HHS Public Access
Author manuscript
Lifetime Data Anal. Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:
Lifetime Data Anal. 2018 April ; 24(2): 224–249. doi:10.1007/s10985-017-9398-z.

Reweighted estimators for additive hazard model with censoring indicators missing at random

Xiaolin Chen and
School of Statistics, Qufu Normal University, Qufu, 273165, China

Jianwen Cai
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, U.S.A

Abstract

Survival data with missing censoring indicators are frequently encountered in biomedical studies. In this paper, we consider statistical inference for this type of data under the additive hazard model. Reweighting methods based on simple and augmented inverse probability are proposed. The asymptotic properties of the proposed estimators are established. Furthermore, we provide a numerical technique for checking adequacy of the fitted model with missing censoring indicators. Our simulation results show that the proposed estimators outperform the simple and augmented inverse probability weighted estimators without reweighting. The proposed methods are illustrated by analyzing a dataset from a breast cancer study.

Key words and phrases

Additive hazard model; Censored data; Inverse probability weighted estimator; Missing censoring indicators; Reweighting

1. Introduction

The analysis of time-to-event data, i.e., survival analysis, is frequently encountered in biomedical studies and clinical trials, where the time-to-event of interest (survival time) is subject to right censoring. For some subjects, however, the censoring indicator may be missing due to various reasons, for example, the medical records are missing.

Under such situation, the standard statistical inference methods can no longer be applied directly. A naive method for analyzing such kind of data is to simply ignore the subjects with missing censoring indicator and make inference using data from those with observed censoring indicator. However, this so-called complete-case (CC) analysis could cause loss in efficiency. Moreover, it may lead to inconsistent estimators when the censoring indicator is not missing completely at random. To overcome drawbacks of CC analysis, some methods were proposed to analyze the survival data with missing censoring indicator, such as regression calibration, multiple imputation, simple and augmented inverse probability weighting and so on. For the relevant terminology, especially the simple and augmented inverse probability weighting (SW and AW), the readers are referred to Tsiatis (2006).

Related literature for survival data with missing censoring indicator includes McKeague and Subramanian (1998), Lu and Tsiatis (2001), Liu and Wang (2010), Hyun et al. (2012) among others. McKeague and Subramanian (1998) proposed a method of estimation when the missing mechanism is MCAR, applicable both in the nonparametric setting and under the semi-parametric proportional hazard model. Liu and Wang (2010) suggested a regression imputation method for the Cox proportional hazard model. Hyun et al. (2012) considered competing risk data with missing causes of failure under the proportional hazard model. Their methods are based on the simple and augmented inverse probability weighting ideas. The simple and augmented inverse probability weighting methods were also applied to the quantile regression model by Sun et al. (2012). Lu and Tsiatis (2001) studied the multiple imputation methods under the proportional hazard model. Wang and Dinse (2010) and Li and Wang (2012) considered several methods under the censored linear regression model, such as regression calibration, multiple imputation and so on.

An important and useful alternative to the Cox proportional hazard model is the additive hazard model, which assumes that the conditional hazard function of survival time T has the form:

$$\lambda(t|Z)=\lambda_0(t)+\beta_0^T Z, \quad (1)$$

where $\lambda_0(t)$ is the baseline hazard function, Z is a p-dimensional covariate vector and β_0 is p-vector of regression coefficients. In contrast to the Cox proportional hazard model, the additive hazard model specifies conditional hazard function as the sum of, instead of the product of, baseline hazard function and regression function of covariates. It has sound biological and empirical bases (Breslow and Day, 1980), and the desirable interpretation of the regression coefficients as hazard differences. Also, the computations for the additive hazard model do not require iteration and do not give rise to numerical problems. Moreover, in many practical applications, the additive hazard model is more appropriate than the multiplicative, particularly with respect to continuous covariates.

Due to these considerations, many authors have investigated inference methods for survival data with missing censoring indicator under the additive hazard model. Zhou and Sun (2003) studied this problem by extending the idea of McKeague and Subramanian (1998). To overcome the limitation of MCAR assumption, Lu and Liang (2008) considered the similar problem under the framework of competing risk data using the SW and AW methods when the cause of death is missing at random (MAR). SW method utilizes the inverse of probability that censoring indicator is observed to enlarge the representation of an individual, while AW method added an augmentation term obtained by considering additional information from missing censoring indicators. Noting that in Lu and Liang (2008) two auxiliary parametric models were imposed for the observation probability and the conditional probability of an uncensored observation respectively, Song et al. (2010) proposed the kernel-assisted SW and AW methods to avoid model misspecification.

Tsiatis (2006) noted that if some observation probabilities are small, i.e., heavy weights are imposed on these observations, the SW estimators will be unstable and have inflated variances. Sometimes even AW estimators could have poor performances, as can be seen in our simulation studies. A simple and efficient method of reducing influence of overweighting is to assign another weight function to the complete observation to offset overweighting caused by a small observation probability. We refer to this method as reweighting. In this paper, based on a weight function suggested in Xu et al. (2009), we propose new approaches to analyze survival data with missing censoring indicators under the additive hazard model. We consider simple and augmented inverse probability weighting with reweighting. We refer to them as the simple and augmented reweighting (SRW and ARW) methods subsequently. As was noted by a referee, the proposed reweighting procedure can be deemed as a two-stage approach, in which the first weighting is done on each complete observation by the inverse of the observation probability, followed by a further adjustment by a new weight function on the pseudo-unbiased sample generated from previous step. Under the MAR assumption, the asymptotic properties of the proposed estimators are established. Our simulation results show that the proposed estimators significantly improve the SW and AW estimators respectively. Although we focus on the survival data with missing censoring indicators, the methods proposed here can be easily generalized to the competing risk data with missing causes of death.

The rest of the article is organized as follows. In Section 2, SRW and ARW methods are developed and the asymptotic properties of the proposed estimators are presented. Section 3 provides a simple procedure to check the adequacy of the additive hazard model with missing censoring indicators. We report the simulation results and illustrate the proposed procedures by applying them to a dataset from a clinical trial in Sections 4 and 5, respectively. A brief discussion is included in Section 6. The regularity conditions and the outlines of the proofs of the theorems are given in the Appendix.

2. The reweighting method

Let T denote the survival time, C the censoring time, $X = T \wedge C$ the observed time and $\delta = I(T \leq C)$ the failure indicator, where $I(A)$ is the indicator function of the set A. Suppose that T is conditionally independent of C given Z. Define $\xi = 1$ if δ is observed, and $\xi = 0$ otherwise. Throughout the paper, we assume that δ is missing at random, thus the observation probability is

$$\Pr(\xi=1|X,Z,R,\delta)=\Pr(\xi=1|X,Z,R) \equiv \pi(W),$$

where R is an auxiliary covariate which is used to predict the observation probability and $W = (X, Z, R)$. Furthermore, we assume that $\pi(W)$ could be modeled by a parametric model $\pi(W, \alpha)$, where α is an unknown vector of finite-dimensional parameters, e.g. the logistic regression model. The observed data are independent and identically distributed (i.i.d.) random vector $(X_i, \xi_i, \xi_i\delta_i, Z_i, R_i)$, $i = 1, \ldots, n$. Define $W_i = (X_i, Z_i, R_i)$ for $i = 1, \ldots, n$.

To facilitate expressions using counting processes, a few more notations are needed. Let $N_i(t) = I(X_i \leq t, \delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$ be the counting and at risk processes for the ith subject, respectively. Define $M_i(t)=N_i(t) - \int_0^t Y_i(s)\{d\Lambda_0(s)+\beta_0^T Z_i ds\}$ for $0 \leq t \leq \tau$, where

$\Lambda_0(t)$ is the true cumulative hazard function and τ is the maximum follow-up time. Using the fact that $M_i(t)$, $i = 1, \ldots, n$ are zero-mean martingales with respect to the σ-filtration $\sigma\{N_i(u), Y_i(u+), Z_i : 0 \le u \le t, i = 1, \ldots, n\}$, when the censoring indicators are fully observed, Lin and Ying (1994) proposed the following estimating function for β,

$$U_n(\beta)=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\overline{Z}(t)\}\{\mathrm{dN}_i(t)-Y_i(t)\beta^T Z_i \mathrm{dt}\}, \quad (2)$$

where $\bar{Z}(t) = S^{(1)}(t)/S^{(0)}(t)$ with $S^{(k)}(t)=n^{-1}\sum_{i=1}^{n} Y_i(t) Z_i^{\otimes k}$ for $k = 0$ and 1. Here $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^T$ for a vector a.

2.1 Simple reweighting method

When some censoring indicators are missing, the estimating function (2) is not applicable. Motivated by the inverse probability weighting idea of Horvitz and Thompson (1952), a routine choice is to weigh the complete observations with the inverse of their observation probabilities to construct an unbiased estimating function (Lu and Liang, 2008; Song et al., 2010). However, this method will not work well when some of the observation probabilities are close to zero as mentioned in Section 1. In this paper, we take the further step of reweighting the inverse probability weighted observations by another weight function, proposed by Xu et al. (2009), to offset the overweighting caused by the small observation probabilities. Specifically, in addition to the observation probability, the reweighting method imposes an additional weighting function, i.e., the marginal observation probability given the risk set at time t, to the complete observations. Hence the additional weighting function is the same for each observation in the same risk set at time t, denoted by $\pi^*(t)$. As in Xu et al. (2009), we estimate $\pi^*(t)$ by the empirical estimator, i.e.,

$$\hat{\pi}^*(t)=\frac{\sum_{i=1}^{n}\xi_i Y_i(t)}{\sum_{j=1}^{n} Y_j(t)}.$$

Combining the simple inverse probability weighting idea of Horvitz and Thompson (1952), the reweighting idea of Xu et al. (2009) and the estimating function (2), the simple reweighting estimating function can be obtained as

$$U_{\mathrm{sr}}(\beta,\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\xi_i\frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{\mathrm{sr}}(t,\hat{\alpha})\}\{\mathrm{dN}_i(t)-Y_i(t)\beta^T Z_i \mathrm{dt}\}, \quad (3)$$

where $\overline{Z}_{\mathrm{sr}}(t,\hat{\alpha})=S_{\mathrm{sr}}^{(1)}(t,\hat{\alpha})/S_{\mathrm{sr}}^{(0)}(t,\hat{\alpha})$ with $S_{\mathrm{sr}}^{(k)}(t,\hat{\alpha})=n^{-1}\sum_{i=1}^{n}\xi_i\pi^{-1}(W_i,\hat{\alpha})Y_i(t)Z_i^{\otimes k}$ for $k = 0$, 1 and $\hat{\alpha}$ is the maximum likelihood estimator (MLE) of α. It is well known that $\hat{\alpha}$ is maximizer of the following likelihood

$$\prod_{i=1}^{n}\pi(W_i,\alpha)^{\xi_i}\{1-\pi(W_i,\alpha)\}^{(1-\xi_i)},$$

or solution of the following score equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i-\pi(W_i,\alpha)}{\pi(W_i,\alpha)\{1-\pi(W_i,\alpha)\}}\frac{\partial\pi(W_i,\alpha)}{\partial\alpha}=0,$$

and $\hat{\alpha}$ consistently estimate α_0 when $\pi(W, \alpha_0)$ is correctly specified for $\pi(W)$. Denote by $\hat{\beta}_{sr}$ the solution of the estimating equation $U_{sr}(\beta, \hat{\alpha}) = 0$, which has the closed form

$$\hat{\beta}_{\mathrm{sr}}=\left[\sum_{i=1}^{n}\int_0^{\tau}\xi_i\frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{\mathrm{sr}}(t,\hat{\alpha})\}^{\otimes 2}Y_i(t)\mathrm{d}t\right]^{-1}\times\left[\sum_{j=1}^{n}\int_0^{\tau}\xi_j\frac{\hat{\pi}^*(t)}{\pi(W_j,\hat{\alpha})}\{Z_j-\overline{Z}_{\mathrm{sr}}(t,\hat{\alpha})\}\mathrm{dN}_j(t)\right].$$

With the estimator $\hat{\beta}_{sr}$ of β_0, the cumulative baseline hazard function $\Lambda_0(t)$ can be estimated by

$$\hat{\Lambda}_{\mathrm{sr}}(t)=\frac{1}{n}\sum_{i=1}^{n}\int_0^{t}\frac{\xi_i}{\pi(W_i,\hat{\alpha})}\frac{1}{S_{\mathrm{sr}}^{(0)}(s,\hat{\alpha})}\{\mathrm{dN}_i(s)-Y_i(s)\hat{\beta}_{\mathrm{sr}}^{T}Z_i\mathrm{ds}\}. \quad (4)$$

To facilitate the presentation of the asymptotic results, we introduce the following notations. Define

$$s^{(k)}(t)=E\{Y(t)Z^{\otimes k}\},\qquad k=0,1,2,$$

$$\overline{z}(t)=s^{(1)}(t)/s^{(0)}(t),$$

$$A=E\left[\int_0^{\tau}\pi^*(t)\{Z-\overline{z}(t)\}^{\otimes 2}Y(t)\mathrm{d}t\right].$$

We study the asymptotic distribution in the following Theorem 1, proof of which is provided in the Appendix.

Theorem 1—Under regularity conditions (C1) to (C4) in the Appendix, if $\pi(W, \alpha_0)$ is correctly specified for $\pi(W)$, $\hat{\beta}_{sr}$ is a consistent estimator of β_0, and $n^{\frac{1}{2}}(\hat{\beta}_{\mathrm{sr}}-\beta_0)$ is asymptotically normal with mean zero and covariance matrix $A_{\mathrm{sr}}^{-1}\sum_{\mathrm{sr}}A_{\mathrm{sr}}^{-1}$, where

$$\sum_{sr}=E\left\{\left[\int_0^\tau \xi \frac{\pi^*(t)}{\pi(W,\alpha_0)}\{Z-\bar{z}(t)\}\mathrm{dM}(t)-V_{\alpha_0}I_{\alpha_0}^{-1}S_{\alpha_0}\right]^{\otimes 2}\right\}$$

with $V_{\alpha 0}$, $I_{\alpha 0}$ and $S_{\alpha 0}$ being given in the Appendix.

The asymptotic covariance matrix can be consistently estimated by $\hat{A}_{sr}^{-1}\hat{\sum}_{sr}\hat{A}_{sr}^{-1}$, where

$$\hat{A}_{sr}=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \xi_i \frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{sr}(t,\hat{\alpha})\}^{\otimes 2}Y_i(t)\mathrm{dt}.$$

$$\hat{\sum}_{sr}=\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \xi_i \frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{sr}(t,\hat{\alpha})\}d\hat{M}_i(t)-\hat{V}_{\alpha_0}\hat{I}_{\alpha_0}^{-1}\hat{S}_{\alpha_0,i}\right]^{\otimes 2}$$

with $\hat{M}_i(t)=N_i(t)-\int_0^\tau Y_i(s)\{d\hat{\Lambda}_{sr}(s)+\hat{\beta}_{sr}^T Z_i \mathrm{ds}\}$, $\hat{V}_{\alpha 0}$, $\hat{I}_{\alpha 0}$ and $\hat{S}_{\alpha 0,i}$ being obtained by their empirical counterparts and replacing (α_0, β_0) by $(\hat{\alpha}, \hat{\beta}_{sr})$ accordingly.

2.2 Augmented reweighting method

It is well-known that the AW estimators are more efficient than the SW estimators because they incorporate more information from data and enjoy the double robustness property, i.e. they are consistent if either the model for the observation probability or the conditional probability of missing data given observed data is correctly specified (Robins et al., 1994). Here we further consider the augmented reweighting method to improve the simple reweighting method.

Define $\rho(W) = \Pr(\delta = 1 | W)$. Under the MAR assumption, it can be shown that

$$\rho(W)=\Pr(\delta=1|W,\xi=1). \quad (5)$$

Furthermore, a parametric model $\rho(W, \gamma)$ can be posited for $\rho(W)$, in which the true value γ_0 can be estimated by the complete observations according to (5). Specifically, the maximum likelihood estimator $\hat{\gamma}$ is maximizer of the following likelihood function

$$\prod_{i=1}^{n}\rho(W_i,\gamma)^{\xi_i\delta_i}\{1-\rho(W_i,\gamma)\}^{\xi_i(1-\delta_i)},$$

or solution of the following score equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i(\delta_i-\rho(W_i,\gamma))}{\rho(W_i,\gamma)\{1-\rho(W_i,\gamma)\}}\frac{\partial\rho(W_i,\gamma)}{\partial\gamma}=0.$$

Adding the augmentation term to the estimating function (3), we propose the following augmented reweighting estimating function

$$U_{\mathrm{ar}}(\beta,\hat{\alpha},\hat{\gamma})=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}\{Z_i-\overline{Z}_{\mathrm{ar}}(t)\}\hat{\pi}^*(t)\{\frac{\xi_i}{\pi(W_i,\hat{\alpha})}\mathrm{dN}_i(t)$$
$$+(1-\frac{\xi_i}{\pi(W_i,\hat{\alpha})})\rho(W_i,\hat{\gamma})\mathrm{dN}_i^*(t)-Y_i(t)\beta^TZ_i\mathrm{dt}\}, \qquad (6)$$

where $N_i^*(t)=I(X_i\le t)$ and $\overline{Z}_{\mathrm{ar}}(t)=S_{\mathrm{ar}}^{(1)}(t)/S_{\mathrm{ar}}^{(0)}(t)$ with $S_{\mathrm{ar}}^{(k)}(t)=n^{-1}\sum_{i=1}^{n}Y_i(t)Z_i^{\otimes k}$ for $k = 0, 1, 2$. Denote by $\hat{\beta}_{ar}$ the solution of the estimating equation $U_{ar}(\beta, \hat{\alpha}, \hat{\gamma}) = 0$, which has the following explicit form

$$\hat{\beta}_{\mathrm{ar}}$$
$$=\left[\sum_{i=1}^{n}\int_{0}^{\tau}\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\mathrm{ar}}(t)\}^{\otimes 2}Y_i(t)\mathrm{dt}\right]^{-1}$$
$$\times\left[\sum_{j=1}^{n}\int_{0}^{\tau}\{Z_j-\overline{Z}_{\mathrm{ar}}(t)\}\hat{\pi}^*(t)\{\frac{\xi_j}{\pi(W_j,\hat{\alpha})}\mathrm{dN}_j(t)+(1-\frac{\xi_j}{\pi(W_j,\hat{\alpha})})\rho(W_j,\hat{\gamma})\mathrm{dN}_j^*(t)\}\right].$$

Then $\Lambda_0(t)$ can be estimated by

$$\hat{\Lambda}_{\mathrm{ar}}(t)=\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{t}\frac{1}{S_{\mathrm{ar}}^{(0)}(s)}\{\frac{\xi_i}{\pi(W_i,\hat{\alpha})}\mathrm{dN}_i(s)$$
$$+(1-\frac{\xi_i}{\pi(W_i,\hat{\alpha})})\rho(W_i,\hat{\gamma})\mathrm{dN}_i^*(s)-Y_i(s)\hat{\beta}_{\mathrm{ar}}^TZ_i\mathrm{ds}\}.$$

It is well-known that there exist α_* and γ_* such that $\hat{\alpha}\xrightarrow{\mathrm{P}}\alpha_*$ and $\hat{\gamma}\xrightarrow{\mathrm{P}}\gamma_*$ under the conditions (C4) and (C5) in the Appendix (White, 1982). Furthermore, $\alpha_* = \alpha_0$ if $\pi(W, \alpha_0)$ is correctly specified, while $\gamma_* = \gamma_0$ if $\rho(W, \gamma_0)$ is correctly specified. The asymptotic results of regression coefficients are presented in the following theorem with the proof being postponed to the Appendix.

Theorem 2—Under regularity conditions (C1) to (C5) in the Appendix, if either $\pi(W, \alpha_0)$ or $\rho(W, \gamma_0)$ is correctly specified, $\hat{\beta}_{ar}$ is a consistent estimator of β_0, and $n^{\frac{1}{2}}(\hat{\beta}_{\mathrm{ar}}-\beta_0)$ is asymptotically normal with mean zero and covariance matrix $A_{\mathrm{ar}}^{-1}\sum_{\mathrm{ar}}A_{\mathrm{ar}}^{-1}$, where

$$\sum\nolimits_{\mathrm{ar}}=E\left\{\left[\int_{0}^{\tau}\pi^*(t)\{Z-\overline{z}(t)\}\mathrm{dM}^*(t)-V_{\alpha_*}^*(I_{\alpha_*}^*)^{-1}S_{\alpha_*}^*-V_{\gamma_*}^*(I_{\gamma_*}^*)^{-1}S_{\gamma_*}^*\right]^{\otimes 2}\right\}$$

with $M^*(t)$, $V_{\alpha_*}^*, I_{\alpha_*}^*, S_{\alpha_*}^*, V_{\gamma_*}^*, I_{\gamma_*}^*$ and $S_{\gamma_*}^*$ being given in the Appendix.

The asymptotic covariance matrix can be consistently estimated by $\hat{A}_{\text{ar}}^{-1}\hat{\sum}_{\text{ar}}\hat{A}_{\text{ar}}^{-1}$, where

$$\hat{A}_{\text{ar}}=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}_{\text{ar}}(t)\}^{\otimes 2}Y_i(t)\text{dt},$$

$$\hat{\sum}_{\text{ar}}=\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\text{ar}}(t)\}d\hat{M}_i^*(t)-\hat{V}_{\alpha*,i}^*(\hat{I}_{\alpha*}^*)^{-1}\hat{S}_{\alpha*,i}^*-\hat{V}_{\gamma*,i}^*(\hat{I}_{\gamma*}^*)^{-1}\hat{S}_{\gamma*,i}^*\right]^{\otimes 2}$$

with

$$\hat{M}_i^*(t)=\frac{\xi_i}{\pi(W_i,\hat{\alpha})}N_i(t)+(1-\frac{\xi_i}{\pi(W_i,\hat{\alpha})})\rho(W_i,\hat{\gamma})N_i^*(t)$$
$$-\int_0^t Y_i(s)d\hat{\Lambda}_{\text{ar}}(s)-\int_0^t\hat{\beta}_{\text{ar}}^T Z_iY_i(s)\text{ds},$$

and $\hat{V}_{\alpha*}^*, \hat{I}_{\alpha*}^*, \hat{S}_{\alpha*,i}^*, \hat{V}_{\gamma*}^*, \hat{I}_{\gamma*}^*, \hat{S}_{\gamma*,i}^*$ being obtained by their empirical counterparts and replacing $(\alpha_*, \beta_0, \gamma_*)$ by $(\hat{\alpha}, \hat{\beta}_{ar}, \hat{\gamma})$ accordingly.

3. Model Checking

As suggested by a referee, it is useful to perform a lack-of-fit test to justify the use of the additive hazard model (1). In this section, we propose a model checking procedure based on the simple reweighted cumulative sums of martingale-based residuals (Lin et al., 1993). An approach based on the augmented reweighted cumulative sums could be derived similarly. Specifically, the simple reweighted cumulative sums is defined as

$$\mathscr{F}(t,z)=n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}I(Z_i\le z)\hat{M}_i(t),$$

where $I(Z_i \le z)$ means that each component of Z_i is no larger than the corresponding component of z. The null hypothesis here is that the additive hazard model (1) is correctly specified. Under the null hypothesis, we have shown in the Appendix that distribution of $\mathscr{F}(t, z)$ can be approximated by the following zero-mean Gaussian process

$$\overline{\mathscr{F}}(t,z)=n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\int_0^t\left[I(Z_i\le z)-\frac{\hat{f}_1(s,z,\hat{\alpha})}{S_{\text{sr}}^{(0)}(s,\hat{\alpha})}\right]d\hat{M}_i(s)$$
$$-n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\hat{f}_2^T(t,z,\hat{\alpha})\hat{A}_{\text{sr}}^{-1}\int_0^\tau\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\text{sr}}(t,\hat{\alpha})\}d\hat{M}_i(t)$$
$$+n^{-1/2}\sum_{i=1}^{n}\left[\hat{f}_3^T(t,z,\hat{\alpha})+\hat{f}_2^T(t,z,\hat{\alpha})\hat{A}_{\text{sr}}^{-1}\hat{V}_\alpha-\hat{f}_4^T(t,z,\hat{\alpha})\right]\hat{I}_\alpha^{-1}\hat{S}_{\alpha,i}, \quad (7)$$

where

$$\hat{f}_1(t,z,\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i,\hat{\alpha})}I(Z_i\le z)Y_i(t),$$

$$\hat{f}_2(t,z,\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\int_0^t I(Z_i\le z)Y_i(s)[Z_i-\overline{Z}_{\rm sr}(s,\hat{\alpha})]\mathrm{ds},$$

$$\hat{f}_3(t,z,\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\hat{\pi}^*(t)}{\pi^2(W_i,\hat{\alpha})}\dot{\pi}(W_i,\hat{\alpha})\int_0^t\frac{\hat{f}_1(s,z,\hat{\alpha})}{S_{\rm sr}^{(0)}(s,\hat{\alpha})}d\hat{M}_i(s)$$

and

$$\hat{f}_4(t,z,\hat{\alpha})=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\hat{\pi}^*(t)}{\pi^2(W_i,\hat{\alpha})}\dot{\pi}(W_i,\hat{\alpha})I(Z_i\le z)\hat{M}_i(t).$$

For $i = 1, \cdots, n$, define

$$\begin{aligned}\Phi_i(t,z)=&\frac{\xi_i\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\int_0^t\left[I(Z_i\le z)-\frac{\hat{f}_1(s,z,\hat{\alpha})}{S_{\rm sr}^{(0)}(s,\hat{\alpha})}\right]d\hat{M}_i(s)\\&-\frac{\xi_i\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\hat{f}_2^T(t,z,\hat{\alpha})\hat{A}_{\rm sr}^{-1}\int_0^\tau\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\rm sr}(t,\hat{\alpha})\}d\hat{M}_i(t)\\&+\left[\hat{f}_3^T(t,z,\hat{\alpha})+\hat{f}_2^T(t,z,\hat{\alpha})\hat{A}_{\rm sr}^{-1}\hat{V}_\alpha-\hat{f}_4^T(t,z,\hat{\alpha})\right]\hat{I}_\alpha^{-1}\hat{S}_{\alpha,i}.\end{aligned}$$

Then $\overline{\mathscr{F}}(t,z)=n^{-1/2}\sum_{i=1}^{n}\Phi_i(t,z)$. Apparently, it is difficult to obtain the analytical expression for the distribution of $\bar{\mathscr{F}}(t,z)$. In the following, we utilize the resampling method to obtain its asymptotic null distribution. In particular, we define

$$\hat{\mathscr{F}}(t,z)=n^{-1/2}\sum_{i=1}^{n}\Phi_i(t,z)G_i, \qquad (8)$$

where G_i, $i = 1, \cdots, n$ are i.i.d. standard normal random variables, which are independent of the data. It can be shown that the null distribution of $\mathscr{F}(t,z)$ can be approximated by the conditional distribution of $\hat{\mathscr{F}}(t,z)$ given the data. Therefore, the null distribution of $\mathscr{F}(t,z)$ can be approximated by a large number of realizations of $\hat{\mathscr{F}}(t,z)$ by repeatedly generating random numbers G_i, $i = 1, \ldots, n$, from standard normal distribution while fixing the observed data. To check the fit of model (1), one can use the supremum test $\sup_{t,z}|\mathscr{F}(t,z)|$ to obtain the p-value of the test, which can be obtained by comparing the observed value of $\sup_{t,z}|\mathscr{F}(t,z)|$ to a large number of realizations from $\sup_{t,z}|\hat{\mathscr{F}}(t,z)|$.

4. Simulation Studies

In this section, we report extensive simulation studies to investigate the performance of the proposed methods in Section 2 and compare them with the corresponding methods without reweighting. Let $Z = (Z_1, Z_2)^T$, where Z_1 and Z_2 are independent and identically distributed as the Bernoulli distribution with success probability 0.5. Given Z, the survival times were generated from distribution with the hazard function

$$\lambda(t|Z)=\lambda_0(t)+\beta_1 Z_1+\beta_2 Z_2, \quad (9)$$

where $\lambda_0(t) = 1$ and $(\beta_1, \beta_2) = (1, -1)$. The censoring time follows the uniform distribution on (0, 2.5), which produces approximately 48% censoring. The observation probability for δ was specified by the logistic regression model,

$$\pi(W,\alpha)=\frac{\exp(\alpha^T W)}{1+\exp(\alpha^T W)}, \quad (10)$$

where $W = (1, X, Z_1, Z_2)^T$. In our simulation, different α's were chosen to produce the desired missing rate. Furthermore, it can be shown that

$$\rho(W)=\frac{(2.5-X)(1+Z_1-Z_2)}{(2.5-X)(1+Z_1-Z_2)+1},$$

which is not a logistic regression model. However in our simulation, we still specify a logistic regression model $\rho(W, \gamma)$ for $\rho(W)$, because the true model is always unknown in practical problems and the logistic regression is most frequently used instead. The parameters in both of $\pi(W, \alpha)$ and $\rho(W, \gamma)$ are estimated by maximizing their likelihoods respectively. The sample size was set to be $n = 300$ or 500. All the simulation results were based on 5000 replications of independently simulated datasets. The following cases were considered:

Case 1: $\alpha = (0, -3, 1, 1)^T$, which produces approximately 65% missing rate;

Case 2: $\alpha = (1, -3, 1, 1)^T$, which produces approximately 50% missing rate;

Case 3: $\alpha = (1.6, -3, 1, 1)^T$, which produces approximately 35% missing rate.

Note that the second component of α, i.e. the coefficient for the observed survival time, is relatively large. This means that the missing mechanisms rely on the observed survival time heavily. Under these circumstances, the proportion of very small observation probabilities are high and some of the very small observation probabilities are very near zeros. The simulation results are reported in Tables 1 to 3. In these tables, bias is the sample mean of the estimator minus the true value; ESE denotes the average of the estimated standard errors; SSE is the sample standard error over the 5000 replications; CP represents the empirical coverage probability of 95% Wald-type confidence interval. These summary statistics are obtained for various methods including the full data analysis (denoted by Full), the complete case analysis (denoted by CC), the simple inverse probability weighted estimator (denoted by SW), the simple inverse probability weighted estimator with reweighting (denoted by SRW), the augmented inverse probability weighted estimator (denoted by AW), and the augmented inverse probability weighted estimator with reweighting (denoted by ARW).

As expected, from the three tables we can see that full data method produces the best results, while CC method obtains the worst including the largest biases and the smallest CP's. In addition, the SRW and ARW estimators improve the SW and AW estimators substantially, respectively. Note that the Full data method is only achievable in simulation when there is no missing censoring indicator. Full data results are presented here to provide an upper limit for the other methods.

For all the three cases, the CC method produces the biased estimates, which cannot be improved through increasing sample size. Although the ESEs agree with SSEs well, biases lead to the very low CP's. Tables 1 to 3 also show that when some observation probabilities are small, the inverse probability weighted estimators may be inflated and the inverse probability weighting methods do not perform well, although they perform better than the CC method. In the meanwhile, our proposed ARW estimators perform very well. The improvements of reweighting methods over methods without reweighting are substantial, especially for the cases with high missing rate. Moreover, it is easily observed that SSEs of the proposed ARW estimators are very close to those of full data analysis, implying that ARW method is very efficient. We further compared the relative efficiency of ARW relative to AW, Full relative to AW and summarized the results in Table 4. From Table 4, we can see that with the same sample size, as missing rate increases, the improvement in the efficiency increases. This could be due to the fact that when there is more missingness, there is more room for improvement. The amount of the efficiency gain can also depend on the value of the regression coefficient.

In order to further assess the performance of the proposed methods, we conducted some additional simulation studies for the setting with a continuous covariate and another distribution for censoring time. Define $Z = (Z_1, Z_2)^T$, where Z_1 follows the Bernoulli distribution with success probability 0.5 and Z_2, independent of Z_1, is a random variable following standard uniform distribution. Given Z, the survival times were generated from (9) with β_1, β_2 and λ_0 being the same as in the previous cases. The censoring time follows a distribution with the hazard function $\lambda(t) = 0.5t$, for $t > 0$, which leads to approximately 33% censoring. The observation probability was specified by model (10) with $W = (X, Z_1, Z_2, R)^T$, where R is the auxiliary variable distributed as the Bernoulli distribution with success probability 0.5. In this setting, we can obtain that

$$\rho(W)=\frac{1+Z_1-Z_2}{1+Z_1-Z_2+0.5X}$$

Here we still specify a logistic regression model $\rho(W, \gamma)$ for $\rho(W)$ as discussed previously. We chose $\alpha = (-5, 2, 2, 2)^T$ to produce approximately 56% missing rate. Denote this setting by Case 4. The simulation results based on 5000 replications are reported in Table 5, from which we can see that all the methods except ARW method do not perform well. The proposed ARW has negligible biases and the coverage rate is close to the nominal level.

From the above simulation studies and results in Tables 1 to 3 and 5, we can see that SW and SRW did not perform well. This is because the proportion of subjects with small observation probabilities are high and many of them are very close to zero. To further explore this, we conducted some additional simulations. In these additional simulations, all the settings are the same as those in Cases 1 to 3, except for the values of α in the observation probability model (10). Specifically, we consider the following cases:

Case 5: $\alpha = (-0.5, -1, 1, 1)^T$, which produces approximately 54.09% missingness;

Case 6: $\alpha = (-0.35, -1.2, 1, 1)^T$, which produces approximately 54.79% missingness;

Case 7: $\alpha = (-0.18, -1.5, 1, 1)^T$, which produces approximately 54.16% missingness;

Case 8: $\alpha = (0, -1.8, 1, 1)^T$, which produces approximately 54.64% missingness.

Case 9: $\alpha = (0.5, -2.5, 1, 1)^T$, which produces approximately 53.18% missingness.

In these cases, all the overall missing rates are approximately 54%. However the proportion of subjects with small observation probabilities become higher and higher from Cases 5 to 9. Specifically, the proportions of subjects with observation probabilities smaller than 0.1 are around 1.49%, 3.38%, 8.02%, 11.42% and 16.57% for Cases 5 to 9, respectively. The simulation results for sample size of 300 based on 5000 replications are listed in Table 6. From this table, we can see that the CP's of SW and SRW are reasonable for Cases 5 to 7. As the proportion of small observation probabilities increases, performances of SW, SRW and AW become worse. However, the ARW estimator behaves rather well in all the cases.

All the above simulation results demonstrate that the SRW and ARW methods could improve upon the SW and AW methods substantially, respectively, and the ARW methods perform very well for the survival data with missing censoring indicators under the additive hazard model under the cases we consider.

5. Real Data Analysis

In this section, we applied model (1) and the methods in Sections 2 and 3 to a dataset from a breast cancer clinical trial. This trial was conducted by the Eastern Cooperative Oncology Group (Cummings et al., 1986) with the aim to evaluate tamoxifen as a treatment for stage II breast cancer among elderly women. We are interested in the difference of the cause-specific hazard functions for death due to breast cancer of the placebo and treatment groups. This can be easily modeled by the additive hazard model (1).

There are a total of 169 elderly women participating in this trial, among whom 44 women died from breast cancer, 107 died from other known causes (n=17) or censored (n=90), and the cause of death was unknown for 18 patients. Let X denote the observed survival time of a patient, δ the indicator showing whether death was due to breast cancer, ξ the indicator of whether cause of death was known. The observed time for every subject was scaled by year. For every patient, we considered the following two variables: the treatment ($Z_1 = 0$, placebo group; $Z_1 = 1$, tamoxifen group) and whether the subject had 4 or more positive axillarylymph nodes ($Z_2 = 0$, no; $Z_2 = 1$, yes). Therefore the additive hazard model is

$$\lambda(t|Z)=\lambda_0(t)+\beta_1 Z_1+\beta_2 Z_2.$$

Let $W = (1, X, Z_1, Z_2, XZ_1, XZ_2, Z_1Z_2)^T$. We use the "glm" function in R to fit logistic regression for the observation probability with covariates W, then the "step" function to select the best model by the Akaike information criterion (AIC). The final chosen model only includes a single covariate X. Furthermore, the same W is used to fit logistic regression for the conditional probability of missing censoring indicator given the fully observed variables. We also fitted these two models by the probit regression, which is another popular alternative to model the binary outcome variable. We obtained almost the same modeling results. Hence we only used the logistic regression for further analysis. The results are presented in the first half of Table 7. The missing indicator rate is 11% in this example. With such low missing rate, our proposed reweighting methods could not exhibit their advantages over those without reweighting. For illustration purposes, we artificially deleted some censoring indicators among the observed ones in addition to the original missing. As discussed above, the observation probability is influenced only by one covariate X. Thus to introduce some additional missingness, the observation probability for δ was specified by $\pi(W) = \exp(20.5 - 2.8X)/(1 + \exp(20.5 - 2.8X))$. This model produces about 40% for missing censoring indicator.

The results with 40% missing rate are presented in the second half of Table 7. From the results, we can see that both of the reweighting methods and the methods without reweighting arrive at similar conclusions, i.e. treatment is not effective while the number of positive axillary lymph nodes is positively associated with breast-cancer survival. For the number of positive axillary lymph nodes, the reweighting methods produce similar point estimates, but smaller standard errors, and consequently smaller P-values than those without reweighting. This phenomenon is consistent with the simulation results presented and discussed in Section 4.

At last, we apply the approach developed in Section 3 to check the adequacy of the additive hazard model (1) for the original data. The supremum test $\sup_{t,z}|\mathcal{T}(t, z)| = 0.246$ and the P-value is 0.197 based on 1000 realizations of the $\sup_{t,z}|\hat{\mathcal{T}}(t, z)|$. Thus there is no evidence against the hazard model (1).

6. Discussion

In this paper, we investigated simple and augmented reweighting methods for survival data with missing censoring indicator under the additive hazard model. Generally speaking, the simple and augmented inverse probability weighted estimators without reweighting could become unstable when the probabilities of missingness for some subjects are very high, or, in other words, when the observation probabilities of some subjects are very small. The "reweighting" is motivated by the suggestion that it may enhance stability and ameliorate the variance inflation under such situation. Based on our experiences, the overall missing proportion in the entire cohort is not the key factor that affects the extent of improvement of reweighting methods over those without reweighting. The simple and augmented inverse probability weighted estimators without reweighting could behave comparably to the reweighting methods under the circumstances of high overall missing proportion. The most important factors are the proportion of subjects with very small observation probabilities and how small some of the observation probabilities are. When the observation probabilities for some subjects are very close to zero and the proportion of subjects with small observation probabilities is high, simple and augmented inverse probability weighted estimators without reweighting are very unstable. This point has been illustrated by Cases 5 to 9 in our simulation studies. From these simulation results, we can see that SW and AW estimators are rather unstable when the proportion of subjects with observation probabilities smaller than 0.1 is larger than 10%, i.e. Cases 8 and 9. In addition, we note that some of the observation probabilities are smaller than 0.012, and 0.004 under Cases 8 and 9, respectively. Under these circumstances, SW and AW have inflated variances and low coverage probabilities, while the proposed reweighting estimators can improve the estimators without reweighting significantly.

The proposed reweighting function is motivated from Xu et al. (2009), which considered the reweighting estimators for the Cox model with missing covariates. The proposed reweighting procedure could be extended to allow a general weighting function $\hat{\pi}^*(t)$. It is easy to see that the simple and augmented inverse probability weighted estimators without reweighting are special cases of the proposed reweighting estimators when the weight function $\hat{\pi}^* = 1$. As stated above, our reweighting estimator improves the estimator without reweighting when some of the observation probabilities are very close to zero and the proportion of very small observation probabilities are high. The improvement of the proposed reweighting methods over those without reweighting has been demonstrated in our simulation studies. There could be other choices for the reweighting function and it will be worthwhile to explore to find an optimal weight function, if it exists, in future studies.

Acknowledgments

The authors thank Editor-in-Chief, the Associate Editor, and two referees for their insightful comments and suggestions that greatly improved the article.

Appendix

Regularity conditions and sketch proofs of the main results

For proofs of the theorems, we list the following regularity conditions.

- **(C1)** $\Lambda_0(\tau) < \infty$ and $\Pr\{Y(\tau) = 1\} > 0$;
- **(C2)** Z is bounded with probability 1 and time-independent;
- **(C3)** The matrix A is positive definite.
- **(C4)** The observation probability $\pi(W, \alpha)$ is bounded away from 0; $\pi(W, \alpha)$ is twice continuously differential in α; There exists a compact neighborhood $\mathscr{A}$ of α_0 such that $E[\sup_{\alpha \in \mathscr{A}}\{\|\dot{\pi}(W, \alpha)\|^2 + \|\ddot{\pi}(W, \alpha)\|\}] < \infty$, where $\dot{\pi}(W, \alpha) = \partial\pi(W, \alpha)/\partial\alpha$ and $\ddot{\pi}(W, \alpha) = \partial^2\pi(W, \alpha)/\partial\alpha\partial\alpha^T$; There exists α_* satisfying the

equations $E(S^*_{\alpha_*})=0$, where $S^*_{\alpha}=\xi-\pi(W,\alpha)[\pi(W,\alpha)(1-\pi(W,\alpha))]^{-1}\dot{\pi}(W,\alpha)$.

(C5) $\rho(W,\gamma)$ is twice continuously differentiable in γ; There exists γ_* satisfying the equations $E(S^*_{\gamma_*})=0$, where $S^*_{\gamma}=\xi(\delta-\rho(W,\gamma))[\rho(W,\gamma)(1-\rho(W,\gamma))]^{-1}\dot{\rho}(W,\gamma)$.

All these conditions are standard for the derivation of asymptotic results in the survival analysis and parametric inference.

Proof of Theorem 1

By some simple algebraic calculations, it can be seen that

$$n^{\frac{1}{2}}(\hat{\beta}_{\text{sr}}-\beta_0)=\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\xi_i\frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{\text{sr}}(t,\hat{\alpha})\}^{\otimes 2}Y_i(t)\text{dt}\right]^{-1}\times\left[n^{-\frac{1}{2}}\sum_{j=1}^{n}\int_0^{\tau}\xi_j\frac{\hat{\pi}^*(t)}{\pi(W_j,\hat{\alpha})}\{Z_j-\overline{Z}_{\text{sr}}(t,\hat{\alpha})\}\text{dM}_j(t)\right].$$

Under conditions (C1), (C2) and (C4), it can be shown that

$$\sup_{t\in[0,\tau]}\|\hat{\pi}^*(t)-\hat{\pi}^*(t)\|=o_p(1). \quad \text{(A.1)}$$

$$\sup_{t\in[0,\tau]}\|S^{(k)}_{\text{sr}}(t,\hat{\alpha})-s^{(k)}(t)\|=o_p(1), k=0,1,2. \quad \text{(A.2)}$$

$$\sup_{t\in[0,\tau]}\|\overline{Z}_{\text{sr}}(t,\hat{\alpha})-\overline{z}(t)\|=o_p(1). \quad \text{(A.3)}$$

By (A.1), (A.3) and the fact that $\hat{\alpha}\xrightarrow{\text{P}}\alpha_0$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\xi_i\frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{\text{sr}}(t,\hat{\alpha})\}^{\otimes 2}Y_i(t)\text{dt}\xrightarrow{\text{P}}A.$$

It is easy to see that

$$n^{\frac{1}{2}}U_{\text{sr}}(\beta_0,\hat{\alpha})=n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\xi_i\frac{\hat{\pi}^*(t)}{\pi(W_i,\hat{\alpha})}\{Z_i-\overline{Z}_{\text{sr}}(t,\hat{\alpha})\}\text{dM}_i(t).$$

So we can conclude that

$$n^{\frac{1}{2}}(\hat{\beta}_{\rm sr}-\beta_0)=(A+o_p(1))^{-1}n^{\frac{1}{2}}U_{\rm sr}(\beta_0,\hat{\alpha}). \quad \text{(A.4)}$$

By the Taylor expansion of $n^{\frac{1}{2}}U_{\rm sr}(\beta_0,\hat{\alpha})$ at α_0,

$$n^{\frac{1}{2}}U_{\rm sr}(\beta_0,\hat{\alpha})=n^{\frac{1}{2}}U_{\rm sr}(\beta_0,\alpha_0)+\frac{\partial U_{\rm sr}(\beta_0,\alpha)}{\partial\alpha^T}\Big|_{\alpha=\alpha_0}n^{\frac{1}{2}}(\hat{\alpha}-\alpha_0)+o_p(1), \quad \text{(A.5)}$$

where

$$\begin{aligned}&-\frac{\partial U_{\rm sr}(\beta_0,\alpha)}{\partial\alpha^T}\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\pi^2(W_i,\alpha)}\int_0^\tau\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\rm sr}(t,\alpha)\}{\rm dM}_i(t)\dot{\pi}^T(W_i,\alpha)\\&+\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i,\alpha)}\int_0^\tau\hat{\pi}^*(t)\left\{\frac{S_{\rm sr}^{(1)}(t,\alpha)}{\left(S_{\rm sr}^{(0)}(t,\alpha)\right)^2}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{\xi_j}{\pi^2(W_j,\alpha)}\dot{\pi}^T(W_j,\alpha)Y_j(t)\right]\right.\\&\left.-\frac{1}{S_{\rm sr}^{(0)}(t,\alpha)}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{\xi_j}{\pi^2(W_j,\alpha)}Y_j(t)Z_j\dot{\pi}^T(W_j,\alpha)\right]\right\}{\rm dM}_i(t).\end{aligned}$$

By (A.1) to (A.3) and the law of large numbers, it can be proven that

$$-\frac{\partial U_{\rm sr}(\beta_0,\alpha)}{\partial\alpha^T}\Big|_{\alpha=\alpha_0}\xrightarrow{\rm P}V_{\alpha_0}, \quad \text{(A.6)}$$

where

$$\begin{aligned}V_{\alpha_0}=&E\left[\int_0^\tau\{Z-\overline{z}(t)\}\frac{1}{\pi(W,\alpha_0)}\dot{\pi}^T(W,\alpha_0)\pi^*(t){\rm dM}(t)\right]\\&-E\left[\int_0^\tau\left\{\frac{h^{(2)}(t)}{s^{(0)}(t)}-\frac{s^{(1)}(t)h^{(1)}(t)}{\left(s^{(0)}(t)\right)^2}\right\}\pi^*(t){\rm dM}(t)\right],\end{aligned}$$

$$h^{(1)}(t)={\rm E}\left[\frac{1}{\pi(W,\alpha_0)}\dot{\pi}^T(W,\alpha_0)Y(t)\right]$$

and

$$h^{(2)}(t)={\rm E}\left[\frac{1}{\pi(W,\alpha_0)}Y(t)Z\dot{\pi}^T(W,\alpha_0)\right].$$

By (A.1), (A.3) and Lemma A.1 of Qi et al. (2005), we can obtain that

$$n^{\frac{1}{2}}U_{\mathrm{sr}}(\beta_0,\alpha_0)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i,\alpha_0)}\int_0^{\tau}\pi^*(t)\{Z_i-\overline{z}(t)\}\mathrm{dM}_i(t)+o_p(1). \quad (A.7)$$

Define

$$S_{\alpha_0}=\frac{\xi-\pi(W,\alpha_0)}{\pi(W,\alpha_0)\{1-\pi(W,\alpha_0)\}}\dot{\pi}(W,\alpha_0)$$

and

$$I_{\alpha_0}=E\left[S_{\alpha_0}S_{\alpha_0}^{T}-\frac{\xi-\pi(W,\alpha_0)}{\pi(W,\alpha_0)\{1-\pi(W,\alpha_0)\}}\ddot{\pi}(W,\alpha_0)\right],$$

which are score and information matrices of $\pi(W,\alpha)$ respectively. Then under condition (C4), it can be shown that

$$n^{\frac{1}{2}}(\hat{\alpha}-\alpha_0)=n^{-\frac{1}{2}}\sum_{i=1}^{n}I_{\alpha_0}^{-1}S_{\alpha_0,i}+o_p(1), \quad (A.8)$$

where $S_{\alpha,i}$ is obtained through replacing ξ and W by ξ_i and W_i in S_α respectively.

By (A.4) to (A.8), we finally arrive at

$$\begin{aligned} &n^{\frac{1}{2}}(\hat{\beta}_{\mathrm{sr}}-\beta_0)\\ &=A^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[\frac{\xi_i}{\pi(W_i,\alpha_0)}\int_0^{\tau}\pi^*(t)\{Z_i-\overline{z}(t)\}\mathrm{dM}_i(t)-V_{\alpha_0}I_{\alpha_0}^{-1}S_{\alpha_0,i}\right]+o_p(1). \end{aligned}$$

By the central limit theorem, the desired result is proved.

Proof of Theorem 2

It is easily verified that

$$
\begin{aligned}
&n^{\frac{1}{2}}(\hat{\beta}_{\mathrm{ar}}-\beta_0)\\
&=\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\mathrm{ar}}(t)\}^{\otimes 2}Y_i(t)\mathrm{dt}\right]^{-1}\\
&\times\left[n^{-\frac{1}{2}}\sum_{j=1}^{n}\int_0^{\tau}\{Z_j-\overline{Z}_{\mathrm{ar}}(t)\}\hat{\pi}^*(t)\left\{\frac{\xi_j}{\pi(W_j,\hat{\alpha})}\mathrm{dN}_j(t)\right.\right.\\
&\quad\left.\left.+(1-\frac{\xi_j}{\pi(W_j,\hat{\alpha})})\rho(W_j,\hat{\gamma})\mathrm{dN}_j^*(t)-Y_j(t)\beta_0^TZ_j\mathrm{dt}-Y_j(t)d\Lambda_0(t)\right\}\right]\\
&=\left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\mathrm{ar}}(t)\}^{\otimes 2}Y_i(t)\mathrm{dt}\right]^{-1}\left[n^{-\frac{1}{2}}\sum_{j=1}^{n}\int_0^{\tau}\{Z_j-\overline{Z}_{\mathrm{ar}}(t)\}\hat{\pi}^*(t)d\hat{M}_j^*(t)\right],
\end{aligned}
\quad (A.9)
$$

where

$$
\begin{aligned}
&d\hat{M}_j^*(t)\\
&=\frac{\xi_j}{\pi(W_j,\hat{\alpha})}\mathrm{dN}_j(t)+(1-\frac{\xi_j}{\pi(W_j,\hat{\alpha})})\rho(W_j,\hat{\gamma})\mathrm{dN}_j^*(t)-Y_j(t)\beta_0^TZ_j\mathrm{dt}-Y_j(t)d\Lambda_0(t).
\end{aligned}
$$

By the fact that

$$\sup_{t\in[0,\tau]}\|\overline{Z}_{\mathrm{ar}}(t)-\overline{z}(t)\|=o_p(1) \quad (A.10)$$

and (A.1), we can conclude that

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\hat{\pi}^*(t)\{Z_i-\overline{Z}_{\mathrm{ar}}(t)\}^{\otimes 2}Y_i(t)\mathrm{dt}\xrightarrow{\mathrm{P}}A. \quad (A.11)$$

By (A.11), we have

$$n^{\frac{1}{2}}(\hat{\beta}_{\mathrm{ar}}-\beta_0)=(A+o_p(1))^{-1}n^{\frac{1}{2}}U_{\mathrm{ar}}(\beta_0,\hat{\alpha},\hat{\gamma}). \quad (A.12)$$

By the Taylor expansion of $n^{\frac{1}{2}}U_{\mathrm{ar}}(\beta_0,\hat{\alpha},\hat{\gamma})$ at α_* and γ_*,

$$n^{\frac{1}{2}}U_{\mathrm{ar}}(\beta_0,\hat{\alpha},\hat{\gamma})$$
$$=n^{\frac{1}{2}}U_{\mathrm{ar}}(\beta_0,\alpha_*,\gamma_*)+\frac{\partial U_{\mathrm{ar}}(\beta_0,\alpha,\gamma)}{\partial\alpha^T}\Big|_{\alpha=\alpha_*,\gamma=\gamma_*}n^{\frac{1}{2}}(\hat{\alpha}-\alpha_*)$$
$$+\frac{\partial U_{\mathrm{ar}}(\beta_0,\alpha,\gamma)}{\partial\gamma^T}\Big|_{\alpha=\alpha_*,\gamma=\gamma_*}n^{\frac{1}{2}}(\hat{\gamma}-\gamma_*)+o_p(1), \quad (A.13)$$

where

$$-\frac{\partial U_{\mathrm{ar}}(\beta_0,\alpha,\gamma)}{\partial\alpha^T}$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}_{\mathrm{ar}}\}\hat{\pi}^*(t)\frac{\xi_i}{\pi^2(W_i,\alpha)}\dot{\pi}^T(W_i,\alpha)\{\mathrm{dN}_i(t)-\rho(W_i,\gamma)\mathrm{dN}_i^*(t)\}$$

and

$$\frac{\partial U_{\mathrm{ar}}(\beta_0,\alpha,\gamma)}{\partial\gamma^T}=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\{Z_i-\overline{Z}_{\mathrm{ar}}(t)\}\hat{\pi}^*(t)(1-\frac{\xi_i}{\pi(W_i,\alpha)})\dot{\rho}^T(W_i,\gamma)\mathrm{dN}_i^*(t).$$

By (A.1), $\hat{\alpha}\xrightarrow{\mathrm{P}}\alpha_*$ and $\hat{\gamma}\xrightarrow{\mathrm{P}}\gamma_*$, we have

$$-\frac{\partial U_{\mathrm{ar}}(\beta_0,\alpha,\gamma)}{\partial\alpha^T}\Big|_{\alpha=\alpha_*,\gamma=\gamma_*}\xrightarrow{\mathrm{P}}V_{\alpha_*}^* \quad (A.14)$$

and

$$-\frac{\partial U_{\mathrm{ar}}(\beta_0,\alpha,\gamma)}{\partial\gamma^T}\Big|_{\alpha=\alpha_*,\gamma=\gamma_*}\xrightarrow{\mathrm{P}}V_{\gamma_*}^*, \quad (A.15)$$

where

$$V_{\alpha_*}^*=E\left[\int_0^\tau\{Z-\overline{z}(t)\}\pi^*(t)\frac{\dot{\pi}^T(W,\alpha_*)}{\pi(W,\alpha_*)}\{\mathrm{dN}(t)-\rho(W,\gamma_*)\mathrm{dN}^*(t)\}\right]$$

and

$$V_{\gamma_*}^*=E\left[\int_0^\tau\{Z-\overline{z}(t)\}\pi^*(t)\frac{\xi-\pi(W,\alpha_*)}{\pi(W,\alpha_*)}\dot{\rho}^T(W,\gamma_*)\mathrm{dN}^*(t)\right].$$

Similar to (A.7), we have

$$n^{\frac{1}{2}}U_{\text{ar}}(\beta_0,\alpha_*,\gamma_*)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\pi^*(t)\{Z_i-\overline{z}(t)\}\text{dM}_i^*(t)+o_p(1), \quad \text{(A.16)}$$

where

$$\text{dM}_i^*(t) = \frac{\xi_i}{\pi(W_i,\alpha_*)}\text{dN}_i(t)+(1-\frac{\xi_i}{\pi(W_i,\alpha_*)})\rho(W_i,\gamma_*)\text{dN}_i^*(t)-Y_i(t)\beta_0^T Z_i\text{dt}-Y_i(t)d\Lambda_0(t).$$

Define

$$S_{\alpha_*}^*=\frac{\xi-\pi(W,\alpha_*)}{\pi(W,\alpha_*)(1-\pi(W,\alpha_*))}\dot{\pi}(W,\alpha_*),$$

$$S_{\gamma_*}^*=\frac{\xi(\delta-\rho(W,\gamma_*))}{\rho(W,\gamma_*)(1-\rho(W,\gamma_*))}\dot{\rho}(W,\gamma_*),$$

$$I_{\alpha_*}^*=E\left[S_{\alpha_*}^*S_{\alpha_*}^{*T}-\frac{\xi-\pi(W,\alpha_*)}{\pi(W,\alpha_*)(1-\pi(W,\alpha_*))}\ddot{\pi}(W,\alpha_*)\right],$$

and

$$I_{\gamma_*}^*=E\left[S_{\gamma_*}^*S_{\gamma_*}^{*T}-\frac{\xi(\delta-\rho(W,\gamma_*))}{\rho(W,\gamma_*)(1-\rho(W,\gamma_*))}\ddot{\rho}(W,\gamma_*)\right].$$

Then under Condition (C4) and (C5), we have

$$n^{\frac{1}{2}}(\hat{\alpha}-\alpha_*)=n^{-\frac{1}{2}}\sum_{i=1}^{n}(I_{\alpha_*}^*)^{-1}S_{\alpha_*,i}^*+o_p(1) \quad \text{(A.17)}$$

and

$$n^{\frac{1}{2}}(\hat{\gamma}-\gamma_*)=n^{-\frac{1}{2}}\sum_{i=1}^{n}(I_{\gamma_*}^*)^{-1}S_{\gamma_*,i}^*+o_p(1). \quad \text{(A.18)}$$

where $S_{\alpha_*,i}^*$ and $S_{\gamma_*,i}^*$ is obtained through replacing ξ, δ and W by ξ_i, δ_i and W_i in $S_{\alpha_*}^*$ and $S_{\gamma_*}^*$. By (A.9) to (A.18), we can finally conclude that

$$
\begin{aligned}
& n^{\frac{1}{2}}(\hat{\beta}_{\text{ar}} - \beta_0) \\
=& A^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \left[\int_0^\tau \pi^*(t)\{Z_i - \overline{z}(t)\} \mathrm{dM}_i^*(t) - V_{\alpha_*}^*(I_{\alpha_*}^*)^{-1} S_{\alpha_*,i}^* - V_{\gamma_*}^*(I_{\gamma_*}^*)^{-1} S_{\gamma_*,i}^* \right] + o_p(1).
\end{aligned}
$$

By the central limit theorem, the desired result is proved.

Proof of (7)

It is easy to see that

$$
\begin{aligned}
& \mathscr{F}(t,z) = n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i \hat{\pi}^*(t)}{\pi(W_i, \hat{\alpha})} I(Z_i \le z) M_i(t) \\
& - n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i \hat{\pi}^*(t)}{\pi(W_i, \hat{\alpha})} I(Z_i \le z) \int_0^t Y_i(s) d\{\hat{\Lambda}_{\text{sr}}(s) - \Lambda_0(s)\} \\
& - n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i \hat{\pi}^*(t)}{\pi(W_i, \hat{\alpha})} I(Z_i \le z) \int_0^t Y_i(s) \mathrm{ds} \mathrm{Z}_i^T (\hat{\beta}_{\text{sr}} - \beta_0) \\
& \doteq I - \mathrm{II} - \mathrm{III}. \qquad \text{(A.19)}
\end{aligned}
$$

By (A.1) and Taylor expansion of I at α_0, we obtain

$$
\begin{aligned}
I =& n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i \pi^*(t)}{\pi(W_i, \alpha_0)} I(Z_i \le z) M_i(t) \\
& - \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\xi_i \pi^*(t)}{\pi^2(W_i, \alpha_0)} \dot{\pi}^T(W_i, \alpha_0) I(Z_i \le z) M_i(t) \right] n^{-1/2} \sum_{i=1}^{n} I_{\alpha_0}^{-1} S_{\alpha_0,i} + o_p(1) \\
=& n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i \pi^*(t)}{\pi(W_i, \alpha_0)} I(Z_i \le z) M_i(t) - f_4^T(t,z,\alpha_0) n^{-1/2} \sum_{i=1}^{n} I_{\alpha_0}^{-1} S_{\alpha_0,i} + o_p(1), \qquad \text{(A.20)}
\end{aligned}
$$

where $f_4(t, z, \alpha_0)$ is the limit of $\hat{f}_4(t, z, \hat{\alpha})$.

Similar to the proof of Theorem 2.4 in Lin (2011), by (4), it can be proven that

$$
\begin{aligned}
& n^{-1/2}\{\hat{\Lambda}_{\text{sr}}(t) - \Lambda_0(t)\} \\
=& n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i, \alpha_0)} \int_0^t \frac{1}{s^{(0)}(t)} \mathrm{dM}_i(s) \\
& - \frac{1}{n} \sum_{i=1}^{n} \frac{\xi_i}{\pi^2(W_i, \alpha_0)} \dot{\pi}^T(W_i, \alpha_0) \int_0^t \frac{1}{s^{(0)}(t)} \mathrm{dM}_i(s) \\
& \times n^{-1/2} \sum_{i=1}^{n} I_{\alpha_0}^{-1} S_{\alpha_0,i} + o_p(1). \qquad \text{(A.21)}
\end{aligned}
$$

By (A.21) and Taylor expansion, we have

$$\mathrm{II}=n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i\pi^*(t)}{\pi(W_i,\alpha_0)}\int_0^t\frac{f_1(s,z,\alpha_0)}{s^{(0)}(t)}\mathrm{dM}_i(s)$$
$$-f_3^T(t,z,\alpha_0)n^{-1/2}\sum_{i=1}^{n}I_\alpha^{-1}S_{\alpha,i}$$
$$-\pi^*(t)\int_0^t f_1(s,z,\alpha_0)\overline{z}^T(s)\mathrm{ds}n^{-1/2}(\hat{\beta}-\beta_0)+o_p(1). \quad (\mathrm{A}.22)$$

where $f_1(t, z, \alpha_0)$ and $f_3(t, z, \alpha_0)$ are the limits of $\hat{f}_1(t, z, \hat{\alpha})$ and $\hat{f}_3(t, z, \hat{\alpha})$ respectively.It is easy to see that

$$\mathrm{III}=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\pi^*(t)}{\pi(W_i,\alpha_0)}I(Z_i\le z)Z_i^T\int_0^t Y_i(s)\mathrm{ds}\right]n^{-1/2}(\hat{\beta}-\beta_0)+o_p(1). \quad (\mathrm{A}.23)$$

From (A.19) to (A.23), we finally arrive at

$$\mathscr{F}(t,z)=n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i\pi^*(t)}{\pi(W_i,\alpha_0)}\int_0^t\left[I(Z_i\le z)-\frac{f_1(s,z,\alpha_0)}{S_{\mathrm{sr}}^{(0)}(s,\alpha_0)}\right]\mathrm{dM}_i(s)$$
$$-n^{-1/2}\sum_{i=1}^{n}\frac{\xi_i\pi^*(t)}{\pi(W_i,\alpha_0)}f_2^T(t,z,\alpha_0)A^{-1}\int_0^\tau\pi^*(t)\{Z_i-\overline{z}_{\mathrm{sr}}(t)\}\mathrm{dM}_i(t)$$
$$+n^{-1/2}\sum_{i=1}^{n}\left[f_3^T(t,z,\alpha_0)+f_2^T(t,z,\alpha_0)A^{-1}V_\alpha-f_4^T(t,z,\alpha_0)\right]I_{\alpha_0}^{-1}S_{\alpha_0,i}+o_p(1), \quad (\mathrm{A}.24)$$

where $f_2(t, z, \alpha_0)$ is the limit of $\hat{f}_2(t, z, \hat{\alpha})$. The finite dimensional convergence of $\mathscr{F}(t, z)$ can be proven by the multivariate cental limit theorem. By the techniques in Lin (2001), it can be proven that $\mathscr{F}(t, z)$ is tight. So $\mathscr{F}(t, z)$ converges weakly to a zero-mean Gaussian process which can be approximately by (7).

References

Breslow, N., Day, N. Statistical Models in Cancer Research, 1, The Design and Analysis of Case-Control Studies. IARC; Lyon: 1980.

Dinse G. Nonparametric estimation for partially-complete time and type of failure data. Biometrics. 1982; 38(2):417–431. [PubMed: 6896831]

Goetghebeur E, Ryan L. Analysis of competing risks survival data when some failure types are missing. Biometrika. 1995; 82(4):821–833.

Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc. 1952; 47(260):663–685.

Hyun S, Lee J, Sun Y. Proportional hazards model for competing risks data with missing cause of failure. J Stat Plan Infer. 2012; 142(7):1767–1779.

Li X, Wang Q. The weighted least square based estimators with censoring indicators missing at random. J Stat Plan Infer. 2012; 142(11):2913–2925.

Lin D, Wei L, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika. 1993; 80(3):557–572.

Lin D, Ying Z. Semiparametric analysis of the additive risk model. Biometrika. 1994; 81(1):61–71.

Liu C, Wang Q. Semiparametric estimation for regression coefficients in the Cox model with failure indicators missing at random. Stat Sinica. 2010; 20(3):1125–1142.

Lu K, Tsiatis A. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. Biometrics. 2001; 57(4):1191–1197. [PubMed: 11764260]

Lu W, Liang Y. Analysis of competing risks data with missing cause of failure under additive hazards model. Stat Sinica. 2008; 18(1):219–234.

McKeague I, Subramanian S. Product-limit estimators and Cox regression with missing censoring information. Scand J Stat. 1998; 25(4):589–601.

Qi L, Wang C, Prentice R. Weighted estimators for proportional hazards regression with missing covariates. J Am Stat Assoc. 2005; 100(472):1250–1263.

Robins J, Rotnitzky A, Zhao L. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc. 1994; 89(427):846–866.

Song X, Sun L, Mu X, Dinse G. Additive hazards regression with censoring indicators missing at random. Can J Stat. 2010; 38(3):333–351. [PubMed: 21197117]

Sun Y, Wang H, Gilbert P. Quantile regression for competing risks data with missing cause of failure. Stat Sinica. 2012; 22(2):703–728.

Sun Z, Xie T, Liang H. Statistical inference for right-censored data with nonignorable missing censoring indicators. Sci China-Math. 2013; 56(6):1263–1278.

Wang Q, Dinse G. Linear regression analysis of survival data with missing censoring indicators. Lifetime Data Anal. 2011; 17(2):256–279. [PubMed: 20559722]

White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982; 50(1):1–25.

Xu Q, Paik M, Luo X, Tsai W. Reweighting estimators for Cox regression with missing covariates. J Am Stat Assoc. 2009; 104(487):1155–1167.

Tsiatis, A. Semiparametric theory and missing data. Springer; New York: 2006.

Zhou X, Sun L. Additive hazards regression with missing censoring information. Stat Sinica. 2003; 13(4):1237–1257.

Table 1

Simulation results for Case 1.

n	**Method**	β_1				β_2			
		Bias	**ESE**	**SSE**	**CP**	**Bias**	**ESE**	**SSE**	**CP**
300	Full	0.005	0.147	0.148	0.95	−0.004	0.146	0.147	0.95
	CC	0.435	0.485	0.479	0.85	−0.893	0.567	0.573	0.68
	SW	0.067	0.443	0.570	0.84	−0.394	0.569	0.631	0.82
	SRW	0.053	0.421	0.532	0.87	−0.304	0.469	0.544	0.87
	AW	0.009	0.224	0.256	0.93	0.000	0.276	0.427	0.89
	ARW	0.008	0.201	0.201	0.95	0.000	0.217	0.280	0.95
500	Full	0.002	0.113	0.113	0.95	−0.002	0.113	0.113	0.95
	CC	0.440	0.366	0.354	0.77	−0.883	0.435	0.444	0.48
	SW	0.046	0.348	0.451	0.84	−0.317	0.455	0.512	0.83
	SRW	0.026	0.326	0.401	0.88	−0.222	0.364	0.427	0.87
	AW	0.003	0.180	0.200	0.93	−0.007	0.219	0.297	0.89
	ARW	0.002	0.153	0.151	0.96	0.000	0.160	0.161	0.95

Table 2

Simulation results for Case 2.

n	Method	β_1				β_2			
		Bias	**ESE**	**SSE**	**CP**	**Bias**	**ESE**	**SSE**	**CP**
300	Full	0.005	0.147	0.148	0.95	−0.004	0.146	0.147	0.95
	CC	0.240	0.338	0.328	0.91	−0.730	0.396	0.395	0.56
	SW	0.045	0.317	0.396	0.86	−0.266	0.416	0.447	0.85
	SRW	0.028	0.301	0.362	0.90	−0.195	0.338	0.382	0.88
	AW	0.070	0.186	0.199	0.94	−0.002	0.225	0.361	0.91
	ARW	0.006	0.177	0.177	0.95	0.000	0.187	0.236	0.95
500	Full	0.002	0.113	0.113	0.95	−0.002	0.113	0.113	0.95
	CC	0.236	0.257	0.248	0.86	−0.715	0.304	0.301	0.34
	SW	0.032	0.255	0.320	0.86	−0.199	0.332	0.378	0.84
	SRW	0.017	0.239	0.282	0.90	−0.130	0.265	0.311	0.88
	AW	0.001	0.146	0.156	0.94	−0.006	0.180	0.228	0.91
	ARW	0.001	0.136	0.135	0.96	0.000	0.141	0.144	0.95

Table 3

Simulation results for Case 3.

n	Method	β_1				β_2			
		Bias	ESE	SSE	CP	Bias	ESE	SSE	CP
300	Full	0.005	0.147	0.148	0.95	−0.004	0.146	0.147	0.95
	CC	0.154	0.283	0.275	0.93	−0.620	0.331	0.327	0.55
	SW	0.035	0.269	0.330	0.87	−0.209	0.353	0.372	0.86
	SRW	0.021	0.256	0.302	0.91	−0.150	0.288	0.320	0.89
	AW	0.008	0.171	0.181	0.94	−0.008	0.201	0.285	0.91
	ARW	0.005	0.168	0.168	0.95	−0.002	0.174	0.184	0.95
500	Full	0.002	0.113	0.113	0.95	−0.002	0.113	0.115	0.95
	CC	0.154	0.216	0.210	0.90	−0.616	0.255	0.253	0.32
	SW	0.024	0.218	0.270	0.88	−0.155	0.282	0.313	0.87
	SRW	0.010	0.205	0.239	0.91	−0.103	0.227	0.260	0.90
	AW	0.004	0.133	0.137	0.94	−0.007	0.160	0.194	0.92
	ARW	0.000	0.129	0.129	0.95	−0.003	0.133	0.136	0.95

Table 4

Summary of the relative efficiency for Cases 1 to 3

Missing Rate	n	$RE_{ARW:AW}$		$RE_{Full:AW}$	
		β_1	β_2	β_1	β_2
60%	300	1.62	2.33	2.99	8.44
	500	1.75	3.40	3.13	6.90
50%	300	1.26	2.34	1.81	6.03
	500	1.34	2.51	1.90	4.07
35%	300	1.16	2.40	1.50	3.76
	500	1.12	2.03	1.47	2.85

Table 5

Simulation results for Case 4.

n	Method	β_1				β_2			
		Bias	**ESE**	**SSE**	**CP**	**Bias**	**ESE**	**SSE**	**CP**
300	Full	0.007	0.143	0.145	0.95	−0.001	0.205	0.207	0.95
	CC	−0.446	0.510	0.526	0.89	−1.122	0.827	0.831	0.73
	SW	−0.250	0.493	0.712	0.78	−0.583	0.983	1.140	0.85
	SRW	−0.232	0.442	0.625	0.81	−0.456	0.818	1.002	0.86
	AW	−0.002	0.248	0.500	0.94	0.071	0.455	1.875	0.82
	ARW	0.005	0.183	0.193	0.95	−0.005	0.324	0.411	0.95
500	Full	0.004	0.111	0.110	0.95	−0.002	0.159	0.158	0.95
	CC	−0.425	0.389	0.398	0.84	−1.111	0.635	0.639	0.58
	SW	−0.201	0.407	0.592	0.79	−0.513	0.804	0.966	0.83
	SRW	−0.190	0.359	0.502	0.81	−0.371	0.653	0.825	0.85
	AW	0.003	0.224	0.521	0.94	0.063	0.418	1.240	0.82
	ARW	0.005	0.142	0.155	0.95	−0.006	0.254	0.337	0.95

Table 6

Simulation results for Cases 5 to 9 with sample size $n = 300$.

Case	Method	β_1				β_2			
		Bias	**ESE**	**SSE**	**CP**	**Bias**	**ESE**	**SSE**	**CP**
Case 5	Full	0.004	0.147	0.148	0.950	−0.005	0.146	0.146	0.950
	CC	0.129	0.240	0.242	0.930	−0.336	0.336	0.338	0.870
	SW	0.016	0.242	0.267	0.930	−0.094	0.304	0.318	0.940
	SRW	0.018	0.244	0.264	0.930	−0.083	0.286	0.299	0.950
	AW	0.007	0.167	0.172	0.950	−0.005	0.193	0.200	0.930
	ARW	0.006	0.170	0.171	0.950	−0.002	0.186	0.186	0.950
Case 6	Full	0.009	0.147	0.150	0.945	−0.004	0.146	0.147	0.947
	CC	0.155	0.252	0.249	0.920	−0.395	0.347	0.354	0.837
	SW	0.020	0.249	0.274	0.920	−0.121	0.319	0.338	0.936
	SRW	0.020	0.251	0.271	0.930	−0.101	0.294	0.313	0.940
	AW	0.011	0.168	0.174	0.937	−0.003	0.194	0.205	0.919
	ARW	0.009	0.170	0.173	0.948	0.000	0.184	0.187	0.942
Case 7	Full	0.005	0.146	0.147	0.952	−0.004	0.145	0.146	0.951
	CC	0.189	0.274	0.272	0.912	−0.477	0.366	0.375	0.780
	SW	0.022	0.265	0.300	0.917	−0.151	0.349	0.372	0.923
	SRW	0.021	0.264	0.291	0.924	−0.122	0.309	0.337	0.931
	AW	0.006	0.170	0.173	0.944	−0.007	0.201	0.217	0.917
	ARW	0.006	0.172	0.171	0.951	−0.002	0.185	0.187	0.943
Case 8	Full	0.005	0.146	0.147	0.952	−0.004	0.145	0.146	0.951
	CC	0.219	0.295	0.292	0.902	−0.555	0.384	0.389	0.728
	SW	0.028	0.280	0.326	0.899	−0.188	0.378	0.396	0.910
	SRW	0.024	0.276	0.312	0.917	−0.147	0.324	0.353	0.924
	AW	0.007	0.173	0.178	0.943	−0.007	0.207	0.235	0.912
	ARW	0.006	0.173	0.173	0.954	−0.002	0.186	0.189	0.947
Case 9	Full	−0.005	0.146	0.147	0.952	−0.004	0.145	0.146	0.951
	CC	0.248	0.332	0.322	0.893	−0.685	0.405	0.404	0.624
	SW	0.038	0.314	0.375	0.877	−0.248	0.425	0.434	0.877

Case	Method	β_1				β_2			
		Bias	ESE	SSE	CP	Bias	ESE	SSE	CP
	SRW	0.027	0.300	0.349	0.901	−0.186	0.346	0.376	0.896
	AW	0.007	0.182	0.192	0.937	−0.007	0.223	0.342	0.904
	ARW	−0.006	0.177	0.176	0.952	−0.002	0.188	0.208	0.949

Table 7

Analysis results for the breast cancer study.

Missing Rate	Method	Treatment			#Nodes		
		EST($\times 10^2$)	ESE($\times 10^2$)	*P*-value	EST($\times 10^2$)	ESE($\times 10^2$)	*P*-value
11%	CC	−0.845	1.452	0.561	3.442	1.528	0.024
	SW	−0.921	1.520	0.545	3.708	1.598	0.020
	SRW	−0.887	1.524	0.561	3.697	1.602	0.021
	AW	−0.466	1.470	0.751	3.020	1.527	0.048
	ARW	−0.438	1.476	0.767	3.018	1.532	0.049
40%	CC	−1.845	2.700	0.494	6.331	2.842	0.025
	SW	−0.639	2.159	0.767	5.329	2.302	0.021
	SRW	−1.110	2.006	0.580	4.891	2.120	0.021
	AW	−0.091	1.692	0.956	3.258	1.695	0.054
	ARW	−0.258	1.562	0.868	3.134	1.594	0.049

Note: EST is the estimator of regression coefficient; ESE is the estimated standard error.