



Published in final edited form as:

Can J Stat. 2010 September ; 38(3): 333–351. doi:10.1002/cjs.10072.

Additive hazards regression with censoring indicators missing at random

Xinyuan SONG,

Department of Statistics, Shatin, N. T., Hong Kong, P. R. China

Liuquan SUN,

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China

Xiaoyun MU, and

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China

Gregg E. DINSE

Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

Xinyuan SONG: xysong@sta.cuhk.edu.hk; Liuquan SUN: slq@amt.ac.cn; Xiaoyun MU: muxy@amss.ac.cn; Gregg E. DINSE: dinse@niehs.nih.gov

Abstract

In this article, the authors consider a semiparametric additive hazards regression model for right-censored data that allows some censoring indicators to be missing at random. They develop a class of estimating equations and use an inverse probability weighted approach to estimate the regression parameters. Nonparametric smoothing techniques are employed to estimate the probability of non-missingness and the conditional probability of an uncensored observation. The asymptotic properties of the resulting estimators are derived. Simulation studies show that the proposed estimators perform well. They motivate and illustrate their methods with data from a brain cancer clinical trial.

Keywords

Additive hazards model; censoring; kernel smoother; missing at random; weighted estimating equation

1. INTRODUCTION

In the analysis of failure time data, the cause of failure may be unknown for some subjects for a variety of reasons (e.g., autopsies were not performed or medical records were missing). We motivate and illustrate our methods with data on patients from a brain cancer clinical trial, where we evaluate the effect of two potential explanatory variables on a measure of quality of life. All patients were initially ambulatory, but over time some lost their mobility, some had a progression of their cancer, and some experienced both events. To assess quality of life, we define “survival time” as the time to non-ambulatory progression. Thus, patients who progressed and were no longer ambulatory contributed

uncensored times, patients who progressed but were still ambulatory or who had not progressed by the end of the study contributed censored times, and patients who progressed but whose ambulatory status was unknown contributed times with missing censoring indicators. We apply our regression analysis to evaluate the effects of sex and age on the time to non-ambulatory progression.

Specifically, let T be the failure time, let Z be a $p \times 1$ vector of covariates, and let C be a censoring time that is assumed to be conditionally independent of T given Z . Data are available on Z and $X = T \wedge C$, but the censoring indicator $\delta = I(T \leq C)$ may be missing. If the probability that δ is missing does not depend on either the true value of δ or the values of X and Z , then a missing δ is said to be missing completely at random (MCAR). Alternatively, if the probability that δ is missing depends on the values of X and Z but not on the true value of δ , then a missing δ is said to be missing at random (MAR); see Little & Rubin (1987).

Under the MCAR assumption and in the absence of covariates, Dinse (1982) obtained a nonparametric maximum likelihood estimator (NPML) of the survival function using an EM algorithm. Lo (1991) proved that there are infinitely many NPMLs and some of them may be inconsistent; he consequently constructed a consistent and asymptotically normal estimator. Gijbels, Lin & Ying (1993, 2007) and McKeague & Subramanian (1998) proposed further improvements on these estimators. When covariates are present, Gijbels, Lin & Ying (1993) initiated research on estimation under the Cox model. McKeague & Subramanian (1998) provided an alternative approach to estimation. Subramanian (2000) considered estimation under proportionality of conditional hazards. Zhou & Sun (2003) studied the additive hazards regression model.

Under the MAR assumption, van der Laan & McKeague (1998) first addressed efficient estimation of the survival function and proposed a sieved nonparametric maximum likelihood estimator. Further developments along the lines of efficient estimation can be found in Subramanian (2004, 2006) and Wang & Ng (2008). Goetghebeur & Ryan (1995) and Lu & Tsiatis (2001) analyzed competing risks data with missing cause of failure under proportional hazards regression models. Gao & Tsiatis (2005) considered the linear transformation competing risks model with missing cause of failure. Recently, Lu & Liang (2008) studied competing risks data with missing cause of failure under the semiparametric additive hazards model, and suggested the inverse probability weighted (IPW) and double robust (DR) estimators. To obtain these estimators, however, they imposed parametric models for two components: the probability that the censoring indicator is not missing and the conditional probability of a given failure type.

In this article, we propose estimators for the regression parameters in a semiparametric additive hazards model, where the failure times are subject to right censoring and some censoring indicators are missing at random. We provide simple and fully augmented weighted estimators that incorporate incomplete data nonparametrically. Unlike Lu & Liang (2008), no parametric models are assumed for the missingness probability or the conditional probability of an uncensored observation; instead, we use nonparametric kernel smoothing techniques to estimate these probabilities. The resulting estimators have closed forms and are easy to implement. Under the usual MAR assumption, both the simple and fully augmented weighted estimators are consistent and asymptotically equivalent, i.e., they have the same asymptotic normal distribution. In addition, the asymptotic properties of the estimated baseline cumulative hazard function are also established for the model.

The remainder of the paper is organized as follows. Section 2 presents the simple and fully augmented weighted estimators and their asymptotic properties under the MAR assumption. Section 3 reports simulation results that show the proposed estimators perform well. In

Section 4, our methods are applied to analyze the brain cancer data described earlier. Our concluding remarks follow in Section 5 and technical proofs are relegated to the Appendix.

2. ESTIMATION PROCEDURE

Under an additive hazards model, the hazard function for failure time T given covariate Z is assumed to be of the form

$$\lambda(t|Z) = \lambda_0(t) + \beta_0' Z, \quad (1)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and β_0 is a p -vector of unknown regression parameters. In the case where all data are observed, Lin & Ying (1994) introduced a pseudoscore function for the parameter vector β_0 and showed that the resulting estimator is consistent and asymptotically normal, with an easily estimated covariance matrix.

When censoring indicators are missing for right-censored data, we observe n independent and identically distributed vectors $(X_i, \xi_i, \delta_i, Z_i, R_i)$ ($i = 1, \dots, n$), where ξ_i is an indicator that δ_i is not missing, and R_i is an auxiliary covariate that is not used to model the hazard but may be used to describe the probability that δ_i is missing. The probability that δ_i is missing is characterized by the distribution of ξ_i given δ_i and $W_i = (X_i, Z_i, R_i)$, which is Bernoulli with probability $P\{\xi_i = 1 | \delta_i, W_i = w\}$. Under the MAR assumption (Little & Rubin 1987), we have

$$P\{\xi_i = 1 | \delta_i, W_i = w\} = P\{\xi_i = 1 | W_i = w\} \equiv \rho(w). \quad (2)$$

Another function of interest is $\pi(w) = P\{\delta_i = 1 | W_i = w, \xi_i = 1\}$, which is the conditional probability of an uncensored observation, given that δ_i is observed and $W_i = w$.

A naive method for estimating β_0 is to simply ignore the missing data and to apply the pseudoscore function of Lin & Ying (1994) to the complete data only. Such a procedure (called the complete case estimator) may not only lose efficiency due to discarding incomplete observations, but may also generate biased estimators, even when the censoring indicators are MAR. If either $\rho(w)$ or $\pi(w)$ is modeled correctly, we can use the approach of Lu & Liang (2008) to obtain the IPW and DR estimators. In many situations, however, knowledge of $\rho(w)$ and $\pi(w)$ is limited, and thus both models may be misspecified. In this article, no parametric models are assumed for these two probabilities; rather, both are estimated nonparametrically by kernel smoothers. We begin by introducing the simple weighted estimator, which is derived under the MAR assumption.

Because $\rho(W_i)$ is a function of continuous variables such as X_i , we estimate it with the Nadaraya-Watson estimator based on complete observations. Specifically, let d denote the number of continuous elements of W_i and let K be an r th-order ($r > d$) kernel function of d variables with finite support that satisfies $\int K(u) du = 1$, $\int u^m K(u) du = 0$, $m = 1, \dots, r-1$, $\int u^r K(u) du \neq 0$, and $\int K(u)^2 du < \infty$, where u can be a scalar or a vector. If u is a vector, say $u = (u_1, \dots, u_d)'$, then u^m denotes $(u_1^m, \dots, u_d^m)'$. The motivation for using higher-order kernels is to reduce the order of magnitude of the bias of the curve estimator, leading to a faster rate of convergence of the mean integrated squared error (Wand & Schucany 1990). This type of

kernel function may be constructed in various manners. For instance, Wand & Schucany (1990) gave a univariate Gaussian-based kernel of order $2r$:

$$K(u_1) = \frac{(-1)^r \varphi^{(2r-1)}(u_1)}{2^{r-1}(r-1)!u_1},$$

where $\varphi^{(2r-1)}(u_1)$ is the $(2r-1)$ -th derivative of the standard normal density function $\varphi(u_1)$. Hall & Marron (1988) proposed a class of univariate kernels of order r :

$$K(u_1) = \pi^{-1} \int_0^\infty \cos(tu_1) \exp(-t^r) dt.$$

Some higher-order polynomial kernels can be found in Müller (1984) and Gasser, Müller & Mammitzsch (1985).

Define $K_h(\cdot) = K(\cdot/h)$, where h is a bandwidth sequence, and $K(u/h) = K(u_1/h, \dots, u_d/h)$ for $u = (u_1, \dots, u_d)'$. Write $W_i = (W_{1i}, W_{2i})$, where W_{1i} and W_{2i} include all continuous and discrete elements of W_i , respectively. Then the Nadaraya-Watson estimator of $\rho(w)$ is given by

$$\hat{\rho}(w) = \frac{\sum_{i=1}^n \xi_i K_h(w_1 - W_{1i}) I(W_{2i} = w_2)}{\sum_{i=1}^n K_h(w_1 - W_{1i}) I(W_{2i} = w_2)}, \tag{3}$$

where $w = (w_1, w_2)$. The choice of the kernel function K usually has little effect on the estimator $\hat{\rho}(w)$, and thus the estimator of β_0 , but the bandwidth sequence h typically does influence these estimators, both theoretically and practically. We assume that h satisfies $nh^{2r} \rightarrow 0$ and $nh^{2d} \rightarrow \infty$ as $n \rightarrow \infty$. If $h = O(n^{-1/p})$ for some integer $p > 2d$, then a reasonable choice for r is the smallest even integer such that $r \geq p - d$ (Qi, Wang & Prentice 2005). For example, when $d = 2$, we might choose $p = 5$ and $r = 4$. In a similar manner, we can estimate $\pi(w)$ by

$$\hat{\pi}(w) = \frac{\sum_{i=1}^n \xi_i \delta_i K_h(w_1 - W_{1i}) I(W_{2i} = w_2)}{\sum_{i=1}^n \xi_i K_h(w_1 - W_{1i}) I(W_{2i} = w_2)}. \tag{4}$$

Note that the kernel function K and bandwidth sequence h used in (3) need not be identical to those used in (4), and the bandwidth can be different for each component of W_{1i} . For example, we can define $h = (h_1, \dots, h_d)'$ for different bandwidths, and write $K(u/h) = K(u_1/h_1, \dots, u_d/h_d)$. Here, we use the same K and h in both for notational convenience.

Let $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ denote the baseline cumulative hazard function. Using the inverse probability weighted approach, consider the following estimating equations for β_0 and Λ_0 :

$$\sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\hat{\rho}(W_i)} Z_i \left[dN_i^u(t) - Y_i(t) \beta' Z_i dt - Y_i(t) d\Lambda_0(t) \right] = 0, \tag{5}$$

$$\sum_{i=1}^n \frac{\xi_i}{\widehat{\rho}(W_i)} \left[dN_i^u(t) - Y_i(t)\beta' Z_i dt - Y_i(t)d\Lambda_0(t) \right] = 0, \tag{6}$$

where $N_i^u(t) = I(X_i \leq t, \delta_i = 1)$, $Y_i(t) = I(X_i \geq t)$, and τ is a prespecified positive constant such that $P(X_i \geq \tau) > 0$. The resulting simple weighted estimators for β_0 and Λ_0 have the following closed forms:

$$\widehat{\beta} = \left[\sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\widehat{\rho}(W_i)} Y_i(t) \{Z_i - \bar{Z}(t)\}^{\otimes 2} dt \right]^{-1} \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\widehat{\rho}(W_i)} \{Z_i - \bar{Z}(t)\} dN_i^u(t)$$

and

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n \widehat{\rho}(W_i)^{-1} \xi_i [dN_i^u(s) - Y_i(s)\widehat{\beta}' Z_i ds]}{\sum_{i=1}^n \widehat{\rho}(W_i)^{-1} \xi_i Y_i(s)},$$

where $a^{\otimes 2} = aa'$ for any vector a and

$$\bar{Z}(t) = \frac{\sum_{i=1}^n \widehat{\rho}(W_i)^{-1} \xi_i Y_i(t) Z_i}{\sum_{i=1}^n \widehat{\rho}(W_i)^{-1} \xi_i Y_i(t)}.$$

In practice, we often choose τ to be the largest observation time, say $\tau = \max\{X_i\}$.

Let $\bar{z}(t) = E[Y_i(t)Z_i(t) | E[Y_i(t)]]$. Define $N_i(t) = I(X_i \leq t)$ and $B_i = \int_0^\tau \{Z_i - \bar{z}(t)\} dN_i(t)$. The asymptotic properties of $\widehat{\beta}$ are given in the following theorem.

Theorem 1

Under regularity conditions (C1)–(C6), which are stated in the Appendix, $\widehat{\beta}$ is consistent and $n^{1/2}(\widehat{\beta} - \beta_0)$ is asymptotically normal with mean zero and covariance matrix $V = A^{-1}\Sigma A^{-1} + A^{-1}\Sigma^* A^{-1}$, where

$$\Sigma = E \left[\int_0^\tau \{Z_i - \bar{z}(t)\}^{\otimes 2} dN_i^u(t) \right],$$

$$\Sigma^* = E \left[\pi(W_i)(1 - \pi(W_i))(1 - \rho(W_i))\rho(W_i)^{-1} B_i^{\otimes 2} \right],$$

and

$$A = E \left[\int_0^\tau Y_i(t) \{Z_i - \bar{z}(t)\}^{\otimes 2} dt \right].$$

Note that the first term in V is the asymptotic variance of the Lin & Ying (1994) estimator based only on the complete data ($\xi_i \equiv 1$) and the second term represents the effect of the

missing censoring indicators. If we let $\widehat{B}_i = \int_0^\tau \{Z_i - \bar{Z}(t)\} dN_i(t)$, then the covariance matrix V can be consistently estimated by $\widehat{V} = \widehat{A}^{-1}(\widehat{\Sigma} + \widehat{\Sigma}^*)\widehat{A}^{-1}$, where

$$\begin{aligned} \widehat{\Sigma} &= n^{-1} \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\widehat{\rho}(W_i)} \{Z_i - \bar{Z}(t)\}^{\otimes 2} dN_i^u(t), \\ \widehat{\Sigma}^* &= n^{-1} \sum_{i=1}^n \widehat{\pi}(W_i)(1 - \widehat{\pi}(W_i))(1 - \widehat{\rho}(W_i))\widehat{\rho}(W_i)^{-1}\widehat{B}_i^{\otimes 2}, \end{aligned}$$

and

$$\widehat{A} = n^{-1} \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\widehat{\rho}(W_i)} Y_i(t) \{Z_i - \bar{Z}(t)\}^{\otimes 2} dt.$$

Define $d(t) = \int_0^t \bar{z}(s) ds$ and

$$D(t) = \int_0^t \left\{ \frac{E[Y_i(s)Z_i^{\otimes 2}]}{E[Y_i(s)]} - \bar{z}(s)^{\otimes 2} \right\} \beta_0 ds.$$

The asymptotic properties of $\widehat{\Lambda}_0(t)$ are given in the next theorem.

Theorem 2

Under the assumptions of Theorem 1, $\widehat{\Lambda}_0(t)$ converges in probability to $\Lambda_0(t)$ uniformly in $t \in [0, \tau]$, and $n^{1/2}\{\widehat{\Lambda}_0(t) - \Lambda_0(t)\}$ converges weakly on $[0, \tau]$ to a zero-mean Gaussian process with covariance function at (t, s) ($t \leq s$) equal to

$$\begin{aligned} \Gamma(t, s) &= \int_0^t \frac{dE\{N_i^u(u)\}}{E\{Y_i(u)\}^2} + E \left\{ \pi(W_i)(1 - \pi(W_i))(1 - \rho(W_i))\rho(W_i)^{-1} \int_0^t \frac{dN_i(u)}{E\{Y_i(u)\}^2} \right\} - d(t)' A^{-1} \left\{ \pi(W_i)(1 - \pi(W_i))(1 - \rho(W_i))\rho(W_i)^{-1} \int_0^s \frac{(Z_i - \bar{z}(u))dN_i(u)}{E\{Y_i(u)\}} \right. \\ &\quad \left. - d(s)' A^{-1} \left\{ \pi(W_i)(1 - \pi(W_i))(1 - \rho(W_i))\rho(W_i)^{-1} \int_0^t \frac{(Z_i - \bar{z}(u))dN_i(u)}{E\{Y_i(u)\}} \right\} + d(t)' A^{-1}(\Sigma + \Sigma^*)A^{-1}d(s) - d(t)' A^{-1}D(s) - d(s)' A^{-1}D(t) \right\}. \end{aligned}$$

The covariance function $\Gamma(t, s)$ can be consistently estimated by substituting $\widehat{\beta}$, $\widehat{\rho}$ and $\widehat{\pi}$ for the unknowns β_0, ρ and π in the appropriate empirical estimators, and by replacing the (unobserved) processes N_i^u with $\widehat{\rho}(W_i)^{-1}\xi_i N_i^u$. For an individual with a given covariate vector z_0 , the corresponding estimator of the survival function $S(t, z_0)$ is

$$\widehat{S}(t, z_0) = \exp\{-\widehat{\Lambda}_0(t) - \widehat{\beta}' z_0 t\}.$$

Using the functional delta-method and Theorem 2, we can obtain the asymptotic properties of $\widehat{S}(t, z_0)$, which can be applied to construct confidence bands for $S(t, z_0)$.

When the missingness probability $\rho(w)$ is known or a parametric model is specified for $\rho(w)$, the simple weighted estimator uses only the complete case data (i.e., only individuals with $\xi_i = 1$), and the fully augmented weighted estimator (also called the double robust estimator)

incorporates contributions from the incomplete observations (i.e., individuals with $\xi_i = 0$), thus the fully augmented weighted estimator is more efficient than the corresponding simple weighted estimator (Lu & Liang 2008). In addition, the fully augmented weighted estimator has the so-called double-robustness property; that is, the estimator is consistent if one can correctly specify either the missingness probability $\rho(w)$ or the conditional probability of an uncensored observation $\pi(w)$ (Wang & Chen 2001). However, estimating $\rho(w)$ nonparametrically enables the simple weighted estimator $\hat{\beta}$ to follow the same asymptotic distribution as the fully augmented weighted estimator $\hat{\beta}_a$ (described next). This indicates that $\hat{\beta}$ is equivalent asymptotically to $\hat{\beta}_a$. These conclusions are consistent with the results of Qi, Wang & Prentice (2005) for proportional hazards regression with missing covariates.

The fully augmented weighted estimators for β_0 and Λ_0 are the solutions to the following estimating equations:

$$\sum_{i=1}^n \int_0^r Z_i \left[\frac{\xi_i}{\hat{\rho}(W_i)} dN_i^u(t) + \left(1 - \frac{\xi_i}{\hat{\rho}(W_i)} \right) \hat{\pi}(W_i) dN_i(t) - Y_i(t) \beta' Z_i dt - Y_i(t) d\Lambda_0(t) \right] = 0, \tag{7}$$

$$\sum_{i=1}^n \left[\frac{\xi_i}{\hat{\rho}(W_i)} dN_i^u(t) + \left(1 - \frac{\xi_i}{\hat{\rho}(W_i)} \right) \hat{\pi}(W_i) dN_i(t) - Y_i(t) \beta' Z_i dt - Y_i(t) d\Lambda_0(t) \right] = 0. \tag{8}$$

The resulting fully augmented weighted estimators for β_0 and Λ_0 have the following closed forms:

$$\hat{\beta}_a = \left[\sum_{i=1}^n \int_0^r Y_i(t) \{Z_i - \bar{Z}^*(t)\}^{\otimes 2} dt \right]^{-1} \sum_{i=1}^n \int_0^r \{Z_i - \bar{Z}^*(t)\} \times \left[\frac{\xi_i}{\hat{\rho}(W_i)} dN_i^u(t) + \left(1 - \frac{\xi_i}{\hat{\rho}(W_i)} \right) \hat{\pi}(W_i) dN_i(t) \right]$$

and

$$\hat{\Lambda}_a(t) = \int_0^t \frac{\sum_{i=1}^n \left[\hat{\rho}(W_i)^{-1} \xi_i dN_i^u(s) + (1 - \hat{\rho}(W_i)^{-1} \xi_i) \hat{\pi}(W_i) dN_i(s) - Y_i(s) \hat{\beta}_a' Z_i ds \right]}{\sum_{i=1}^n Y_i(s)},$$

where

$$\bar{Z}^*(t) = \frac{\sum_{i=1}^n Z_i Y_i(t)}{\sum_{i=1}^n Y_i(t)}.$$

Similar to Theorems 1 and 2, the asymptotic properties of $\hat{\beta}_a$ and $\hat{\Lambda}_a$ are given in the following theorem.

Theorem 3

Under the assumptions of Theorem 1, we have:

- i. $\hat{\beta}_a$ is consistent and $n^{1/2}(\hat{\beta}_a - \beta_0)$ is asymptotically normal with mean zero and covariance matrix V:

- ii. $\hat{\Lambda}_a(t)$ converges in probability to $\Lambda_0(t)$ uniformly in $t \in [0, \tau]$, and $n^{1/2}\{\hat{\Lambda}_a(t) - \Lambda_0(t)\}$ converges weakly on $[0, \tau]$ to a zero-mean Gaussian process with covariance function $\Gamma(t, s)$ at (t, s) ($t \leq s$).

For the fully augmented weighted method, the covariance matrix V and covariance function $\Gamma(t, s)$ can be consistently estimated by substituting $\hat{\beta}_a$, $\hat{\rho}$ and $\hat{\pi}$ for β_0 , ρ and π in the appropriate empirical estimators, and by replacing the processes N_i^u with

$$\hat{\rho}(W_i)^{-1} \xi_i N_i^u + (1 - \xi_i \hat{\rho}(W_i)^{-1}) \hat{\pi}(W_i) N_i.$$

Theorems 1, 2 and 3 show that both the simple and fully augmented weighted estimators have the same asymptotic normal distribution, and the resulting estimators of the baseline cumulative hazard function converge to the same Gaussian process. This means that the simple weighted estimators with nonparametric $\hat{\rho}(w)$ are as efficient as the kernel-assisted fully augmented weighted estimators. One intuitive explanation for this is that the incomplete observations are indirectly incorporated in the simple weighted estimator by using the inverse of $\hat{\rho}(w)$ as a weight.

Note that $\hat{\Lambda}_0(t)$ and $\hat{\Lambda}_a(t)$ may not always be monotonic in t , in which case simple modifications such as those discussed in Lin & Ying (1994) can be made to ensure monotonicity while preserving asymptotic properties.

3. SIMULATION STUDIES

We conducted simulation studies to examine and compare the finite-sample performance of the simple and fully augmented weighted estimators proposed in Section 2, and also to compare their performance with that of the full data and complete-case analyses under the MAR model. In these studies, we considered three situations for the covariate Z : (a) Z was assumed to follow a Bernoulli distribution with success probability 0.5; (b) Z was generated from a uniform distribution on $(0, 1)$; (c) $Z = (Z_1, Z_2)'$, where Z_1 follows a uniform distribution on $(0, 1)$ and Z_2 follows a Bernoulli distribution with success probability 0.5.

The underlying additive hazards model for the failure time T was taken to be $\lambda(t|Z) = 1 + \beta_0' Z$, where $\beta_0 = 0, 0.5$ and 1 for the case Z is a scalar, and $\beta_0 = (0, 0)'$ and $\beta_0 = (1, -1)'$ for the two-dimensional covariate. The censoring time C was generated from a uniform distribution on $(0, c)$, where c was selected to give a censoring rate of either 15% or 55%.

The missingness indicators were generated from the logistic model

$$\rho(W) = \frac{\exp(\theta' W)}{1 + \exp(\theta' W)}, \quad (9)$$

where $W = (X, Z)$, $X = T \wedge C$, and θ was chosen to produce a missingness rate of 50% under each censoring level. When Z was a Bernoulli random variable, there was only one ($d = 1$) continuous element in W , and we used the univariate Gaussian kernel function $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ and a bandwidth of $h = 0.5n^{-1/3}$, with sample size of $n = 100$. When Z was a uniform random variable or a two-dimensional covariate as in (c), there were two ($d = 2$) continuous elements in W , and we used the bivariate Gaussian-based kernel function of order 4 (Wand & Schucany 1990)

$$K(u_1, u_2) = \frac{1}{8\pi} (3 - u_1^2)(3 - u_2^2) \exp(-(u_1^2 + u_2^2)/2) \quad (10)$$

and a bandwidth vector of $h = (h_1, h_2)' = (1.5n^{-1/5}, n^{-1/5})'$, with sample size of $n = 400$. We took τ to be the largest observed value of X , so that all data were used in the analysis. All simulation studies were based on 1000 replications for each combination of parameters.

Our simulation results are summarized in Tables 1 and 2. In these tables, Bias is the sample mean of the estimate minus the true value; MSE is the sample mean of the squared differences between the estimate and the true value; and CP is the 95% empirical coverage probability for β_0 based on a normal approximation. Similar summaries for the full-data and complete-case estimators are calculated for comparison.

Tables 1 and 2 show that the complete case estimator is highly biased in all situations, with coverage probabilities that are too small, whereas the simple and fully augmented weighted estimators are nearly unbiased, with very reasonable coverage probabilities. Furthermore, the simple and fully augmented weighted estimators have similar MSE values, which are only slightly larger than those of the full data estimator and are often much smaller than those of the complete case estimator. These results suggest that our proposed estimators are more efficient than the complete case estimator and are adequate for practical use. We also simulated data under different parameter configurations and obtained similar results.

We compared the proposed methods and the parametric approach of Lu & Liang (2008) under MAR and MCAR assumptions. Data were simulated under correctly and incorrectly specified parametric models, using the same setup as in Table 1 with a censoring rate of 55% and a missingness rate of 50%, where Z follows a Bernoulli distribution with a sample size of $n = 200$ and $\beta_0 = 0$ and 1. The results are presented in Table 3. In Table 3, LIPW1 and LIPW2 stand for the inverse probability weighted (IPW) estimators of Lu & Liang (2008) when using the logistic model and the constant model for $\rho(w)$, respectively; LDR1 and LDR2 stand for the double robust (DR) estimators of Lu & Liang (2008) when using the logistic model and the constant model for $\rho(w)$, respectively. In all cases, we used a constant model for $\pi(w)$, which is misspecified.

It can be seen from Table 3 that the proposed methods are essentially unbiased in all the settings, and the parametric approach of Lu & Liang (2008) is also unbiased when the parametric model for $\rho(w)$ is correctly specified. Furthermore, the proposed estimators are as efficient as the DR estimator of Lu & Liang (2008), and are more efficient than the IPW estimator of Lu & Liang (2008). When both $\rho(w)$ and $\pi(w)$ are misspecified, however, both the IPW and DR estimators of Lu & Liang (2008) are biased under MAR. The key advantage of our method is that it provides reasonable estimation without making parametric modeling assumptions about $\rho(w)$ and $\pi(w)$. Rather than assuming parametric models, our approach uses nonparametric smoothing techniques to estimate these probabilities. In addition, the proposed estimators are more efficient than the complete case estimator under MCAR. So, if MCAR is true, our proposed approach still works well and does not lose efficiency.

We also conducted simulation studies to examine the performance of the proposed methods when MNAR (missing not at random) is true. In the study, the setup was the same as in Table 1, where Z follows a uniform distribution on $(0, 1)$ with $\beta_0 = 0$ and $n = 400$, except that the censoring rate was set to be 20%, and the missingness probability was given by

$$\rho(W, \delta) = \frac{\exp(\theta_1' W + \theta_2 \delta)}{1 + \exp(\theta_1' W + \theta_2 \delta)},$$

where θ_1 and θ_2 were chosen to produce a missingness rate of either 20% or 50%. The results are summarized in Table 4. It can be seen from Table 4 that the proposed estimation procedures perform well when the missingness rate is low (say, 20%), but when the missingness rate is high (say, 50%), the proposed estimators are a little biased. However, the biases are relatively small compared to those of the complete case estimator.

4. EXAMPLE: ANALYSIS OF BRAIN CANCER DATA

We applied our methods to the brain cancer data mentioned earlier. We analyzed the data on all 387 patients who entered the clinical trial with a form of brain cancer known as glioblastoma. Dinse (1982) used a subset of these data to illustrate his nonparametric maximum likelihood analysis, which did not account for covariates. All patients were ambulatory when they entered the trial, but over time some lost their mobility, some had a progression of their cancer, and some experienced both events. As a measure of quality of life, we defined “survival time” as the time to non-ambulatory progression, and we evaluated the effects of sex and age on this event time.

Of the 387 patients, 86 progressed and were non-ambulatory, 24 progressed but were still ambulatory, 220 did not progress by the end of the study, and 57 progressed but had an unknown ambulatory status. Thus, our analysis treated these outcomes as 86 uncensored times, 244 censored times, and 57 times with a missing censoring indicator. There were 144 women and 243 men, ranging in age from 14 to 74 years, and the length of time on study (or until progression) varied from 2 to 1088 days.

Let X be the observed time (in days), measured from the beginning of the trial, and let δ indicate whether the patient had progressed and was non-ambulatory. We defined Z_1 to be a binary indicator of the patient’s sex, which was 1 for men and 0 for women, and Z_2 to be the age at trial entry (in years), which was treated as a continuous covariate. Since $W = (X, Z_1, Z_2)$ contains two continuous elements, we used the bivariate Gaussian-based kernel function of order 4 for K , as defined in (10), with a bandwidth vector of $h = (h_1, h_2)' = (34, 10)'$. We used $\tau = 1088$, which was the largest observed value of X .

The analysis of the brain cancer data is summarized in Table 5, which gives the results for our simple weighted estimator (SWE) and our fully augmented weighted estimator (FAWE). For comparison, Table 5 also gives the results of the complete case (CC) analysis. None of the three methods suggested that men and women had different hazard rates for non-ambulatory progression. On the other hand, our two estimators showed that age is important ($p = 0.037$ for SWE and $p = 0.011$ for FAWE), but the CC analysis did not ($p = 0.367$). Specifically, the hazard rate for non-ambulatory progression increased as patients grew older, which is consistent with worsening quality of life. The age coefficients were of similar magnitude for all three methods, but the standard error was much larger for the CC analysis than for our SWE and FAWE analyses. Thus, as a result of excluding data, the complete case analysis missed the age effect on non-ambulatory progression that our approaches appropriately identified.

5. CONCLUDING REMARKS

Model (1) has the limitation that the linear predictor $\beta_0'Z$ needs to be constrained to ensure non-negativity for the right side of (1). One may avoid this constraint by using a nonnegative link function, such as $\lambda_0(t) + \exp(\beta_0'Z)$. The ideas presented in this paper can be applied to any regression function $g(\beta_0'Z)$, where $g(\cdot)$ is a known link function. In addition, Our approach can be extended to incorporate missing covariates (Qi, Wang & Prentice 2005) in the situation where both the failure indicators and the covariates are partially observed.

Nonparametric kernel estimation can be done for a small number of continuous covariates, but for categorical covariates, it would usually require stratified kernel estimation within each strata defined by the categorical covariates. In practice, when there are too many categories, it may be desirable to specify a more flexible model for the missingness probability, such as a partially linear additive model, and then use local kernel regression to estimate the missingness probability. Here we focus on a kernel estimation approach for $\rho(w)$ and $\pi(w)$. Of course, other smoothing techniques such as the local polynomial method (Fan & Gijbels 1996) may be used and require the same assumptions. Furthermore, $n^{1/2}$ -rate asymptotic normality of the proposed estimators indicates that an appropriate choice for the bandwidth sequence h depends only on the second order terms of the mean square error of the estimators, and thus bandwidth selection may not be critical for estimating β_0 and Λ_0 .

Since the estimating functions in (5) to (8) were obtained in a somewhat ad hoc fashion, it might be worthwhile investigating possible improvements that could result from other approaches, such as the one suggested by McKeague & Sasieni (1994) or perhaps a nonparametric maximum likelihood approach. Alternatively, estimation procedures based on the general Aalen additive model (Aalen 1980) or the linear transformation model (Gao & Tsiatis 2005) with missing censoring information might also be worthy of investigation.

Another limitation of the approach given here is that the covariates Z are time-invariant. In some applications, we might want to incorporate time-dependent covariates. Thus, a more general approach might extend model (1) to a time-varying version:

$$\lambda(t|Z(t)) = \lambda_0(t) + \beta_0(t)'Z(t),$$

where $\beta_0(t)$ is an unknown p -vector of time-varying regression coefficients and $Z(t)$ is a vector of covariates that may depend on time. However, the proposed estimation procedure cannot be extended in a straightforward manner to deal with time-dependent covariates because of the curse of dimensionality created by $Z(t)$ and a need for alternative smoothing techniques for estimating $\beta_0(t)$. In addition, when the dimension of $Z(t)$ is high, the probabilities $\rho(w)$ and $\pi(w)$ can be modeled parametrically (Lu & Liang 2008). As a different approach, perhaps dimension-reduction techniques could be extended in conjunction with a partially linear model (Liang, Härdle & Carroll 1999) for $\rho(w)$ and $\pi(w)$.

Acknowledgments

The authors would like to thank the Editor (Paul Gustafson), the Associate Editor, two reviewers and Shyamal Peddada for their constructive and insightful comments and suggestions that greatly improved the paper. Xinyuan Song's research was fully supported by two grants from the Research Grant Council of the Hong Kong Special Administration Region. Liuquan Sun's research was fully supported by the National Natural Science Foundation of China Grants, the National Basic Research Program of China (973 Program) and Key Laboratory of RCSDS, CAS. Gregg Dinse's research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

References

- Aalen, OO. A model for nonparametric regression analysis of counting processes. In: Klonecki, W.; Kozek, A.; Rosinski, J., editors. *Mathematical Statistics and Probability Theory, Lecture Notes in Statistics*. 2. Springer-Verlag; New York: 1980. p. 1-25.
- Dinse GE. Nonparametric estimation for partially-complete time and type of failure data. *Biometrics* 1982;38:417-431. [PubMed: 6896831]
- Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*. Chapman & Hall; London: 1996.
- Gao G, Tsiatis AA. Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika* 2005;92:875-891.
- Gasser T, Müller HG, Mammitzsch V. Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society Series B* 1985;47:238-252.
- Gijbels, I.; Lin, D.; Ying, Z. Tech Report 039-93. Mathematical Sciences Research Institute; Berkeley: 1993. Non- and semi-parametric analysis of failure time data with missing failure indicators.
- Gijbels I, Lin D, Ying Z. Non- and semi-parametric analysis of failure time data with missing failure indicators. *IMS Lecture Notes-Monograph Series. Inverse Problems: Tomography, Networks and Beyond* 2007;54:203-223.
- Goetghebeur EJ, Ryan L. Analysis of competing risks survival data when some failure types are missing. *Biometrika* 1995;82:821-833.
- Hall P, Marron JS. Choice of kernel order in density estimation. *The Annals of Statistics* 1988;16:161-173.
- van der Laan MJ, McKeague IW. Efficient estimation from right-censored data when failure indicators are missing at random. *The Annals of Statistics* 1998;26:164-182.
- Liang H, Härdle W, Carroll RJ. Estimation in a semiparametric partially linear errors-invariable model. *The Annals of Statistics* 1999;27:1519-1535.
- Lin DY, Ying Z. Semiparametric analysis of the additive risk model. *Biometrika* 1994;81:61-71.
- Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. Wiley; New York: 1987.
- Lo S-H. Estimating a survival function with incomplete cause-of-death data. *Journal of Multivariate Analysis* 1991;39:217-235.
- Lu K, Tsiatis AA. 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 IW, Sasieni PD. A partly parametric additive risk model. *Biometrika* 1994;81:501-514.
- McKeague IW, Subramanian S. Product-limit estimators and Cox regression with missing censoring information. *Scandinavian Journal of Statistics* 1998;25:589-601.
- Müller HG. Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics* 1984;12:766-774.
- Pollard, D. *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics; Hayward, California: 1990.
- Qi L, Wang CY, Prentice RL. Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association* 2005;100:1250-1263.
- 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. Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *Journal of Nonparametric Statistics* 2004;16:797-817.
- Subramanian S. Survival analysis for the missing censoring indicator model using kernel density estimation techniques. *Statistical Methodology* 2006;3:125-136. [PubMed: 18953423]
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. Springer-Verlag; New York: 1996.
- Wand MP, Schucany WR. Gaussian-based kernels. *The Canadian Journal of Statistics* 1990;18:197-204.

Wang CY, Chen HY. Augmented inverse probability weighted estimator for Cox missing covariate regression. *Biometrics* 2001;57:414–419. [PubMed: 11414564]

Wang Q, Ng KW. Asymptotically efficient product-limit estimators with censoring indicators missing at random. *Statistica Sinica* 2008;18:749–768.

Zhou X, Sun L. Additive hazards regression with missing censoring information. *Statistica Sinica* 2003;13:1237–1257.

APPENDIX

We will use the same notation defined in the previous sections and assume that the following regularity conditions hold:

(C1) $\Lambda_0(\tau) < \infty$, $Pr(X \geq \tau) > 0$, Z is bounded, and $\inf_{0 \leq t \leq \tau} [\lambda_0(t) + \beta_0' Z] > 0$ a.e.

(C2) The probability (density) $f(w)$ of W_i is bounded away from 0, and has r continuous and bounded partial derivatives with respect to the continuous components of W_i a.e.

(C3) The missingness probability $\rho(w)$ is bounded away from 0, and has r continuous and bounded partial derivatives with respect to the continuous components of W_i a.e.

(C4) The conditional probability $\pi(w)$ has r continuous and bounded partial derivatives with respect to the continuous components of W_i a.e.

(C5) $A = E[\int_0^\tau Y_i(t) \{Z_i - \bar{Z}(t)\}^{\otimes 2} dt]$ is nonsingular.

(C6) $nh^{2r} \rightarrow 0$ and $nh^{2d} \rightarrow \infty$, as $n \rightarrow \infty$.

We give the proof of Theorem 3 and outline the proof of Theorem 1; Theorem 2 can be proven in the same manner. For notational convenience, we assume that all components of W_i are continuous in the following proof.

Proof of Theorem 3(i)

Substituting $\hat{\Lambda}_a$ into equation (7), we find that $\hat{\beta}_a$ is the solution to $U(\beta) = 0$, where

$$U(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}^*(t)\} \left[\frac{\xi_i}{\rho(W_i)} dN_i^u(t) + \left(1 - \frac{\xi_i}{\rho(W_i)}\right) \widehat{\pi}(W_i) dN_i(t) - Y_i(t) \beta' Z_i dt \right].$$

Let $M_i(t) = N_i^u(t) - \int_0^t Y_i(s) \{\lambda_0(s) + \beta_0' Z_i\} ds$. Then we can write

$$U(\beta_0) = U_1(\beta_0) + U_2(\beta_0) + U_3(\beta_0), \tag{A.1}$$

where

$$U_1(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\} dM_i(t),$$

$$U_2(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\} \left(\frac{\xi_i}{\rho(W_i)} - 1 \right) dN_i^u(t),$$

and

$$U_3(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(t)\} \left(1 - \frac{\xi_i}{\widehat{\rho}(W_i)}\right) \widehat{\pi}(W_i) dN_i(t).$$

Note that $U_1(\beta_0)$ is a martingale integral. Thus, it follows that

$$U_1(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) + o_p(n^{1/2}). \tag{A.2}$$

Define $\Phi_n(t) = n^{-1} \sum_{i=1}^n (\widehat{\rho}(W_i)^{-1} \xi_i - 1) N_i^u(t)$, and write $\Phi_n(t) = \Phi_{n1}(t) + \Phi_{n2}(t) + \Phi_{n3}(t)$, where

$$\begin{aligned} \Phi_{n1}(t) &= n^{-1} \sum_{i=1}^n (\xi_i - \rho(W_i)) \frac{N_i^u(t)}{\rho(W_i)}, \\ \Phi_{n2}(t) &= n^{-1} \sum_{i=1}^n (\rho(W_i) - \widehat{\rho}(W_i)) \frac{N_i^u(t)}{\rho(W_i)}, \end{aligned}$$

and

$$\Phi_{n3}(t) = n^{-1} \sum_{i=1}^n (\xi_i - \widehat{\rho}(W_i)) (\rho(W_i) - \widehat{\rho}(W_i)) \frac{N_i^u(t)}{\widehat{\rho}(W_i) \rho(W_i)}.$$

By the uniform strong law of large numbers (Pollard, 1990), $\sup_{0 \leq t \leq \tau} |\Phi_{n1}(t)| = o(1)$ almost surely. It can be checked that

$$\Phi_{n2}(t) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \frac{(\rho(W_i) - \xi_j) K_h(W_i - W_j) N_i^u(t)}{\rho(W_i) h^d \widehat{f}(W_i)},$$

where $\widehat{f}(w) = (nh^d)^{-1} K_h(w - W_i)$, which is a kernel density estimate of $f(w)$. By a Taylor expansion of $1/\widehat{f}(W_i)$ about $f(W_i)$, $\Phi_{n2}(t)$ can be written as $\Phi_{n21}(t) - \Phi_{n22}(t) + o_p(1)$, where

$$\Phi_{n21}(t) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \frac{(\rho(W_i) - \xi_j) K_h(W_i - W_j) N_i^u(t)}{\rho(W_i) h^d f(W_i)},$$

and

$$\Phi_{n22}(t) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \frac{(\rho(W_i) - \xi_j) K_h(W_i - W_j) N_i^u(t) (\widehat{f}(W_i) - f(W_i))}{\rho(W_i) h^d f(W_i)^2}.$$

A straightforward calculation yields that $E\{\Phi_{n21}(t)\} = O(h^r) \rightarrow 0$, and $Var\{\Phi_{n21}(t)\} = O(h^{2r} + (nh^{2d})^{-1}) \rightarrow 0$, which imply $\Phi_{n21}(t) = o_p(1)$. Similarly, $\Phi_{n22}(t) = o_p(1)$, and thus it follows that $\Phi_{n2}(t) = o_p(1)$. Likewise, $\Phi_{n3}(t) = o_p(1)$. Therefore, we have $\Phi_n(t) = o_p(1)$. Note that $\Phi_n(t)$ is monotone and bounded in t . Consequently, we obtain

$$\sup_{0 \leq t \leq \tau} |\Phi_n(t)| = o_p(1). \tag{A.3}$$

The functional central limit theorem (Pollard 1990) implies that

$$\sup_{0 \leq t \leq \tau} |\bar{Z}^*(t) - \bar{z}(t)| = O_p(n^{-1/2}). \tag{A.4}$$

Using (A.3) and (A.4), we have

$$\int_0^\tau \{\bar{Z}^*(t) - \bar{z}(t)\} d\Phi_n(t) = o_p(n^{-1/2}).$$

Hence,

$$\begin{aligned} U_2(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} \left(\frac{\xi_i}{\rho(W_i)} - 1\right) dN_i^u(t) - n \int_0^\tau \{\bar{Z}^*(t) - \bar{z}(t)\} d\Phi_n(t) \\ &= \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1\right) \delta_i B_i + o_p(n^{1/2}). \end{aligned} \tag{A.5}$$

In a similar manner, we obtain

$$U_3(\beta_0) = \sum_{i=1}^n \left(1 - \frac{\xi_i}{\rho(W_i)}\right) \widehat{\pi}(W_i) B_i + o_p(n^{1/2}). \tag{A.6}$$

Thus, it follows from (A.1), (A.2), (A.5) and (A.6) that

$$\begin{aligned} U(\beta_0) &= \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) + \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1\right) (\delta_i - \widehat{\pi}(W_i)) B_i + o_p(n^{1/2}) \\ &= \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) + \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1\right) (\delta_i - \pi(W_i)) B_i + R_{n1} + R_{n2} + o_p(n^{1/2}), \end{aligned}$$

where

$$R_{n1} = \sum_{i=1}^n \left(1 - \frac{\xi_i}{\rho(W_i)}\right) (\widehat{\pi}(W_i) - \pi(W_i)) B_i,$$

and

$$R_{n2} = \sum_{i=1}^n (\widehat{\rho}(W_i) - \rho(W_i)) (\widehat{\pi}(W_i) - \delta_i) \frac{\xi_i B_i}{\widehat{\rho}(W_i) \rho(W_i)}.$$

Let $m(w) = \rho(w)f(w)$ and $\widehat{m}(w) = (nh^d)^{-1} \sum_{i=1}^n \xi_i K_h(w - W_i)$. Then by the Taylor expansion of $1/\widehat{m}(W_i)$ at $m(W_i)$, we can write

$$\sum_{i=1}^n (\widehat{\pi}(W_i) - \pi(W_i)) B_i = R_{n11} + R_{n12} + o_p(n^{1/2}),$$

where

$$R_{n11} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\xi_j (\delta_j - \pi(W_i)) K_h(W_i - W_j) B_i}{h^d m(W_i)},$$

and

$$R_{n12} = -n^{-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\xi_j (\delta_j - \pi(W_i)) K_h(W_i - W_j) (\widehat{m}(W_i) - m(W_i)) B_i}{h^d m(W_i)^2}.$$

Define

$$R_{n11}^* = n^{-1/2} R_{n11} - n^{-1/2} \sum_{i=1}^n \frac{\xi_i (\delta_i - \pi(W_i)) B_i}{\rho(W_i)}.$$

Some straightforward calculation gives $E\{R_{n11}^*\} = O(n^{1/2} h^r) \rightarrow 0$, and $Var\{R_{n11}^*\} = O(nh^{2r} + (nh^2)^{-1}) \rightarrow 0$, which imply that

$$R_{n11} = \sum_{i=1}^n \frac{\xi_i (\delta_i - \pi_i(W_i)) B_i}{\rho(W_i)} + o_p(n^{1/2}).$$

Similarly, we have $R_{n12} = o_p(n^{1/2})$, and thus

$$\sum_{i=1}^n (\widehat{\pi}(W_i) - \pi(W_i)) B_i = \sum_{i=1}^n \frac{\xi_i (\delta_i - \pi_i(W_i)) B_i}{\rho(W_i)} + o_p(n^{1/2}). \tag{A.7}$$

In a similar manner, we obtain

$$\sum_{i=1}^n \frac{\xi_i}{\rho(W_i)} (\widehat{\pi}(W_i) - \pi(W_i)) B_i = \sum_{i=1}^n \frac{\xi_i(\delta_i - \pi_i(W_i))B_i}{\rho(W_i)} + o_p(n^{1/2}). \tag{A.8}$$

It follows from (A.7) and (A.8) that $R_{n1} = o_p(n^{1/2})$. Likewise, $R_{n2} = o_p(n^{1/2})$. Thus,

$$U(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) + \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1 \right) (\delta_i - \pi(W_i)) B_i + o_p(n^{1/2}). \tag{A.9}$$

The law of large numbers and the multivariate central limit theorem show that $n^{-1}U(\beta_0) \rightarrow 0$ in probability and $n^{-1/2}U(\beta_0)$ converges in distribution to a normal random variable with mean zero and variance matrix $\Sigma + \Sigma^*$. Note that

$$\widehat{\beta}_a - \beta_0 = - \frac{\partial U(\beta)}{\partial \beta} U(\beta_0),$$

and

$$-n^{-1} \frac{\partial U(\beta)}{\partial \beta} = n^{-1} \sum_{i=1}^n \int_0^\tau Y_i(t) \{Z_i - \bar{Z}^*(t)\}^{\otimes 2} dt \rightarrow A$$

almost surely by the uniform strong law of large numbers (Pollard 1990). Then it follows from (A.9) that $\widehat{\beta}_a$ is consistent and $n^{1/2}(\widehat{\beta}_a - \beta_0)$ is asymptotically normal with mean zero and covariance matrix $V = A^{-1}(\Sigma + \Sigma^*)A^{-1}$.

Proof of Theorem 3(ii)

First write

$$\widehat{\Lambda}_a(t) - \Lambda_0(t) = \int_0^t \frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)} - (\widehat{\beta}_a - \beta_0)' \int_0^t \bar{Z}^*(s) ds + \sum_{i=1}^n \left(\frac{\xi_i}{\widehat{\rho}(W_i)} - 1 \right) (\delta_i - \widehat{\pi}(W_i)) \int_0^t \frac{dN_i(s)}{\sum_{i=1}^n Y_i(s)}.$$

Note that

$$\sup_{0 \leq t \leq \tau} \left| n^{-1} \sum_{i=1}^n Y_i(t) - E[Y_1(t)] \right| = O_p(n^{-1/2}).$$

Following similar arguments as in the proof of (i), we obtain

$$\widehat{\Lambda}_a(t) - \Lambda_0(t) = n^{-1} \int_0^t \frac{\sum_{i=1}^n dM_i(s)}{E\{Y_i(s)\}} - d(t)'(\widehat{\beta}_a - \beta_0) + n^{-1} \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1 \right) (\delta_i - \pi(W_i)) \int_0^t \frac{dN_i(s)}{E\{Y_i(s)\}} + o_p(n^{-1/2}) \tag{A.10}$$

uniformly on $[0, \tau]$. In view of the consistency of $\widehat{\beta}_a$, it follows from the uniform strong law of large numbers and the multivariate central limit theorem that $\sup_{0 \leq t \leq \tau} |\widehat{\Lambda}_a(t) - \Lambda_0(t)| \rightarrow 0$ in probability, and $n^{1/2}\{\widehat{\Lambda}_a(t) - \Lambda_0(t)\}$ converges in finite dimensional distributions to a zero-mean Gaussian process. The first term on the right-hand side of (A.10) is tight as it is a martingale integral. The second term is tight because $n^{1/2}(\widehat{\beta}_a - \beta_0)$ converges in distribution and $d(t)$ is a deterministic function. Note that for each i , $(\xi_i \rho(W_i)^{-1} - 1) (\delta_i - \pi(W_i)) \int_0^t dN_i(s)/E\{Y_i(s)\}$ can be written as sums of monotone processes. Then the tightness of the third term follows from Example 2.11.16 of van der Vaart & Wellner (1996). Thus, $n^{1/2}\{\widehat{\Lambda}_a(t) - \Lambda_0(t)\}$ is tight and converges weakly to a zero-mean Gaussian process with covariance function $\Gamma(s, t)$ at (s, t) .

Outlined proof of Theorem 1

Note that $\widehat{\beta}$ is the solution to $U^*(\beta) = 0$, where

$$U^*(\beta) = \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\rho(W_i)} \{Z_i - \bar{Z}(t)\} [dN_i^u(t) - Y_i(t)\beta' Z_i dt].$$

Then it can be checked that

$$U^*(\beta_0) = U_1^*(\beta_0) + U_2^*(\beta_0), \tag{A.11}$$

where

$$U_1^*(\beta_0) = \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\rho(W_i)} \{Z_i - \bar{Z}(t)\} dM_i(t),$$

and

$$U_2^*(\beta_0) = \sum_{i=1}^n \int_0^\tau \xi_i \left(\frac{1}{\widehat{\rho}(W_i)} - \frac{1}{\rho(W_i)} \right) \{Z_i - \bar{Z}(t)\} dM_i(t).$$

Similarly to (A.2), we get

$$U_1^*(\beta_0) = \sum_{i=1}^n \int_0^\tau \frac{\xi_i}{\rho(W_i)} \{Z_i - \bar{z}(t)\} dM_i(t) + o_p(n^{1/2}). \tag{A.12}$$

From an argument similar to that in the proof of (A.7), we have

$$U_2^*(\beta_0) = - \sum_{i=1}^n \int_0^\tau \frac{\xi_i - \rho(W_i)}{\rho(W_i)} \{Z_i - \bar{z}(t)\} dM_i(t) + \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1 \right) (\delta_i - \pi(W_i)) B_i + o_p(n^{1/2}). \quad (\text{A.13})$$

It follows from (A.11)–(A.13) that

$$U^*(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) + \sum_{i=1}^n \left(\frac{\xi_i}{\rho(W_i)} - 1 \right) (\delta_i - \pi(W_i)) B_i + o_p(n^{1/2}),$$

which implies that $n^{-1/2}U^*(\beta_0)$ converges in distribution to a normal random variable with mean zero and variance matrix $\Sigma + \Sigma^*$. Then it follows from the Taylor expansion of $U^*(\hat{\beta})$ that $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero and covariance matrix $V = A^{-1}(\Sigma + \Sigma^*)A^{-1}$.

Table 1

Simulation results for one covariate with a missingness rate of 50%

| CR | Parameters Estimate | $\beta_0 = 1$ | | | $\beta_0 = 0.5$ | | | $\beta_0 = 0$ | | |
|-----|---------------------|---------------|--------|-------|-----------------|--------|-------|---------------|--------|-------|
| | | Bias | MSE | CP | Bias | MSE | CP | Bias | MSE | CP |
| 15% | Full | 0.0322 | 0.1304 | 0.952 | -0.0076 | 0.0816 | 0.948 | 0.0009 | 0.0506 | 0.960 |
| | SWE | -0.0205 | 0.1563 | 0.939 | -0.0033 | 0.1066 | 0.948 | -0.0168 | 0.0603 | 0.943 |
| | FAWE | 0.0336 | 0.1569 | 0.945 | -0.0049 | 0.0992 | 0.948 | 0.0020 | 0.0607 | 0.952 |
| | CC | -0.3630 | 0.3356 | 0.795 | -0.1729 | 0.1461 | 0.882 | -0.0865 | 0.0836 | 0.936 |
| 55% | Full | -0.0154 | 0.2061 | 0.956 | -0.0067 | 0.1459 | 0.944 | -0.0027 | 0.0914 | 0.954 |
| | SWE | 0.0030 | 0.3652 | 0.948 | 0.0138 | 0.2194 | 0.941 | 0.0093 | 0.1631 | 0.951 |
| | FAWE | 0.0042 | 0.3536 | 0.950 | 0.0056 | 0.2199 | 0.945 | -0.0010 | 0.1451 | 0.953 |
| | CC | -0.1918 | 0.4250 | 0.913 | -0.0788 | 0.2489 | 0.923 | -0.1122 | 0.1805 | 0.932 |
| 15% | Full | 0.0063 | 0.0810 | 0.958 | 0.0164 | 0.0565 | 0.961 | 0.0058 | 0.0359 | 0.953 |
| | SWE | 0.0133 | 0.1009 | 0.946 | -0.0094 | 0.0622 | 0.939 | 0.0049 | 0.0401 | 0.934 |
| | FAWE | 0.0049 | 0.0957 | 0.941 | 0.0187 | 0.0615 | 0.956 | 0.0039 | 0.0395 | 0.942 |
| | CC | -0.2304 | 0.1608 | 0.866 | -0.1510 | 0.0988 | 0.918 | -0.0962 | 0.0592 | 0.912 |
| 55% | Full | 0.0199 | 0.1436 | 0.948 | 0.0061 | 0.0845 | 0.964 | 0.0052 | 0.0666 | 0.947 |
| | SWE | -0.0052 | 0.2263 | 0.943 | -0.0094 | 0.1288 | 0.947 | 0.0050 | 0.1130 | 0.933 |
| | FAWE | 0.0195 | 0.2258 | 0.950 | -0.0034 | 0.1248 | 0.952 | 0.0002 | 0.1099 | 0.934 |
| | CC | -0.2371 | 0.2695 | 0.910 | -0.1839 | 0.1541 | 0.910 | 0.0624 | 0.1175 | 0.925 |

CR stands for censoring rate, Full stands for full data estimator, SWE stands for simple weighted estimator, FAWE stands for fully augmented weighted estimator, and CC stands for complete case estimator.

Table 2

Simulation results for a two-dimensional covariate vector with a missingness rate of 50%

| CR | Estimate | $\beta_{10} = 0$ | | | $\beta_{20} = 0$ | | |
|------------------|----------|------------------|--------|-------|------------------|--------|-------|
| | | Bias | MSE | CP | Bias | MSE | CP |
| 15% | Full | 0.0103 | 0.0369 | 0.947 | -0.0050 | 0.0122 | 0.957 |
| | SWE | -0.0129 | 0.0526 | 0.942 | 0.0092 | 0.0144 | 0.949 |
| | FAWE | 0.0091 | 0.0512 | 0.946 | -0.0077 | 0.0145 | 0.938 |
| | CC | -0.2222 | 0.0932 | 0.805 | 0.1091 | 0.0261 | 0.836 |
| 55% | Full | -0.0161 | 0.0640 | 0.958 | 0.0010 | 0.0212 | 0.956 |
| | SWE | -0.0352 | 0.1044 | 0.949 | 0.0048 | 0.0346 | 0.940 |
| | FAWE | -0.0231 | 0.1057 | 0.955 | 0.0014 | 0.0322 | 0.946 |
| | CC | -0.1290 | 0.1122 | 0.809 | 0.1701 | 0.0605 | 0.835 |
| $\beta_{10} = 1$ | | | | | | | |
| 15% | Full | -0.0001 | 0.0537 | 0.957 | -0.0053 | 0.0221 | 0.952 |
| | SWE | -0.0100 | 0.0784 | 0.957 | 0.0062 | 0.0297 | 0.940 |
| | FAWE | -0.0161 | 0.0805 | 0.950 | 0.0027 | 0.0302 | 0.955 |
| | CC | -0.0951 | 0.0950 | 0.902 | 0.0797 | 0.0460 | 0.891 |
| 55% | Full | 0.0113 | 0.0541 | 0.949 | -0.0078 | 0.0221 | 0.953 |
| | SWE | -0.0135 | 0.0796 | 0.961 | 0.0057 | 0.0309 | 0.933 |
| | FAWE | 0.0086 | 0.0810 | 0.948 | -0.0014 | 0.0335 | 0.955 |
| | CC | -0.0806 | 0.0930 | 0.908 | 0.0714 | 0.0450 | 0.887 |

CR stands for censoring rate, Full stands for full data estimator, SWE stands for simple weighted estimator, FAWE stands for fully augmented weighted estimator, and CC stands for complete case estimator.

Table 3

Comparison of the proposed method with the parametric approach of Lu and Liang (2008) under MAR and MCAR for a missingness rate of 50%

| Parameters | $\beta_0 = 0$ | | | | $\beta_0 = 1$ | | | |
|------------|---------------|--------|-------|-------|---------------|--------|--------|-------|
| | Bias | MSE | CP | CP | Bias | MSE | MSE | CP |
| MAR | | | | | | | | |
| Full | 0.0040 | 0.0448 | 0.948 | 0.948 | 0.0063 | 0.1071 | 0.1071 | 0.950 |
| SWE | 0.0018 | 0.0669 | 0.955 | 0.955 | 0.0099 | 0.1670 | 0.1670 | 0.945 |
| FAWE | 0.0051 | 0.0677 | 0.953 | 0.953 | 0.0011 | 0.1669 | 0.1669 | 0.947 |
| LIPW1 | 0.0077 | 0.0909 | 0.945 | 0.945 | 0.0013 | 0.2192 | 0.2192 | 0.950 |
| LDR1 | 0.0031 | 0.0675 | 0.950 | 0.950 | 0.0006 | 0.1683 | 0.1683 | 0.947 |
| LIPW2 | 0.6229 | 0.4311 | 0.139 | 0.139 | 1.1296 | 1.3814 | 1.3814 | 0.087 |
| LDR2 | 0.2121 | 0.0843 | 0.813 | 0.813 | 0.5999 | 0.4515 | 0.4515 | 0.471 |
| MCAR | | | | | | | | |
| Full | 0.0060 | 0.0447 | 0.961 | 0.961 | 0.0108 | 0.1079 | 0.1079 | 0.958 |
| SWE | 0.0089 | 0.0694 | 0.950 | 0.950 | 0.0080 | 0.1679 | 0.1679 | 0.954 |
| FAWE | 0.0104 | 0.0698 | 0.946 | 0.946 | 0.0083 | 0.1661 | 0.1661 | 0.950 |
| LIPW2 | 0.0004 | 0.0906 | 0.944 | 0.944 | 0.0298 | 0.2099 | 0.2099 | 0.942 |
| LDR2 | 0.0089 | 0.0686 | 0.946 | 0.946 | 0.0209 | 0.1686 | 0.1686 | 0.954 |
| CC | 0.0041 | 0.0932 | 0.955 | 0.955 | 0.0041 | 0.2191 | 0.2191 | 0.955 |

CR stands for censoring rate, Full stands for full data estimator, CC stands for complete case estimator, SWE stands for simple weighted estimator, FAWE stands for fully augmented weighted estimator, LIPW1 and LIPW2 stand for the inverse probability weighted estimators of Lu and Liang (2008) when using the logistic model and the constant model for $\rho^{(w)}$, respectively, LDR1 and LDR2 stand for the double robust estimators of Lu and Liang (2008) when using the logistic model and the constant model for $\rho^{(w)}$, respectively, and in all cases a constant model is used for $\pi^{(w)}$.

Table 4

Simulation results for the proposed method under MNAR

| MR | Estimate | Full | SWE | FAWE | CC |
|-----|----------|---------|--------|--------|---------|
| 20% | Bias | 0.0023 | 0.0103 | 0.0184 | -0.0119 |
| | MSE | 0.0385 | 0.0416 | 0.0424 | 0.0430 |
| | CP | 0.936 | 0.935 | 0.938 | 0.938 |
| 50% | Bias | -0.0012 | 0.0167 | 0.0562 | -0.0891 |
| | MSE | 0.0383 | 0.0434 | 0.0491 | 0.0716 |
| | CP | 0.938 | 0.936 | 0.930 | 0.926 |

MR stands for missingness rate, Full stands for full data estimator, SWE stands for simple weighted estimator, FAWE stands for fully augmented weighted estimator, and CC stands for complete case estimator.

Table 5

Estimates of regression coefficients for sex and age (in years), along with estimated standard errors and significance levels, from the analysis of time (in days) to non-ambulatory progression for patients in the brain cancer clinical trial.

| | CC | | SWE | | FAWE | |
|-------------------------|-------|-------|-------|-------|-------|-------|
| | sex | age | sex | age | sex | age |
| $\beta \times 10^4$ | 1.570 | 0.081 | 1.589 | 0.077 | 1.708 | 0.087 |
| $SE(\beta) \times 10^4$ | 2.445 | 0.090 | 2.306 | 0.037 | 2.241 | 0.034 |
| P-Value | 0.521 | 0.367 | 0.491 | 0.037 | 0.446 | 0.011 |

Note: The sample size is $n = 387$; the bandwidth vector is $h = (h_1, h_2)' = (34, 10)'$ CC denotes the complete case estimator; SWE denotes the simple weighted estimator; FAWE denotes the fully augmented weighted estimator.