeun Hyun: shyun@uscupstate.edu; Jimin Lee: jlee@unca.edu; Yanqing Sun: yasun@uncc.edu

^aDivision of Mathematics and Computer Science, University of South Carolina Upstate, Spartanburg, SC 29303, USA

^bDepartment of Mathematics, University of North Carolina Asheville, Asheville, NC 28804, USA

^cDepartment of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

Abstract

We consider the semiparametric proportional hazards model for the cause-specific hazard function in analysis of competing risks data with missing cause of failure. The inverse probability weighted equation and augmented inverse probability weighted equation are proposed for estimating the regression parameters in the model, and their theoretical properties are established for inference. Simulation studies demonstrate that the augmented inverse probability weighted estimator is doubly robust and the proposed method is appropriate for practical use. The simulations also compare the proposed estimators with the multiple imputation estimator of Lu and Tsiatis (2001). The application of the proposed method is illustrated using data from a bone marrow transplant study.

Keywords

Asymptotic property; Augmented inverse probability weighted estimator; Cause-specific hazard function; Double robust property; Inverse probability weighted estimator; Missing cause of failure; Multiple imputation

1. Introduction

Competing risks data are commonly encountered in medical studies. Typically the responses to a treatment can be classified in terms of failure from disease of interest or from non-disease-related causes. Hence, in the competing risks framework, each individual is exposed to K distinct types of risks and the eventual failure can be attributed to precisely one of the risks. Let T^* denote the time to failure, Δ^* the cause of failure, and Z a p -dimensional vector of possibly time-dependent covariates. Then a principal estimable quantity in competing risks data is the cause-specific hazard function of cause k , defined, in the absence of censoring, by

© 2012 Elsevier B.V. All rights reserved.

Correspondence to: Jimin Lee, jlee@unca.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

$$\lambda_k^*(t|Z) = \lim_{h \rightarrow 0} h^{-1} P(t \leq T^* < t+h, \Delta^* = k | T^* \geq t, Z),$$

which is the instantaneous rate of experiencing the event of type k at time t , having not experienced any of the K competing events until time t . Without loss of generality in our study, we consider only two causes of failure, the cause of interest as cause 1 and the other as cause 2 (i.e. $\Delta^* = 1$ or 2). In many applications involving follow-up studies, however, individuals may be subject to censoring. Let C be a censoring time and $T^* = \min(T_1^*, T_2^*)$, where T_1^* and T_2^* denote the latent failure times from causes 1 and 2, respectively. Then the observed data consist of observations of (T, Δ, Z) , where $T = \min(T^*, C)$ and $\Delta = \Delta^* I(T^* \leq C)$. If the failure time T^* is observed, Δ is the cause of failure and $\Delta = 0$ otherwise. The observable cause-specific hazard function of cause k in the presence of censoring is given by

$$\lambda_k(t|Z) = \lim_{h \rightarrow 0} h^{-1} P(t \leq T < t+h, \Delta = k | T \geq t, Z), \quad k=1, 2.$$

Throughout the paper, we assume that Z is an external covariate process (Kalbfleisch and Prentice, 2002) and the censoring time C is conditionally independent of (T^*, Δ^*) given Z (Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Lu and Liang, 2008). Under this assumption, it can be shown that $\lambda_k^*(t|Z) = \lambda_k(t|Z)$ if the distribution of C is continuous at t . A number of statistical models for the relationship between the cause-specific hazard function of interest and regression covariates have been studied, among others, by Benichou and Gail (1990); Prentice et al. (1978); Cheng et al. (1998); Shen and Cheng (1999); Scheike and Zhang (2003). In this article we study the proportional hazards model for describing the relationship,

$$\lambda_1^*(t|Z) = \lambda_0(t) e^{\beta_0^T Z(t)}, \quad (1)$$

where $\lambda_0(\cdot)$ is a nonnegative, but otherwise unspecified baseline hazard function and β_0 is a p -dimensional vector of regression parameters. The parameter β_0 can be consistently estimated by treating all the failure times with $\Delta \neq 1$ as censored observations and using the partial likelihood score equation proposed by Cox (1972, 1975). The estimator will be called the full-case estimator, denoted by $\hat{\beta}_F$ in the paper.

In practice, however, the information needed for the cause of failure may be lost, or it may be difficult to determine the cause of disease or death for some individuals (Andersen et al., 1996). When we have missing causes in data, a naive method for estimating the regression parameter β_0 is to simply ignore the missing data and use the partial likelihood score equation to the complete data only. The so-called complete-case estimator, denoted by $\hat{\beta}_C$, is clearly inefficient and can lead to serious bias. Thus, analysis of competing risks data with missing cause of failure has received considerable attention and a number of models have been proposed. Dinse (1982, 1986) considered nonparametric estimation for incomplete cause of death data with no covariate. Goetghebeur and Ryan (1990) proposed a modified log-rank test to compare survival in two groups, Dewanji (1992) suggested a modification of that approach, and Goetghebeur and Ryan (1995) extended the results of Goetghebeur and Ryan (1990) to proportional hazards regression model. More recently, Lu and Tsiatis (2001) proposed a parametric model to model the probability that the missing cause is the cause of interest while allowing the inclusion of additional auxiliary covariates and then estimated the regression parameters by using a multiple imputation method (Rubin, 1987, 1996). Gao

and Tsiatis (2005) considered linear transformation models and Lu and Liang (2008) considered the additive hazards model for analysis of competing risks data with missing cause of failure.

For right-censored survival data in which the censoring indicator is missing, Lo (1991) considered the problem of nonparametric maximum likelihood estimation of a survival function in the absence of covariates, McKeague and Subramanian (1998) developed a survival function estimator assuming that the censoring indicators are missing completely at random. Subramanian (2000) considered further development of efficient estimation of the regression parameters under proportionality assumptions of the conditional hazards, and Gijbels et al. (2007) proposed a class of estimating functions for the regression parameters of the Cox proportional hazard model, among others.

In this study of analysis of competing risks data with missing cause of failure, we derive two different estimators for the regression parameters in model (1), namely the inverse probability weighted estimator and augmented inverse probability weighted estimator, and establish their theoretical properties. The first approach, following the idea of Horvitz and Thompson (1952), uses the inverse probability weighted complete-case technique to estimate the regression parameter. This approach uses only the complete cases and relies on correct modeling for the probability of missing causes. It has been shown that the inverse probability weighted estimator is inconsistent when the respective parametric model is misspecified, and is inefficient (Gao and Tsiatis, 2005; Lu and Liang, 2008; Scharfstein et al., 1999). It would be desirable, therefore, to obtain improved efficiency over the inverse probability weighted estimator. The second approach, adapting the idea of Robins et al. (1994), augments the inverse probability weighted complete-case estimating equation with a consistent estimator of the conditional distribution of the cause of interest that incorporates information available for individuals whose cause of failure is missing. See Subramanian and Bandyopadhyay (2010) for homogeneous right censored data with missing censoring indicators.

The paper is structured as follows. In Section 2, the inverse probability weighted estimator and augmented inverse probability weighted estimator are developed. The asymptotic properties of the corresponding estimators are established in Section 3. In Section 4, we investigate the finite sample properties of the proposed estimators through simulations, including comparisons with the multiple imputations estimator proposed by Lu and Tsiatis (2001). A bone marrow transplant data set is analyzed in Section 4. Some conclusions and discussions are given in Section 5. Technical derivations are detailed in Appendix.

2. Estimating equations

Since the cause of failure may not be observed for some individuals, we define the missingness indicator R as follows. If an individual's death is observed, then $R = 1$ when the cause of failure information Δ^* is observed and $R = 0$ otherwise. If an individual is censored, we always define $R = 1$. We also introduce auxiliary covariates A which are not of interest for modelling the cause-specific hazard function but may be used to describe the missingness mechanism. The utilization of auxiliary information has been considered by Lu and Tsiatis (2001), Gao and Tsiatis (2005), Lu and Liang (2008), Gilbert, McKeague, and Sun (2008), among others. Then the observed data will consist of

$$O_i = \{T_i, Z_i, A_i, R_i, I(\Delta_i=0), R_i I(\Delta_i=1), R_i I(\Delta_i=2)\}$$

for $i = 1, \dots, n$. We assume that $\{O_i, i = 1, \dots, n\}$ are independent identically distributed. The possible choices are $\{T_i, Z_i, A_i, 1, 0, 1, 0\}$ for the individual who died from the cause 1, $\{T_i, Z_i, A_i, 1, 0, 0, 1\}$ for the individual who died from the cause 2, $\{T_i, Z_i, A_i, 0, 0, 0, 0\}$ for the individual who died with missing cause, and $\{T_i, Z_i, A_i, 1, 1, 0, 0\}$ for the censored individual.

We also assume that the cause of failure is missing at random (MAR) (Rubin, 1976); that is, the probability that the cause of failure is missing given $\Delta > 0$ and $W = (T, Z, A)$ depends only on the observed W , but not on the unobserved Δ ,

$$P(R=1|\Delta, \Delta>0, W)=P(R=1|\Delta>0, W). \quad (2)$$

The assumption implies that

$$\begin{aligned} P(R=1|\Delta>0, W) &= P(R=1, \Delta=1|\Delta>0, W) + P(R=1, \Delta=2|\Delta>0, W) \\ &= P(R=1|\Delta=1, \Delta>0, W)P(\Delta=1|\Delta>0, W) + P(R=1|\Delta=2, \Delta>0, W)P(\Delta=2|\Delta>0, W) \\ &= P(R=1|\Delta=1, \Delta>0, W)[P(\Delta=1|\Delta>0, W) + P(\Delta=2|\Delta>0, W)] \\ &= P(R=1|\Delta=1, \Delta>0, W), \end{aligned}$$

and likewise for the case involving $\Delta = 2$. See also Lu and Tsiatis (2001), Gao and Tsiatis (2005), and Lu and Liang (2008).

2.1. Inverse probability weighted estimator

Following the inverse selection probability idea of Horvitz and Thompson (1952), the method of inversely weighting the probability of complete-case has been commonly used in missing data problems. To do that, we need to estimate the probability of a complete case, $\pi(Q) \equiv P(R = 1|Q)$, where $Q = (W, \Delta)$. By the MAR assumption and $R = 1$ when $\Delta = 0$, we have

$$\begin{aligned} \pi(Q) &= P(R=1|W, \Delta) \\ &= P(R=1|W, \Delta, \Delta>0) \cdot I(\Delta>0) + P(R=1|W, \Delta, \Delta=0) \cdot I(\Delta=0) \\ &= P(R=1|W, \Delta, \Delta>0) \cdot I(\Delta>0) + 1 \cdot I(\Delta=0) \\ &= P(R=1|W, \Delta>0) \cdot I(\Delta>0) + I(\Delta=0) \\ &= r(W) \cdot I(\Delta>0) + I(\Delta=0), \end{aligned} \quad (3)$$

where $r(W) = P(R = 1|W, \Delta > 0)$. We consider that the probability of complete-case $r(W_i)$ may be specified as a parametric model $r(W_i, \psi_0)$, in terms of a few unknown parameters ψ_0 . Accordingly, let $\pi(Q_i, \psi_0) = r(W_i, \psi_0)I(\Delta > 0) + I(\Delta = 0)$. Since R is binary, one can posit the logistic model $\text{logit}\{r(W_i, \psi)\} = W_i^\top \psi$, though other parametric models can also be used. By (2) and (3), the likelihood L regarding to $\pi(Q, \psi_0)$ is

$$\begin{aligned} L(\pi) &= \prod \pi(Q_i)^{I(R_i=1)} (1 - \pi(Q_i))^{I(R_i=0)} \\ &= \prod \pi(Q_i)^{I(R_i=1)I(\Delta_i>0)} \cdot \pi(Q_i)^{I(R_i=1)I(\Delta_i=0)} \cdot (1 - \pi(Q_i))^{I(R_i=0)I(\Delta_i>0)} \\ &= \prod r(W_i)^{I(R_i=1)I(\Delta_i>0)} \cdot 1 \cdot (1 - r(W_i))^{I(R_i=0)I(\Delta_i>0)}, \\ &\text{because } \pi(Q_i) = r(W_i) \text{ when } \Delta_i > 0, \text{ and } \pi(Q_i) = 1 \text{ when } \Delta_i = 0 \\ &= \prod r(W_i)^{I(R_i=1)I(\Delta_i>0)} \cdot (1 - r(W_i))^{(1-R_i)I(\Delta_i>0)}. \end{aligned}$$

This implies that the maximum likelihood estimator $\hat{\psi}$ of ψ can be estimated by maximizing the likelihood based on uncensored data

$$\prod_{i=1}^n \{r(W_i, \psi)\}^{R_i I(\Delta_i > 0)} \{1 - r(W_i, \psi)\}^{(1-R_i) I(\Delta_i > 0)}.$$

It is known that for a correctly specified model $r(W_i, \psi)$, $\hat{\psi}$ consistently estimates ψ_0 , the true value of the parametric component of $r(W_i, \psi)$ (Haberman, 1974, 1977; Gourieroux and Monfort, 1981).

We define the counting process $N_i(t) = I(\Delta_i = 1)I(T_i \leq t)$ and at-risk process $Y_i(t) = I(T_i \geq t)$. Let $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$ for a vector a . Let

$$\begin{aligned} \tilde{S}^{(m)}(t, \beta, \psi) &= \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi(Q_i, \psi)} Y_i(t) e^{\beta^T Z_i(t)} Z_i(t)^{\otimes m}, \\ \tilde{Z}(t, \beta, \psi) &= \tilde{S}^{(1)}(t, \beta, \psi) / \tilde{S}^{(0)}(t, \beta, \psi), \\ \tilde{V}(t, \beta, \psi) &= \frac{\tilde{S}^{(2)}(t, \beta, \psi)}{\tilde{S}^{(0)}(t, \beta, \psi)} - \tilde{Z}(t, \beta, \psi)^{\otimes 2} \end{aligned}$$

for $m = 0, 1, 2$. Then we consider the following inverse probability weighted estimating equation for β_0 :

$$U_i(\beta, \hat{\psi}) = \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \hat{\psi})} (Z_i(t) - \tilde{Z}(t, \beta, \hat{\psi})) dN_i(t), \quad (4)$$

where $\tau > 0$ is the end of follow-up time. The inverse probability weighted estimator (IPW) of β solves the above equation and is denoted by $\hat{\beta}_I$. When there is no missing cause, the equation (4) consequently becomes the partial likelihood score equation proposed by Cox (1972, 1975). The cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ can be estimated by

$$\hat{\Lambda}_{I0}(t) = \sum_{i=1}^n \int_0^t \frac{R_i}{\pi(Q_i, \hat{\psi})} \frac{dN_i(u)}{n \tilde{S}^{(0)}(u, \hat{\beta}_I, \hat{\psi})}.$$

2.2. Augmented inverse probability weighted estimator

The inverse probability weighted estimator $\hat{\beta}_I$ uses only complete cases. Thus, it is inefficient. In addition, its consistency relies on correct modelling of the probability $r(W_i, \psi_0)$. To improve the robustness and efficiency over $\hat{\beta}_I$, we adapt the idea of Robins et al. (1994) and propose to augment the inverse probability weighted estimating equation with a consistent estimator of the conditional distribution of the cause of interest that utilizes available information for individuals with missing cause of failure.

Consequently, we estimate the probability that the cause of failure is the cause of interest $\rho(W) \equiv P(\Delta = 1 | \Delta > 0, W)$. Let $f(t, \Delta = 1 | \Delta > 0, z, a)$ be the conditional joint density of T and $\Delta = 1$ given $(\Delta > 0, Z = z, A = a)$. Then for $w = (t, z, a)$,

$$\begin{aligned}\rho(w) &= \frac{f(t, \Delta=1|\Delta>0, z, a)}{f(t|\Delta>0, z, a)} \\ &= \frac{f(t, \Delta=1|z, a)}{f(t, \Delta=1|z, a) + f(t, \Delta=2|z, a)} \\ &= \frac{\lambda_1(t|z, a)}{\lambda_1(t|z, a) + \lambda_2(t|z, a)},\end{aligned}$$

where $\tilde{\lambda}_k(t|z, a)$ is the conditional cause-specific hazard function of T at t due to cause $\Delta = k$ given $(Z, A) = (z, a)$ for $k = 1, 2$. Here, instead of directly estimating $\rho(W_i)$ which requires the estimation of two unknown cause-specific hazard functions, we posit a parametric model $\rho(W_i, \gamma_0)$ for $\rho(W_i)$ in terms of a few unknown parameters γ_0 . It is natural to use a logistic regression model $\text{logit}\{\rho(W_i, \gamma)\} = W_i^\top \gamma$, but other parametric models can also be accommodated.

There is, however, an issue with obtaining estimates for γ in the presence of missingness. The MAR assumption implies that given $\Delta > 0$ and W , R is independent of Δ ; that is,

$$P(\Delta=1|\Delta>0, W) = P(\Delta=1|R, \Delta>0, W). \quad (5)$$

By (5), $\rho(W_i)$ can be deduced from the complete cases with $R_i = 1$ and $\Delta_i > 0$. This suggests that the maximum likelihood estimator $\hat{\gamma}$ of γ can be obtained by maximizing the likelihood based on complete-case data

$$\prod_{i=1}^n \{\rho(W_i, \gamma)\}^{R_i I(\Delta_i=1)} \{1 - \rho(W_i, \gamma)\}^{R_i I(\Delta_i=2)}.$$

Since $\hat{\gamma}$ is the maximum likelihood estimator, then for a correctly specified model $\rho(W_i, \gamma)$, $\hat{\gamma}$ consistently estimates γ_0 , the true value of the parametric component model $\rho(W_i, \gamma)$ (Haberman, 1974, 1977; Gourieroux and Monfort, 1981).

Now, we define the counting process $N_i^*(t) = I(\Delta_i > 0)I(T_i \leq t)$. Let

$$\begin{aligned}S^{(m)}(t, \beta) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta^\top Z_i(t)} Z_i(t)^{\otimes m}, \\ \bar{Z}(t, \beta) &= S^{(1)}(t, \beta) / S^{(0)}(t, \beta), \\ V(t, \beta) &= \frac{S^{(2)}(t, \beta)}{S^{(0)}(t, \beta)} - \bar{Z}(t, \beta)^{\otimes 2}.\end{aligned}$$

Then we propose the following augmented inverse probability weighted estimating equation

$$U_A(\beta, \hat{\psi}, \hat{\gamma}) = \sum_{i=1}^n \int_0^\tau (Z_i(t) - \bar{Z}(t, \beta)) \left[\frac{R_i}{r(W_i, \hat{\psi})} dN_i(t) - \frac{R_i - r(W_i, \hat{\psi})}{r(W_i, \hat{\psi})} \rho(W_i, \hat{\gamma}) dN_i^*(t) \right], \quad (6)$$

where $\hat{\psi}$ and $\hat{\gamma}$ are the maximum likelihood estimators defined earlier. The augmented inverse probability weighted estimator (AIPW) of β solves the above equation (6) and is denoted by $\hat{\beta}_A$. The cumulative baseline hazard function can be estimated by

$$\widehat{\Lambda}_{A_0}(t) = \sum_{i=1}^n \int_0^t \frac{1}{nS^{(0)}(u, \widehat{\beta}_A)} \left[\frac{R_i}{r(W_i, \widehat{\psi})} dN_i(u) - \frac{R_i - r(W_i, \widehat{\psi})}{r(W_i, \widehat{\psi})} \rho(W_i, \widehat{\gamma}) dN_i^*(u) \right].$$

3. Asymptotic results

When the model for $r(W_i)$ is correctly specified, we let ψ_0 be the true value of ψ such that $r(W_i) = r(W_i, \psi_0)$. Under Condition (A.4) stated in the Appendix, $\widehat{\psi} \xrightarrow{P} \psi_0$. When the model for $\rho(W_i)$ is correctly specified, we let γ_0 be the true value of γ such that $\rho(W_i) = \rho(W_i, \gamma_0)$. In this case, $\widehat{\gamma} \xrightarrow{P} \gamma_0$. In general, under Condition (A.4), there exist ψ^* and γ^* such that $\widehat{\psi} \xrightarrow{P} \psi^*$ and $\widehat{\gamma} \xrightarrow{P} \gamma^*$ (White, 1982). We have $\psi^* = \psi_0$ if $r(W_i)$ is correctly specified, and $\gamma^* = \gamma_0$ if $\rho(W_i)$ is correctly specified.

Let $s^{(m)}(t, \beta) = E[Y_1(t)e^{\beta^\top Z_1(t)} Z_1(t)^{\otimes m}]$, $\bar{z}(t, \beta) = s^{(1)}(t, \beta)/s^{(0)}(t, \beta)$, and $v(t, \beta) = s^{(2)}(t, \beta)/s^{(0)}(t, \beta) - \bar{z}(t, \beta)^{\otimes 2}$.

Theorem 1

Assume Condition A given in the Appendix. If $r(W_i, \psi_0)$ is correctly specified for $r(W_i)$, then $\widehat{\beta}_t \xrightarrow{P} \beta_0$ and $\sqrt{n}(\widehat{\beta}_t - \beta_0)$ converges in distribution to a zero-mean Gaussian random vector with covariance matrix $\Sigma^{-1} E[\omega_i \omega_i^\top] \Sigma^{-1}$, where

$$\Sigma = \int_0^\tau v(t, \beta_0) \lambda_0(t) s^{(0)}(t, \beta_0) dt, \\ \omega_i = \int_0^\tau \{Z_i(t) - \bar{z}(t, \beta_0)\} \frac{R_i}{\pi(Q_i, \psi_0)} dM_i(t) - V_\psi I_\psi^{-1} S_{\psi i},$$

$M_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\beta_0^\top Z_i(u)} d\Lambda_0(u)$, V_ψ is given in (8), I_ψ and $S_{\psi i}$ are given in (11) in the Appendix.

The asymptotic covariance matrix $\Sigma^{-1} E(\omega_i \omega_i^\top) \Sigma^{-1}$ can be consistently estimated by

$$\widehat{\Sigma}_t^{-1} \left(n^{-1} \sum_{i=1}^n \widehat{\omega}_i \widehat{\omega}_i^\top \right) \widehat{\Sigma}_t^{-1},$$

where

$$\widehat{\Sigma}_t = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \widehat{\psi})} \widetilde{V}(t, \widehat{\beta}_t, \widehat{\psi}) dN_i(t), \\ \widehat{\omega}_i = \int_0^\tau \frac{R_i}{\pi(Q_i, \widehat{\psi})} \{Z_i(t) - \widehat{Z}(t, \widehat{\beta}_t, \widehat{\psi})\} d\widehat{M}_i(t) - \widehat{V}_\psi \widehat{I}_\psi^{-1} \widehat{S}_{\psi i},$$

and $\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\widehat{\beta}_t^\top Z_i(u)} d\widehat{\Lambda}_{r_0}(u)$. Here \widehat{V}_ψ , \widehat{I}_ψ and $\widehat{S}_{\psi i}$ are obtained by replacing with their respective sample estimators and substituting $(\widehat{\beta}_t, \widehat{\psi})$ for (β_0, ψ_0) in V_ψ , I_ψ , and $S_{\psi i}$.

The following establishes the asymptotic properties of $\widehat{\beta}_A$.

Theorem 2

Assume Condition A given in the Appendix. If at least one of $r(W_i, \psi_0)$ and $\rho(W_i, \gamma_0)$ is correctly specified for $r(W_i)$ and $\rho(W_i)$, then $\widehat{\beta}_A \xrightarrow{P} \beta_0$ and $\sqrt{n}(\widehat{\beta}_A - \beta_0)$ converges in distribution to a zero-mean Gaussian random vector with covariance matrix

$$\sum^{-1} E[\varphi_i \varphi_i^T] \sum^{-1}, \text{ where}$$

$$\sum = \int_0^\tau v(t, \beta_0) \lambda_0(t) s^{(0)}(t, \beta_0) dt,$$

and

$$\varphi_i = \int_0^\tau \{Z_i(t) - \bar{z}(t, \beta_0)\} dM_i^*(t) - P_\psi (I_\psi^*)^{-1} S_{\psi i}^* - P_\gamma (I_\gamma^*)^{-1} S_{\gamma i}^*.$$

Here $M^*(t)$ is defined in (17), and $P_\psi, I_\psi^*, S_{\psi i}^*, P_\gamma, I_\gamma^*$ and $S_{\gamma i}^*$ are given in (15) and (19) in the Appendix.

It is interesting to notice that if $r(W_i, \psi_0) = r(W_i)$, then $P_\gamma = 0$, and if $\rho(W_i, \gamma_0) = \rho(W_i)$, then $P_\psi = 0$. When both the models for $r(W_i)$ and $\rho(W_i)$ are correctly specified, we have $P_\psi = 0$ and $P_\gamma = 0$ and hence, $\varphi_i = \int_0^\tau \{Z_i(t) - \bar{z}(t, \beta_0)\} dM_i^*(t)$.

The asymptotic covariance matrix $\sum^{-1} E(\varphi_i \varphi_i^T) \sum^{-1}$ can be consistently estimated by

$$\widehat{\sum}_A^{-1} \left(n^{-1} \sum_{i=1}^n \widehat{\varphi}_i \widehat{\varphi}_i^T \right) \widehat{\sum}_A^{-1},$$

where

$$\widehat{\sum}_A = \frac{1}{n} \sum_{i=1}^n \int_0^\tau V(t, \widehat{\beta}_A) \left[\frac{R_i}{r(W_i, \widehat{\psi})} dN_i(t) - \frac{R_i - r(W_i, \widehat{\psi})}{r(W_i, \widehat{\psi})} \rho(W_i, \widehat{\gamma}) dN_i^*(t) \right],$$

$$\widehat{\varphi}_i = \int_0^\tau \{Z_i(t) - \bar{Z}(t, \widehat{\beta}_A)\} d\widehat{M}_i^*(t) - \widehat{P}_\psi (\widehat{I}_\psi^*)^{-1} \widehat{S}_{\psi i}^* - \widehat{P}_\gamma (\widehat{I}_\gamma^*)^{-1} \widehat{S}_{\gamma i}^*,$$

and

$$\widehat{M}_i^*(t) = \frac{R_i}{r(W_i, \widehat{\psi})} N_i(t) - \frac{R_i - r(W_i, \widehat{\psi})}{r(W_i, \widehat{\psi})} \rho(W_i, \widehat{\gamma}) N_i^*(t) - \int_0^t Y_i(u) e^{\widehat{\beta}_A^T Z_i(u)} d\widehat{\Lambda}_{A0}(u).$$

Here \hat{P}_ψ^* , \hat{I}_ψ^* , $\hat{S}_{\psi i}^*$, \hat{P}_γ^* , \hat{I}_γ^* and $\hat{S}_{\gamma i}^*$ are the empirical counterparts of P_ψ , I_ψ^* , $S_{\psi i}^*$, P_γ , I_γ^* and $S_{\gamma i}^*$ given in (15) and (19) in the Appendix, obtained by replacing with their respective sample estimators and substituting $(\hat{\beta}_A, \hat{\rho}, \hat{\gamma})$ for $(\beta_0, \psi^*, \gamma^*)$.

4. Numerical results

4.1. Simulation studies

We present simulation studies conducted to evaluate the performance of our proposed methods. We set $\tau = 2.0$ and consider a univariate covariate Z , where Z follows a uniform distribution on $[0, 1]$. Given Z , the latent failure time T_1^* of interest is generated from the proportional hazards model $\lambda_1^*(t|Z) = \lambda e^{\beta Z}$, where $\lambda = 1$ and $\beta = -0.5$. The other latent failure time T_2^* is generated from a Gompertz distribution with a hazard function $\lambda_2^*(t|Z) = e^{\theta + \nu t}$, where $\theta = -0.5$ and $\nu = 0.2$. The censoring time C is generated from an exponential distribution which yields about 20% censoring level. We consider a single auxiliary covariate A which follows a Bernoulli distribution with success probability of 0.5. We also consider a logistic regression model $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2 T + \psi_3 Z + \psi_4 A$ for missing cause of failure. We have about 20% missingness with $\psi = (0.7, 1, -1, 1)$ and about 45% missingness with $\psi = (-0.8, 1, -1, 1)$. In the settings we consider here, the true model $\rho(W)$ is given by a logistic regression model $\text{logit}\{\rho(W)\} = -\theta - \nu T + \beta Z$. To study the performance of the estimators when $r(W)$ is misspecified, we posit two different parametric models of $r(W, \psi)$, where one is a correctly specified logistic model and the other is a misspecified constant model $r_0 \in (0, 1)$ independent of W . To study the behavior when $\rho(W)$ is misspecified, we consider various model specifications. We posit a correctly specified logistic model $\text{logit}\{\rho(W, \gamma)\} = \gamma_1 + \gamma_2 T + \gamma_3 Z$ (Model 1), a misspecified logistic model $\text{logit}\{\rho(W, \gamma)\} = \gamma_1 + \gamma_2 Z$ (Model 2), a misspecified logistic model $\text{logit}\{\rho(W, \gamma)\} = \gamma_1 + \gamma_2 T$ (Model 3) and a misspecified constant model $\rho_0 \in (0, 1)$ independent of W (Model 4). The simulation studies consist of 1000 runs with the sample size $n = 200$ and $n = 400$. We also conduct comparison with the multiple imputation estimators, studied by Lu and Tsiatis (2001), with the number of imputation $m = 1$ and $m = 5$.

The results from Table 1 and Table 2 show that the complete-case estimator $\hat{\beta}_C$ shows large biases in all the settings. When the parametric model for $r(W)$ is correctly specified, both the IPW estimator $\hat{\beta}_I$ and AIPW estimator $\hat{\beta}_A$ show small biases, but the AIPW estimator has smaller standard errors than the corresponding IPW estimator. When the parametric model for $\rho(W)$ is correctly specified, the multiple imputation estimator has small biases, but the multiple imputation estimator tends to have larger biases when $\rho(W)$ is misspecified. As expected, the AIPW estimator is clearly not sensitive to the misspecification if one of the parametric models for $r(W)$ and $\rho(W)$ is misspecified. In fact, the AIPW estimator performs quite well even when both the parametric models $r(W)$ and $\rho(W)$ are misspecified. The standard errors of the multiple imputation estimator decrease as the number of imputation increases. The standard errors of the multiple imputation estimator and the AIPW estimator are comparable when $\rho(W)$ is correctly specified. However, the multiple imputation estimator seems to have small standard errors than the AIPW estimator under misspecified $\rho(W)$. The estimated standard errors of the AIPW estimator are close to the sample standard errors, and the 95% confidence intervals have reasonable coverage probabilities.

To further study robustness of parameter estimates against misspecification of the parametric models for $r(W)$ and $\rho(W)$, we consider the same model for T_1^* as described above, but here instead we generate T_2^* from a Weibull distribution, log logistic distribution, exponential power distribution and gamma distribution. None of these distributions will induce a simple linear logistic regression model for $\rho(W)$. For example, when T_2^* is generated

from a log logistic distribution with a hazard function $\lambda_2^*(t|Z) = \alpha \lambda t^{\alpha-1} / (1 + \lambda t^\alpha)$, the true logistic model for $\rho(W)$ is $\text{logit}\{\rho(W)\} = -\log(\alpha) - \log(\lambda) + \beta Z + \log((1 + \lambda T^\alpha) / T^{\alpha-1})$. In all cases we misspecify $\rho(W)$ by Model 1 to Model 4 described above. Although not presented here, the findings from these simulations are similar to those from Table 1 and Table 2.

In conclusion, the multiple imputation estimator of Lu and Tsiatis (2001) and the AIPW estimator have similar performance when the parametric models for $r(W)$ and $\rho(W)$ are correctly specified. The AIPW estimator has the advantage of double robustness such that the biases of the AIPW estimator remain small when the parametric model for $r(W)$ or $\rho(W)$ is misspecified. The bias of the multiple imputation estimator can be larger under misspecified $\rho(W)$.

4.2. Bone marrow transplant data

Sierra et al. (2002) described the characteristics and outcomes of 452 patients with primary myelodysplasia (MDS) who received transplants from HLA-identical siblings and were registered with the International Bone Marrow Transplant Registry (IBMTR). The study has two competing risks; treatment related death defined as death in complete remission and relapse defined as recurrence of myelodysplasia. In this example, we consider 408 patients with complete covariate information obtained from the `timereg` package for R. Among these 408 patients, 161 patients died in complete remission, 87 patients relapsed, and 160 patients were censored. The covariates considered in our study are age of patient standardized at mean of 35 years old and platelet before transplantation (1 for more than 100×10^9 per L, or 0 for less). In the data set, the causes of failure are all known. For illustration purposes, we delete some failure causes by the three following missing mechanisms; missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).

For the MCAR, the causes of failure are randomly selected for missing with probability 23%. For the MAR, the logistic model is chosen as $\text{logit}\{r(W)\} = 0.5 + 1.0 * T - 1.0 * \text{age}$ which yields about 23% missing causes, where T is the failure time. For the NMAR, the logistic model is chosen as $\text{logit}\{r(W)\} = 0.5 + 1.0 * T - 1.0 * \text{age} - 0.5 * I(\Delta = 1)$ which yields about 26% missing causes, where $\Delta = 1$ corresponds to the death in complete remission and $\Delta = 2$ does to relapse. We posit the logistic models for both $r(W)$ and $\rho(W)$ with $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2 * T + \psi_3 * \text{age} + \psi_4 * \text{platelet}$ and $\text{logit}\{\rho(W, \gamma)\} = \gamma_1 + \gamma_2 * \log T + \gamma_3 * \text{age} + \gamma_4 * \text{platelet}$.

The results of the estimation of β based on the AIPW estimator, the complete-case estimator and the multiple imputation estimator with the number of imputation $m = 5$ are summarized in Table 3. For comparison, Table 3 also includes the estimation of β based on the original data without artificial missing, namely, the full-case estimator. The results from the AIPW estimator and the multiple imputation estimator are very close under all the missingness mechanisms and they are closer to the full-case estimator than the complete-case estimator. The analyses using the AIPW estimator and the multiple imputation estimator are consistent with the findings from the earlier study; that is, patients with high platelet counts have a lower risk of treatment related mortality than those with low platelet counts, and a higher risk rate is seen among the older patients.

5. Conclusion

We propose the inverse probability weighted estimator and augmented inverse probability weighted estimator for analysis of competing risks data with missing cause of failure, where the Cox proportional hazard model is utilized to examine the covariate effects on the cause-specific hazard function for the failure type of interest. The augmented inverse probability

weighted estimator possesses the double robust property such that it is unbiased as long as one of the parametric models for $r(W)$ and $\rho(W)$ is correctly specified. The inverse probability weighted estimator is unbiased only when the parametric model for $r(W)$ is correctly specified. Under the correctly specified models for $r(W)$ and $\rho(W)$, the augmented inverse probability weighted estimator is more efficient than the inverse probability weighted estimator.

The proposed estimators are compared with the multiple imputation estimator of Lu and Tsiatis (2001) through simulations. The multiple imputation estimator and augmented inverse probability weighted estimator have similar performance when the parametric models for $r(W)$ and $\rho(W)$ are correctly specified. The augmented inverse probability weighted estimator has the advantage of the double robustness over the multiple imputation estimator.

In the competing risks problem, another useful quantity is the cumulative incidence function which is the probability of occurrence by time t for a particular type of failure in the presence of other risks. It is known that the covariate effect on the cause-specific hazard for a particular type of failure can be quite different from its effect on the cumulative incidence function of that type of failure (Gray, 1988; Gaynor et al., 1993). Fine and Gray (1999) developed a direct Cox regression approach for the cumulative incidence curve based on earlier work by Gray (1988) and Pepe (1991). Recently, Andersen et al. (2003) and Klein and Andersen (2005) suggested pseudo-observation approach for direct modeling for cumulative incidence probabilities. It would be interesting to study models that relate the covariates directly to the cumulative incidence function for the failure type of interest in competing risks data with missing cause of failure.

Acknowledgments

The research of Yanqing Sun was partially supported by NSF DMS-0905777 and NIH grant R37 AI054165-09. The authors thank the reviewers for their constructive comments that have improved the paper.

References

- Andersen J, Goetghebuer E, Ryan L. Missing cause of death information in the analysis of survival data. *Statistics in Medicine*. 1996; 15:2191–2201. [PubMed: 8910963]
- Andersen P, Klein J, Rosthøj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*. 2003; 90:15–27.
- Benichou J, Gail M. Estimates of absolute cause-specific risk in cohort studies. *Biometrics*. 1990; 46:813–826. [PubMed: 2242416]
- Cheng S, Fine J, Wei L. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*. 1998; 54:219–228. [PubMed: 9544517]
- Cox D. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B*. 1972; 34:187–220.
- Cox D. Partial likelihood. *Biometrika*. 1975; 62:269–276.
- Dewanji A. A note on a test for competing risks with missing failure type. *Biometrika*. 1992; 79:855–857.
- Dinse G. Nonparametric estimation for partially-complete time and of failure data. *Biometrics*. 1982; 38:417–431. [PubMed: 6896831]
- Dinse G. Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. *Journal of the American Statistical Association*. 1986; 81:328–336.
- Fine J, Gray R. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94:496–509.
- Gao G, Tsiatis A. Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*. 2005; 92:875–891.

- Gaynor J, Feuer E, Tan C, Wu D, Little C, Straus D, Clarkson B, Brennan M. On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. *Journal of the American Statistical Association*. 1993; 88:400–409.
- Gijbels I, Lin D, Ying Z. Non- and semi-parametric analysis of failure time data with missing failure indicators. *IMS Lecture Notes Monograph Series*. 2007; 54:203–223.
- Gilbert P, McKeague I, Sun Y. The two-sample problem for failure rates depending on a continuous mark: An application to vaccine efficacy. *Biostatistics*. 2008; 9:263–276. [PubMed: 17704528]
- Goetghebeur E, Ryan L. A modified logrank test for competing risks with missing failure type. *Biometrika*. 1990; 77:207–211.
- Goetghebeur E, Ryan L. Analysis of competing risks survival data when some failure types are missing. *Biometrika*. 1995; 82:821–834.
- Gourieroux C, Monfort A. Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. *Journal of Econometrics*. 1981; 17:83–97.
- Gray R. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics*. 1988; 16:1040–1154.
- Haberman, S. *The Analysis of Frequency Data*. University of Chicago Press; Chicago: 1974.
- Haberman S. Maximum likelihood estimates in exponential response models. *Annals of Statistics*. 1977; 5:815–841.
- Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; 47:663–685.
- Kalbfleish, J.; Prentice, R. *The Statistical Analysis of Failure Time Data*. 2. Wiley; New York: 2002.
- Klein J, Andersen P. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005; 61:223–229. [PubMed: 15737097]
- Lo SH. Estimating a survival function with incomplete cause-of-death data. *Journal of Multivariate Analysis*. 1991; 39:217–235.
- Lu K, Tsiatis A. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*. 2001; 57:1191–1197. [PubMed: 11764260]
- Lu W, Liang Y. Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica*. 2008; 18:219–234.
- McKeague I, Subramanian S. Product-limit estimators and cox regression with missing censoring information. *Scandinavian Journal of Statistics*. 1998; 25:589–601.
- Pepe M. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*. 1991; 86:770–778.
- Prentice R, Kalbfleisch J, Peterson A, Flournoy N, Farewell V, Breslow N. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978; 34:541–554. [PubMed: 373811]
- Robins J, Rotnitzky A, Zhao L. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
- Rubin D. Inference and missing data. *Biometrika*. 1976; 63:581–592.
- Rubin, D. *Multiple Imputation for Nonresponse in Surveys*. Wiley; New York: 1987.
- Rubin D. Multiple imputation after 18+ years. *Journal of the American Statistical Association*. 1996; 91:473–489.
- Scharfstein D, Rotnitzky A, Robins J. Adjusting for nonignorable drop-out using semiparametric nonresponse models: rejoinder. *Journal of the American Statistical Association*. 1999; 94:1135–1146.
- Scheike T, Zhang M. Extensions and applications of the cox-aalen survival model. *Biometrics*. 2003; 59:1036–1045. [PubMed: 14969483]
- Shen Y, Cheng S. Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*. 1999; 55:1093–1100. [PubMed: 11315053]
- Sierra J, Perez W, Rozman W, Carreras C, Klein J, Rizzo J, Davies J, Lazarus S, Bredeson C, Marks D, Canals C, Boogaerts M, Goldman J, Champlin R, Keating A, Weisdorf D, deWitte T, Horowitz M. Bone marrow transplantation from hla-identical siblings as treatment for myelodysplasia. *Blood*. 2002; 100:1997–2004. [PubMed: 12200358]

- Subramanian S. Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazards. *Journal of Statistical Planning and Inference*. 2000; 84:81–94.
- Subramanian S, Bandyopadhyay D. Doubly robust semiparametric estimation for the missing censoring indicator model. *Statistics and Probability Letters*. 2010; 80:621–630.
- van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press; Cambridge: 1998.
- White H. Maximum likelihood estimation of misspecified models. *Econometrica*. 1982; 50:1–25.

Appendix

Condition A

- (A.1) $\lambda_0(t)$ is continuous on $[0, \tau]$. The distribution of C is continuous on $[0, \tau]$ and $P(C > \tau) > 0$. The covariate processes $Z_i(t)$ have paths that are left continuous and of bounded variation, and satisfy the moment condition $E[\|Z_i(t)\|^4 \exp(2M\|Z_i(t)\|)] < \infty$, where M is a constant such that $\beta \in [-M, M]^p$ and $\|A\| = \max_{k,l} |a_{kl}|$ for a matrix $A = (a_{kl})$.
- (A.2) Each component of $s^{(j)}(t, \beta)$ is continuous on $[0, \tau] \times [-M, M]^p$ for $M > 0, j = 0, 1, 2$ and $s^{(0)}(t, \beta) > 0$ on $[0, \tau] \times [-M, M]^p$. $\sup_{t \in [0, \tau], \beta \in [-M, M]^p} \|S^{(j)}(t, \beta) - s^{(j)}(t, \beta)\| = O_p(n^{-1/2})$, and $\sup_{\psi \in [-L, L]^q} \sup_{t \in [0, \tau], \beta \in [-M, M]^p} \|S^{(j)}(t, \beta, \psi) - s^{(j)}(t, \beta)\| = O_p(n^{-1/2})$ for $j = 0, 1, 2$.
- (A.3) The matrix $\sum_{i=1}^n \int_0^\tau v(t, \beta_0) \lambda_0(t) s^{(0)}(t, \beta_0) dt$ is positive definite.
- (A.4) There is a $\sigma > 0$ such that $r(W_i) \geq \sigma$ for all i with $\Delta_i > 0$. Both $r(W_i, \psi)$ and $\rho(W_i, \gamma)$ are twice continuously differentiable with respect to ψ and γ , respectively. There exist ψ^* and γ^* satisfying the equations $ES_{\psi_i}^* = 0$ and $ES_{\gamma_i}^* = 0$, respectively, where $S_{\psi_i}^*$ and $S_{\gamma_i}^*$ are the corresponding score functions for $r(W_i, \psi)$ and $\rho(W_i, \gamma)$ given in (19). The information matrices I_{ψ}^* and I_{γ}^* also given in (19) are positive definite.

Proof of Theorem 1

Consistency of $\hat{\beta}_1$

Let ψ_0 be the true value of ψ such that $r(W_i) = r(W_i, \psi_0)$ under the correctly specified model for $r(W_i)$. Then $\hat{\psi} \xrightarrow{P} \psi_0$. Let

$$\xi_n(\beta, \psi_0) = n^{-1} \sum_{i=1}^n \int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) \frac{R_i}{\pi(Q_i, \psi_0)} N_i(dt).$$

When $r(W_i, \psi_0)$ is the correct model for $r(W_i)$, $\tilde{z}(t, \beta, \psi_0) = \bar{z}(t, \beta)$, where $\tilde{z}(t, \beta, \psi_0)$ is the limit of $Z(t, \beta, \psi_0)$. Under the conditions of Theorem 1, $n^{-1} U_I(\beta, \hat{\psi}) = \xi_n(\beta, \psi_0) + O_p(n^{-1/2})$ uniformly in $\beta \in [-M, M]^p$ for $M > 0$.

By application of the Glivenko–Cantelli and Donsker theorems, $\xi_n(\beta, \psi_0) = \zeta(\beta, \psi_0) + O_p(n^{-1/2})$ uniformly in $\beta \in [-M, M]^p$, where

$$\xi(\beta, \psi_0) = E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) \frac{R_i}{\pi(Q_i, \psi_0)} N_i(dt) \right].$$

When $r(W_i, \psi_0)$ is the correct model for $r(W_i)$, $\xi(\beta, \psi_0) = E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) N_i(dt) \right]$ by the double expectation formula $E[\cdot] = E\{E[\cdot | W_i, \Delta_i, \Delta_i > 0]\}$ and the missing at random assumption (2). Hence

$$n^{-1} U_I(\beta, \widehat{\psi}) \xrightarrow{P} E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) N_i(dt) \right]$$

uniformly in $\beta \in [-M, M]^p$ for $M > 0$.

Let $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i(s), 0 \leq s \leq t, i = 1, \dots, n\}$ be the σ -field generated by the observed information on the failure/censoring times and covariates up to time t . Then the intensity of the counting process $N_i(t)$ is given by $E\{N_i(dt) | \mathcal{F}_t\} = \lambda_1(t|Z_i(t))Y_i(t)dt$. Under the independent censoring assumption stated in the introduction, $\lambda_1(t|Z_i(t)) = \lambda_1^*(t|Z_i(t))$. Hence $M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda_1(u|Z_i(u))du = N_i(t) - \int_0^t Y_i(u)e^{\beta_0^\top Z_i(u)}\lambda_0(u)du$ is a martingale with respect to \mathcal{F}_t . Since $E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta_0)) Y_i(t) e^{\beta_0^\top Z_i(t)} \lambda_0(t) dt \right] = 0$, it follows that $\xi(\beta_0, \psi_0) = E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta_0)) M_i(dt) \right] = 0$. By Condition (A.3), β_0 is the unique solution to $\xi(\beta, \psi_0) = 0$. We have $\widehat{\beta}_I \xrightarrow{P} \beta_0$ by Theorem 5.9 of van der Vaart (1998).

Asymptotic Normality of $\widehat{\beta}_I$

Since $U_I(\widehat{\beta}_I, \widehat{\psi}) = 0$, $U_I(\widehat{\beta}_I, \widehat{\psi}) = 0$, $U_I(\beta_0, \widehat{\psi}) = -\frac{1}{n} \frac{\partial U_I(\beta, \widehat{\psi})}{\partial \beta} \Big|_{\beta=\widehat{\beta}} (\widehat{\beta}_I - \beta_0)$, where $\widehat{\beta}$ is on the line segment between $\widehat{\beta}_I$ and β_0 . We have

$$\sqrt{n}(\widehat{\beta}_I - \beta_0) = \left(-\frac{1}{n} \frac{\partial U_I(\beta, \widehat{\psi})}{\partial \beta} \Big|_{\beta=\widehat{\beta}} \right)^{-1} \frac{1}{\sqrt{n}} U_I(\beta_0, \widehat{\psi}).$$

By (A.4), $\widehat{\psi} \xrightarrow{P} \psi_0$ under correctly specified model for $r(W_i)$. By Condition (A.2) and the consistency of $\widehat{\beta}_I$, $\widetilde{V}(t, \widehat{\beta}, \widehat{\psi}) \xrightarrow{P} v(t, \beta_0)$ uniformly in $t \in (0, \tau]$. We have

$$\begin{aligned} -\frac{1}{n} \frac{\partial U_I(\beta, \widehat{\psi})}{\partial \beta} \Big|_{\beta=\widehat{\beta}} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \widetilde{V}(t, \widehat{\beta}, \widehat{\psi}) \frac{R_i}{r(W_i, \widehat{\psi})} dN_i(t) \xrightarrow{P} E \left[\int_0^\tau v(t, \beta_0) \frac{R_1}{r(W_1, \psi_0)} dN_1(t) \right] \\ &= \Sigma. \end{aligned}$$

Thus

$$\sqrt{n}(\widehat{\beta}_I - \beta_0) = (\Sigma + o_p(1))^{-1} n^{-1/2} U_I(\beta_0, \widehat{\psi}). \quad (7)$$

Let $\dot{\pi}(\cdot, \psi) = \partial\pi(\cdot, \psi)/\partial\psi$, $\dot{r}(\cdot, \psi) = \partial r(\cdot, \psi)/\partial\psi$, $\ddot{r}(\cdot, \psi) = \partial\dot{r}(\cdot, \psi)/\partial\psi^\top$, and

$$\begin{aligned} H^{(1)}(t, \beta, \psi) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta^\top Z_i(t)} \frac{R_i \dot{\pi}^\top(Q_i, \psi)}{\pi^2(Q_i, \psi)}, \\ H^{(2)}(t, \beta, \psi) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta^\top Z_i(t)} Z_i(t) \frac{R_i \dot{\pi}^\top(Q_i, \psi)}{\pi^2(Q_i, \psi)}, \\ h^{(1)}(t, \beta, \psi) &= E \left[Y_1(t) e^{\beta^\top Z_1(t)} \frac{R_1 \dot{\pi}^\top(Q_1, \psi)}{\pi^2(Q_1, \psi)} \right], \\ h^{(2)}(t, \beta, \psi) &= E \left[Y_1(t) e^{\beta^\top Z_1(t)} Z_1(t) \frac{R_1 \dot{\pi}^\top(Q_1, \psi)}{\pi^2(Q_1, \psi)} \right]. \end{aligned}$$

By the Taylor expansion of $U_I(\beta_0, \hat{\psi})$ around ψ_0 ,

$$n^{-1/2} \sqrt{n} U_I(\beta_0, \hat{\psi}) = n^{-1/2} U_I(\beta_0, \psi_0) + n^{-1} \frac{\partial U_I(\beta_0, \psi)}{\partial \psi} \Big|_{\psi=\psi_0} \sqrt{n}(\hat{\psi} - \psi_0) + o_p(1).$$

Note that

$$n^{-1} \frac{\partial U_I(\beta_0, \psi)}{\partial \psi} \Big|_{\psi=\psi_0} = V_\psi + o_p(1),$$

where

$$V_\psi = E \left[\int_0^\tau (Z_1(t) - \bar{z}(t, \beta_0)) \frac{R_1 \dot{\pi}^\top(Q_1, \psi_0)}{\pi^2(Q_1, \psi_0)} dN_1(t) - \left(\frac{h^{(2)}(t, \beta_0, \psi_0)}{\bar{s}^{(0)}(t, \beta_0, \psi_0)} - \frac{\bar{s}^{(1)}(t, \beta_0, \psi_0) h^{(1)}(t, \beta_0, \psi_0)}{(\bar{s}^{(0)}(t, \beta_0, \psi_0))^2} \right) \frac{R_1}{\pi(Q_1, \psi_0)} dN_1(t) \right].$$

It follows that

$$n^{-1/2} U_I(\beta_0, \hat{\psi}) = n^{-1/2} U_I(\beta_0, \psi_0) + V_\psi \sqrt{n}(\hat{\psi} - \psi_0) + o_p(1). \quad (9)$$

Since

$$\sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \psi_0)} (Z_i(t) - \bar{Z}(t, \beta_0, \psi_0)) Y_i(t) e^{\beta_0^\top Z_i(t)} d\Lambda_0(t) = 0,$$

we have

$$n^{-1/2} U_I(\beta_0, \psi_0) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \psi_0)} (Z_i(t) - \bar{Z}(t, \beta_0, \psi_0)) dM_i(t).$$

Under correctly specified $r(W_i)$, $E[R_i \{\pi(Q_i, \psi_0)\}^{-1} M_i(t)] = 0$. By Glivenko–Cantelli and Donsker theorems, $n^{-1/2} \sum_{i=1}^n R_i \{\pi(Q_i, \psi_0)\}^{-1} M_i(t)$ converges weakly to a mean zero

Gaussian process and $\tilde{Z}(t, \beta_0, \psi_0) \xrightarrow{P} \bar{z}(t, \beta_0)$ uniformly in t under the correctly specified model for $r(W_i)$. Applying Lemma 2 of Gilbert et al. (2008), we have

$$\int_0^\tau \{\tilde{Z}(t, \beta_0, \psi_0) - \bar{z}(t, \beta_0)\} d\{n^{-1/2} \sum_{i=1}^n \frac{R_i}{\pi(Q_i, \psi_0)} M_i(t)\} \xrightarrow{P} 0.$$

Hence

$$n^{-1/2} U_I(\beta_0, \psi_0) = \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \psi_0)} (Z_i(t) - \bar{z}(t, \beta_0)) dM_i(t) + o_p(1). \tag{10}$$

Let S_{ψ_i} and I_ψ be the score vector and the information matrix under the parametric model $r(W_i, \psi)$, respectively. Specifically,

$$S_{\psi_i} = \frac{I(\Delta_i > 0) \{R_i - r(W_i, \psi_0)\} \dot{r}(W_i, \psi_0)}{r(W_i, \psi_0) \{1 - r(W_i, \psi_0)\}},$$

$$I_\psi = E \left[S_{\psi_i} S_{\psi_i}^\top - \frac{I(\Delta_i > 0) \{R_i - r(W_i, \psi_0)\} \dot{r}(W_i, \psi_0)}{r(W_i, \psi_0) \{1 - r(W_i, \psi_0)\}} \right], \tag{11}$$

Under (A.4),

$$\widehat{\psi} - \psi_0 = n^{-1} \sum_{i=1}^n I_\psi^{-1} S_{\psi_i} + o_p(n^{-1/2}). \tag{12}$$

By (9), (10) and (12), we have

$$n^{-1/2} U_I(\beta_0, \widehat{\psi}) = n^{-1/2} \sum_{i=1}^n \omega_i + o_p(1), \tag{13}$$

where

$$\omega_i = \int_0^\tau \{Z_i(t) - \bar{z}(t, \beta_0)\} \frac{R_i}{\pi(Q_i, \psi_0)} dM_i(t) - V_\psi I_\psi^{-1} S_{\psi_i}.$$

By the central limit theorem, $n^{-1/2} U_I(\beta_0, \widehat{\psi})$ converges in distribution to a normal random vector with zero-mean and covariance matrix $E[\omega_1 \omega_1^T]$.

It follows by (7) and (13) that $\sqrt{n}(\widehat{\beta}_t - \beta_0)$ converges in distribution to a zero-mean Gaussian random vector with covariance matrix $\sum^{-1} E[\omega_1 \omega_1^T] \sum^{-1}$.

Proof of Theorem 2

Consistency of $\widehat{\beta}_A$

Under (A.4), there exist ψ^* and γ^* such that $\widehat{\psi} \xrightarrow{P} \psi^*$ and $\widehat{\gamma} \xrightarrow{P} \gamma^*$. Let

$$\eta_n(\beta, \psi^*, \gamma^*) = n^{-1} \sum_{i=1}^n \int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) \left[\frac{R_i}{r(W_i, \psi^*)} dN_i(t) - \frac{R_i - r(W_i, \psi^*)}{r(W_i, \psi^*)} \rho(W_i, \gamma^*) dN_i^*(t) \right].$$

Under Condition A, $n^{-1} U_A(\beta, \hat{\psi}, \hat{\gamma}) = \eta_n(\beta, \psi^*, \gamma^*) + O_p(n^{-1/2})$ uniformly in $\beta \in [-M, M]^p$ for $M > 0$.

By application of the Glivenko–Cantelli and Donsker theorems, $\eta_n(\beta, \psi^*, \gamma^*) = \eta(\beta, \psi^*, \gamma^*) + O_p(n^{-1/2})$ uniformly in $\beta \in [-M, M]^p$, where

$$\eta(\beta, \psi^*, \gamma^*) = E \left\{ \int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) \left[\frac{R_i}{r(W_i, \psi^*)} dN_i(t) - \frac{R_i - r(W_i, \psi^*)}{r(W_i, \psi^*)} \rho(W_i, \gamma^*) dN_i^*(t) \right] \right\}.$$

If at least one of $r(W_i, \psi_0)$ and $\rho(W_i, \gamma_0)$ is correctly specified for $r(W_i)$ and $\rho(W_i)$, then $\eta(\beta, \psi^*, \gamma^*) = E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) N_i(dt) \right]$ by the missing at random assumption (2) and by using the double expectation formula $E[\cdot] = E\{E[\cdot | W_i, \Delta_i, \Delta_i > 0]\}$ if $r(W_i)$ is correctly specified and using $E[\cdot] = E\{E[\cdot | W_i, R_i, \Delta_i > 0]\}$ if $\rho(W_i)$ is correctly specified. Hence

$$n^{-1} U_A(\beta, \hat{\psi}, \hat{\gamma}) \xrightarrow{P} E \left[\int_0^\tau (Z_i(t) - \bar{z}(t, \beta)) N_i(dt) \right]$$

uniformly in $\beta \in [-M, M]$ if at least one of $r(W_i, \psi_0)$ and $\rho(W_i, \gamma_0)$ is correctly specified for $r(W_i)$ and $\rho(W_i)$. Since $\eta(\beta_0, \psi^*, \gamma^*) = 0$ and β_0 is the unique solution to $\eta(\beta_0, \psi^*, \gamma^*) = 0$ by Condition (A.3), we have $\hat{\beta}_A \xrightarrow{P} \beta_0$ by Theorem 5.9 of van der Vaart (1998).

Asymptotic Normality of $\hat{\beta}_A$

Since $U_A(\beta, \hat{\psi}, \hat{\gamma}) = 0$, $U_A(\beta_0, \hat{\psi}, \hat{\gamma}) = -\frac{1}{n} \frac{\partial U_A(\beta, \hat{\psi}, \hat{\gamma})}{\partial \beta} \Big|_{\beta=\tilde{\beta}} (\hat{\beta}_A - \beta_0)$, where $\tilde{\beta}$ is on the line segment between $\hat{\beta}_A$ and β_0 . We have

$$\sqrt{n} (\hat{\beta}_A - \beta_0) = \left(-\frac{1}{n} \frac{\partial U_A(\beta, \hat{\psi}, \hat{\gamma})}{\partial \beta} \Big|_{\beta=\tilde{\beta}} \right)^{-1} \frac{1}{\sqrt{n}} U_A(\beta_0, \hat{\psi}, \hat{\gamma}).$$

By Condition (A.2) and the consistency of $\hat{\beta}_A$, $V(t, \tilde{\beta}) \xrightarrow{P} v(t, \beta_0)$ uniformly in t . Furthermore, by (A.4), $\hat{\psi} \xrightarrow{P} \psi^*$ and $\hat{\gamma} \xrightarrow{P} \gamma^*$, we have

$$-\frac{1}{n} \frac{\partial U_A(\beta, \hat{\psi}, \hat{\gamma})}{\partial \beta} \Big|_{\beta=\tilde{\beta}} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau V(t, \tilde{\beta}) \left[\frac{R_i}{r(W_i, \hat{\psi})} dN_i(t) - \frac{R_i - r(W_i, \hat{\psi})}{r(W_i, \hat{\psi})} \rho(W_i, \hat{\gamma}) dN_i^*(t) \right]$$

converges in probability to

$$\sum_A = E \left[\int_0^\tau V(t, \beta_0) \left(\frac{R_1}{r(W_1, \psi^*)} dN_1(t) - \frac{R_1 - r(W_1, \psi^*)}{r(W_1, \psi^*)} \rho(W_1, \gamma^*) dN_1^*(t) \right) \right],$$

which equals to $\sum = \int_0^\tau v(t, \beta_0) \lambda_0(t) s^{(0)}(t, \beta_0) dt$ if at least one of $r(W)$ and $\rho(W)$ is correctly specified. Thus

$$\sqrt{n}(\widehat{\beta}_A - \beta_0) = (\sum + o_p(1))^{-1} n^{-1/2} U_A(\beta_0, \widehat{\psi}, \widehat{\gamma}). \tag{14}$$

Let $\dot{r}(\cdot, \psi) = \partial r(\cdot, \psi) / \partial \psi$, $\ddot{r}(\cdot, \psi) = \partial \dot{r}(\cdot, \psi) / \partial \psi^\top$, $\dot{\rho}(\cdot, \gamma) = \partial \rho(\cdot, \gamma) / \partial \gamma$, and $\ddot{\rho}(\cdot, \gamma) = \partial \dot{\rho}(\cdot, \gamma) / \partial \gamma^\top$. By the Taylor expansion of $U_A(\beta_0, \widehat{\psi}, \widehat{\gamma})$ around ψ^* and γ^* ,

$$\begin{aligned} & n^{-1/2} U_A(\beta_0, \widehat{\psi}, \widehat{\gamma}) \\ &= n^{-1/2} U_A(\beta_0, \psi^*, \gamma^*) + n^{-1} \frac{\partial U_A(\beta_0, \psi, \gamma)}{\partial \psi} \Big|_{(\psi, \gamma) = (\psi^*, \gamma^*)} \sqrt{n}(\widehat{\psi} - \psi^*) \\ & \quad + n^{-1} \frac{\partial U_A(\beta_0, \psi, \gamma)}{\partial \gamma} \Big|_{(\psi, \gamma) = (\psi^*, \gamma^*)} \sqrt{n}(\widehat{\gamma} - \gamma^*) + o_p(1). \end{aligned}$$

Note that

$$\begin{aligned} n^{-1} \frac{\partial U_A(\beta_0, \psi, \gamma)}{\partial \gamma} \Big|_{(\psi, \gamma) = (\psi^*, \gamma^*)} &= P_\psi + o_p(1), \\ n^{-1} \frac{\partial U_A(\beta_0, \psi, \gamma)}{\partial \psi} \Big|_{(\psi, \gamma) = (\psi^*, \gamma^*)} &= P_\gamma + o_p(1), \end{aligned}$$

where

$$\begin{aligned} P_\psi &= E \left[\int_0^\tau \{Z_1(t) - \bar{z}(t, \beta_0)\} \frac{R_1 \dot{r}^\top(W_1, \psi^*)}{r^2(W_1, \psi^*)} \{dN_1(t) - \rho(W_1, \gamma^*) dN_1^*(t)\} \right], \\ P_\gamma &= E \left[\int_0^\tau \{Z_1(t) - \bar{z}(t, \beta_0)\} \frac{R_1 - r(W_1, \psi^*)}{r(W_1, \psi^*)} \dot{\rho}^\top(W_1, \gamma^*) dN_1^*(t) \right]. \end{aligned} \tag{15}$$

It follows that

$$n^{-1/2} U_A(\beta_0, \widehat{\psi}, \widehat{\gamma}) = n^{-1/2} U_A(\beta_0, \psi^*, \gamma^*) + P_\psi \sqrt{n}(\widehat{\psi} - \psi^*) + P_\gamma \sqrt{n}(\widehat{\gamma} - \gamma^*) + o_p(1). \tag{16}$$

Let $M_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\beta_0^\top Z_i(u)} d\Lambda_0(u)$, and

$$M_i^*(t) = \frac{R_i}{r(W_i, \psi^*)} N_i(t) - \frac{R_i - r(W_i, \psi^*)}{r(W_i, \psi^*)} \rho(W_i, \gamma^*) N_i^*(t) - \int_0^t Y_i(u) e^{\beta_0^\top Z_i(u)} d\Lambda_0(u). \tag{17}$$

Since $\sum_{i=1}^n \int_0^\tau (Z_i(t) - \bar{Z}(t, \beta_0)) Y_i(t) e^{\beta_0^\top Z_i(t)} d\Lambda_0(t) = 0$, it follows that

$$n^{-1/2}U_A(\beta_0, \psi^*, \gamma^*) = n^{-1/2} \sum_{i=1}^n \int_0^\tau (Z_i(t) - \bar{Z}(t, \beta_0)) dM_i^*(t).$$

Similar to the arguments given above, $E[M_i^*(t)] = 0$ if at least one of $r(W_i)$ and $\rho(W_i)$ is correctly specified. By application of the Glivenko-Cantelli and Donsker theorems, $n^{-1/2}M_i^*(t)$ converges weakly to a zero-mean Gaussian process and $\bar{Z}(t, \beta_0) \xrightarrow{P} \bar{z}(t, \beta_0)$ uniformly in t . Applying Lemma 2 of Gilbert et al. (2008), we have

$$\int_0^\tau \{\bar{Z}(t, \beta_0) - \bar{z}(t, \beta_0)\} d\{n^{-1/2} \sum_{i=1}^n M_i^*(t)\} \xrightarrow{P} 0.$$

Hence

$$U_A(\beta_0, \psi^*, \gamma^*) = \sum_{i=1}^n \int_0^\tau (Z_i(t) - \bar{z}(t, \beta_0)) dM_i^*(t) + o_p(1). \quad (18)$$

Let $S_{\psi i}^*$ and I_{ψ}^* be the score vector and the information matrix under the parametric model $r(W_i, \psi)$, respectively, and $S_{\gamma i}^*$ and I_{γ}^* the score vector and the information matrix under the parametric model $\rho(W_i, \gamma)$. Specifically,

$$\begin{aligned} S_{\psi i}^* &= \frac{I(\Delta_i > 0)(R_i - r(W_i, \psi^*))\dot{r}(W_i, \psi^*)}{r(W_i, \psi^*)\{1 - r(W_i, \psi^*)\}}, \\ S_{\gamma i}^* &= \frac{R_i I(\Delta_i > 0)\{I(\Delta_i = 1) - \rho(W_i, \gamma^*)\}\dot{\rho}(W_i, \gamma^*)}{\rho(W_i, \gamma^*)\{1 - \rho(W_i, \gamma^*)\}}, \\ I_{\psi}^* &= E \left[S_{\psi 1} S_{\psi 1}^\top - \frac{I(\Delta_1 > 0)(R_1 - r(W_1, \psi^*))\dot{r}(W_1, \psi^*)}{r(W_1, \psi^*)\{1 - r(W_1, \psi^*)\}} \right], \\ I_{\gamma}^* &= E \left[S_{\gamma 1} S_{\gamma 1}^\top - \frac{R_1 I(\Delta_1 > 0)\{I(\Delta_1 = 1) - \rho(W_1, \gamma^*)\}\dot{\rho}(W_1, \gamma^*)}{\rho(W_1, \gamma^*)\{1 - \rho(W_1, \gamma^*)\}} \right]. \end{aligned} \quad (19)$$

Under (A.4),

$$\widehat{\psi} - \psi^* = n^{-1} \sum_{i=1}^n (I_{\psi}^*)^{-1} S_{\psi i}^* + o_p(n^{-1/2}) \quad (20)$$

and

$$\widehat{\gamma} - \gamma^* = n^{-1} \sum_{i=1}^n (I_{\gamma}^*)^{-1} S_{\gamma i}^* + o_p(n^{-1/2}). \quad (21)$$

By (16), (18), (20), and (21), we have

$$n^{-1/2}U_A(\beta_0, \widehat{\psi}, \widehat{\gamma}) = n^{-1/2} \sum_{i=1}^n \varphi_i + o_p(1), \quad (22)$$

where

$$\varphi_i = \int_0^{\tau} \{Z_i(t) - \bar{z}(t, \beta_0)\} dM_i^*(t) - P_{\psi}(I_{\psi}^*)^{-1} S_{\psi i}^* - P_{\gamma}(I_{\gamma}^*)^{-1} S_{\gamma i}^*.$$

By the central limit theorem, $n^{-1/2}U_A(\beta_0, \hat{\psi}, \hat{\gamma})$ converges in distribution to a normal random vector with zero-mean and covariance matrix $E[\varphi_1 \varphi_1^T]$.

It follows by (14) and (22) that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution to a zero-mean Gaussian random vector with covariance matrix $\sum^{-1} E[\varphi_1 \varphi_1^T] \sum^{-1}$.

Table 1

Summary statistics of simulation results with 20% missingness under various model specifications of $\rho(\cdot)$.

Estimator	$n = 200$					$n = 400$				
	Bias	SSE	MSE	CP	CP	Bias	SSE	MSE	CP	CP
β_C	-0.136	0.442	0.447	96.0	96.0	-0.129	0.322	0.313	92.6	92.6
β_{Ic}	-0.007	0.422	0.426	96.4	96.4	-0.002	0.313	0.299	93.9	93.9
β_{In}	-0.107	0.430	0.430	96.1	96.1	-0.182	0.392	0.385	92.2	92.2
	(Model 1)					(Model 1)				
β_{Ac}	-0.005	0.406	0.416	96.2	96.2	0.002	0.301	0.292	94.4	94.4
β_{An}	-0.003	0.403	0.411	96.5	96.5	0.003	0.351	0.345	95.1	95.1
β_{M^1}	-0.003	0.427	0.437	96.3	96.3	0.003	0.313	0.306	94.3	94.3
β_{M^5}	-0.005	0.406	0.419	96.6	96.6	0.005	0.302	0.294	94.4	94.4
	(Model 2)					(Model 2)				
β_{Ac}	-0.005	0.406	0.416	96.1	96.1	0.002	0.301	0.292	94.4	94.4
β_{An}	-0.007	0.403	0.412	96.4	96.4	0.001	0.297	0.290	94.3	94.3
β_{M^1}	-0.012	0.428	0.438	96.2	96.2	-0.003	0.314	0.307	94.3	94.3
β_{M^5}	-0.011	0.408	0.420	96.6	96.6	-0.001	0.303	0.295	94.0	94.0
	(Model 3)					(Model 3)				
β_{Ac}	-0.004	0.406	0.416	96.2	96.2	0.002	0.300	0.292	94.5	94.5
β_{An}	0.005	0.395	0.405	96.7	96.7	0.010	0.292	0.286	94.5	94.5
β_{M^1}	0.061	0.370	0.383	96.3	96.3	0.066	0.273	0.269	93.8	93.8
β_{M^5}	0.057	0.347	0.364	96.2	96.2	0.067	0.260	0.255	93.8	93.8
	(Model 4)					(Model 4)				
β_{Ac}	-0.005	0.406	0.416	96.2	96.2	0.002	0.301	0.292	94.5	94.5
β_{An}	-0.001	0.397	0.406	96.7	96.7	0.009	0.341	0.336	95.2	95.2
β_{M^1}	0.056	0.367	0.383	96.5	96.5	0.062	0.273	0.269	94.1	94.1
β_{M^5}	0.054	0.348	0.364	96.5	96.5	0.064	0.261	0.256	93.5	93.5

Bias, the mean of the estimates of β ; SSE, the sample standard error of the estimates of β ; MSE, the mean of the standard error estimates; CP, the empirical coverage probability of the corresponding 95% confidence intervals; β_C , the complete-case estimator; β_{IC} and β_{Im} , the IPW estimators; β_{Ac} and β_{Am} , the AIPW estimators; β_{M^1} and β_{M^5} , the multiple imputation estimators with $m = 1$ and $m = 5$ imputations. Here c denotes the correctly specified model and m the misspecified model for $r(\cdot)$, respectively.

Table 2

Summary statistics of simulation results with 45% missingness under various model specifications of $\rho(\cdot)$.

Estimator	$n = 200$					$n = 400$				
	Bias	SSE	MSE	CP	CP	Bias	SSE	MSE	CP	CP
β_C	-0.322	0.595	0.599	94.0	94.0	-0.305	0.420	0.416	88.1	88.1
β_{Ic}	-0.014	0.583	0.554	93.7	93.7	0.002	0.411	0.394	94.7	94.7
β_{In}	-0.194	0.561	0.552	94.6	94.6	-0.182	0.392	0.385	92.2	92.2
	(Model 1)					(Model 1)				
β_{Ac}	-0.022	0.540	0.523	94.3	94.3	0.001	0.380	0.365	93.4	93.4
β_{An}	-0.011	0.493	0.494	95.8	95.8	0.003	0.351	0.345	95.1	95.1
β_{M^1}	-0.007	0.534	0.539	96.2	96.2	-0.006	0.377	0.376	95.2	95.2
β_{M^5}	-0.012	0.502	0.509	96.6	96.6	0.001	0.356	0.354	95.1	95.1
	(Model 2)					(Model 2)				
β_{Ac}	-0.021	0.539	0.523	94.0	94.0	0.001	0.379	0.365	93.6	93.6
β_{An}	-0.017	0.495	0.496	95.9	95.9	-0.003	0.352	0.347	94.8	94.8
β_{M^1}	-0.022	0.538	0.541	96.5	96.5	-0.017	0.377	0.377	95.2	95.2
β_{M^5}	-0.025	0.506	0.510	96.7	96.7	-0.010	0.356	0.355	94.8	94.8
	(Model 3)					(Model 3)				
β_{Ac}	-0.020	0.538	0.524	94.9	94.9	0.002	0.379	0.365	94.0	94.0
β_{An}	0.011	0.463	0.471	96.0	96.0	0.017	0.338	0.333	95.1	95.1
β_{M^1}	0.137	0.368	0.383	94.6	94.6	0.133	0.266	0.268	92.3	92.3
β_{M^5}	0.130	0.325	0.340	94.6	94.6	0.139	0.237	0.238	91.1	91.1
	(Model 4)					(Model 4)				
β_{Ac}	-0.020	0.536	0.525	94.6	94.6	0.002	0.379	0.365	94.2	94.2
β_{An}	0.004	0.469	0.476	96.2	96.2	0.009	0.341	0.336	95.2	95.2
β_{M^1}	0.131	0.371	0.384	95.2	95.2	0.127	0.268	0.269	92.5	92.5
β_{M^5}	0.126	0.327	0.340	94.7	94.7	0.134	0.239	0.239	91.5	91.5

Bias, the mean of the estimates of β ; SSE, the sample standard error of the estimates of β ; MSE, the mean of the standard error estimates; CP, the empirical coverage probability of the corresponding 95% confidence intervals; β_C , the complete-case estimator; β_{IC} and β_{Im} , the IPW estimators; β_{Ac} and β_{Am} , the AIPW estimators; β_{M^1} and β_{M^5} , the multiple imputation estimators with $m = 1$ and $m = 5$ imputations. Here c denotes the correctly specified model and m the misspecified model for $r(\cdot)$, respectively.

Table 3
 Estimation of the effects of platelet and age for the bone marrow transplant data.

Missing	Estimator	Platelet			Age		
		Est.	SEE	p-value	Est.	SEE	p-value
None	β_F	-0.586	0.186	0.002	0.367	0.087	<0.001
	β_C	-0.491	0.236	0.037	0.397	0.112	<0.001
MCAR	β_A	-0.520	0.203	0.010	0.387	0.091	<0.001
	β_{M^5}	-0.530	0.205	0.010	0.380	0.096	<0.001
	β_C	-0.504	0.238	0.034	0.241	0.107	0.024
MAR	β_A	-0.566	0.204	0.006	0.380	0.090	<0.001
	β_{M^5}	-0.614	0.205	0.003	0.380	0.093	<0.001
	β_C	-0.390	0.253	0.123	0.156	0.114	0.171
NMAR	β_A	-0.529	0.220	0.016	0.350	0.098	<0.001
	β_{M^5}	-0.567	0.219	0.010	0.344	0.099	<0.001

Est., the estimate; SEE, the standard error estimate; p-value pertaining to testing no covariate effect; β_F , the full-case estimator with no missing causes; β_C , the complete-case estimator; β_A , the AIPW estimator; β_{M^5} , the multiple imputation estimator with $m = 5$ imputations.