Title

Weighted estimating equations for additive hazards models with missing covariates

Journal

Annals of the Institute of Statistical Mathematics, 71(2)

ISSN

0020-3157

Authors

Qi, Lihong

Zhang, Xu

Sun, Yanqing

et al.

Publication Date

2019-04-01

DOI

10.1007/s10463-018-0648-y

Peer reviewed

HHS Public Access

Author manuscript

Ann Inst Stat Math. Author manuscript; available in PMC 2020 April 01.

Published in final edited form as:

Ann Inst Stat Math. 2019 April ; 71(2): 365–387. doi:10.1007/s10463-018-0648-y.

Weighted Estimating Equations for Additive Hazards Models with Missing Covariates

Lihong Qi[1,*], **Xu Zhang**[2], **Yanqing Sun**[3], **Lu Wang**[4], **Yichuan Zhao**[5]

[1]Division of Biostatistics, Department of Public Health Sciences, The University of California Davis. One Shields Ave, MS1C, Davis, CA 95616

[2]Division of Clinical and Translational Sciences, Department of Internal Medicine, Medical School; Biostatistics/Epidemiology/Research Design Core, Center for Clinical and Translational Sciences, The University of Texas Health Science Center at Houston. 6410 Fannin Street, Houston, TX 77030

[3]Department of Mathematics and Statistics, The University of North Carolina at Charlotte. 9201 Univerity City Boulevard, Charlotte, NC 28223

[4]Department of Statistics, The University of California Davis. 399 Crocker Ln, Davis, CA 95616

[5]Department of Mathematics and Statistics, Room 1342, 25 Park Place Georgia State University. Atlanta, GA 30303

Abstract

This paper presents simple weighted and fully augmented weighted estimators for the additive hazards model with missing covariates when they are missing at random. The additive hazards model estimates the difference in hazards and has an intuitive biological interpretation. The proposed weighted estimators for the additive hazards model use incomplete data nonparametrically and have close-form expressions. We show that they are consistent and asymptotically normal, and are more efficient than the simple weighted estimator which only uses the complete data. We illustrate their finite-sample performance through simulation studies and an application to study the progression from mild cognitive impairment to dementia using data from the Alzheimer's Disease Neuroimaging Initiative as well as an application to the mouse leukemia study.

Keywords

kernel smoother; missing covariates; nonparametric method; weighted estimators; weighted estimating equations

[*]Corresponding author. lhqi@ucdavis.edu.

Supporting information. Additional information for this article is available online. The proofs of the theorems are in the Appendix S1 and the full acknowledgements section of the ADNI study are in the Appendix S2.

"For the Alzheimer's Disease Neuroimaging Initiative**"

The online version of this article contains supplementary material.

1 Introduction

For survival (time-to-event) data, a commonly used model is the Cox proportional hazards (PH) model (Cox, 1972) pertaining to the relative risk of certain covariates. Another well-known but less used method is the additive hazards model (Aalen, 1980; Cox and Oaks, 1984; Thomas, 1986; Breslow and Day, 1987, p.182). Unlike the Cox PH model, the additive hazards model does not assume proportional hazards, and it estimates the difference in hazards instead of the hazard ratios. Although the Cox PH model is very popular, it is desirable to utilize the additive hazards model for several reasons as discussed in Lin and Ying (1997). When describing the covariate-disease association, the hazard difference is complementary to and may be more relevant to public health than the hazard ratio because it translates directly into the number of events (e.g. disease cases) for the covariate. In practice, the additive hazards model may fit certain type of data better than the Cox PH model (Breslow and Day, 1987) and it provides a simple structure for studying frailty models and interval-censored data (Lin and Ying, 1997). Therefore, when the difference in disease risk due to the covariates is of primary interest or the PH assumption does not hold, the additive hazards model may be more proper.

The additive hazards model assumes that the conditional hazard function given a set of covariates is the sum of, rather than the product of, a baseline hazard function and a linear regression function of the covariates. Specifically, the hazard function for the failure time T associated with a column vector of covariates Z has the form

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

where $\lambda_0(t)$ is an unspecified baseline hazard function, and β is a column vector of regression parameters. The additive hazards model has an intuitive biological interpretation. When all covariates are fully observed, Lin and Ying (1994) proposed a simple semiparametric estimating function for β which generates a consistent and asymptotically normally distributed estimator with an explicit form.

Biomedical studies with survival outcomes frequently have missing covariates and some components of Z are not observed for all study subjects. Discarding the subjects with missing covariates may lead to either biased or inefficient estimates when the missing-data mechanism depends on the outcomes. Assuming missing at random (MAR) (Little and Rubin, 1987), i.e., the missing-data mechanism (or the selection probability) depends on the observed data but not on the missing data, Qi et al. (2005) proposed the simple weighted estimating equations using nonparametrically estimated selection probabilities and the kernel-assisted fully augmented weighted estimating equations for the Cox PH models. Their resultant fully augmented estimators (FAWEs) have the double-robustness property and also improve efficiency compared to most of the simple weighted estimators (SWEs); and the SWEs with selection probability estimated using all observed data are asymptotically equivalent to the FAWEs. For two-stage studies, Mark and Katki (2006) incorporated auxiliary information as weights and used model based approaches for estimating the sampling probabilities, which was further extended to general semi-

parametric models by several authors including Breslow et al. (2009) and Sun et al. (2017). For the additive hazards models, Kulich and Lin (2000) proposed a simple weighted estimator for case-cohort studies (Prentice, 1986), a special case of MAR where all cases (failures) and a subset of controls (censored subjects) are selected and have complete observations.

In this paper, we propose the simple weighted and fully augmented weighted estimating equations for the additive hazards models with missing covariates under the MAR assumption. An advantage is that the SWEs and FAWEs for the additive hazards model take explicit forms. We also propose to estimate the selection probabilities in the simple weighted and fully augmented weighted estimating equations, and the unknown conditional expectations in the fully augmented weighted estimating equations using the nonparametric kernel smoothing techniques similar to those used by Wang et al. (1997), Wang and Wang (2001) and Qi et al. (2005). Under certain regularity conditions, the resultant SWEs and FAWEs are consistent and asymptotical normal. We examine the finite sample performance of these estimators through a simulation study, and also demonstrate these methods using an Alzheimer's Disease Neuroimaging Initiative data set (adni.loni.usc.edu) and the data from the mouse leukemia study (Kalbfleisch and Prentice, 1980).

The remainder of the paper is organized as follows: Section 2 presents the SWEs and the FAWEs and their asymptotic properties as well as the discussions of the relationships between the SWEs and the FAWEs. Section 3 examines some properties of the proposed estimators through a simulation study, and illustrates the proposed methods with real examples using the Alzheimer's Disease Neuroimaging Initiative (ADNI) data and the mouse leukemia data. Section 4 provides further discussions and some practical recommendations. The regularity conditions are given in the Appendix and the proofs of the theorems are given in the online Appendix S1.

2 Method

2.1 Simple Weighted Estimating Equations

In the additive hazards model (1), let T, C and $X = min(T,C)$ be the failure, censoring and observed time for a subject, respectively. The failure indicator $\delta = I(T \leq C)$ is 1 if the subject experiences a failure and $\delta = 0$ if censored. Let Z denote a set of time-independent covariates. We assume that, given Z, T and C are independent, and all study subjects have X,δ observed. Suppose some elements of Z, denoted by Z^c, are observed for all n study subjects, while other elements may be missing for some subjects, denoted by Z^m. Let the selection indicator V equal 1 if Z^m is available, and 0 otherwise. Then the selection probability π is defined by $pr(V = 1 \mid X, \delta, Z^c, Z^m)$, equal to $pr(V = 1 \mid X, \delta, Z^c)$ under the MAR assumption.

Let $N(t) = \delta I(X \leq t)$ and $Y(t) = I(X \geq t)$ be the counting process and the at-risk process, respectively, corresponding to (X, δ). Let $(X_i, \delta_i, Z_i^m, Z_i^c, V_i)$, $i = 1,\ldots,n$, be independent and identically distributed copies of (X, δ, Z^m, Z^c, V). When the selection probability π is known, a simple weighted estimating function can have the following form:

$$U_{sw}(\beta,\pi)=\sum_{i=1}^{n}\frac{V_i}{\pi_i}\int_0^{\tau}\{Z_i-\bar{Z}_{sw}(\pi,t)\}\{dN_i(t)-Y_i(t)\beta^T Z_i dt\} \quad (2)$$

where $\bar{Z}_{sw}(\pi,t)=\sum_{j=1}^{n}(V_j/\pi_j)Y_j(t)Z_j/\sum_{l=1}^{n}(V_l/\pi_l)Y_l(t)$ and $\tau=sup\{t: pr(Y(t)=1)>0\}$. This estimating function can be regarded as a weighted complete-data pseudolikelihood score vector. By solving equation $U_{sw}(\beta,\pi)=0$, we obtain the estimator

$$\hat{\beta}_{sw}(\pi)=\left[\sum_{i=1}^{n}\frac{V_i}{\pi_i}\int_0^{\tau}\{Z_i-\bar{Z}_{sw}(\pi,t)\}^{\otimes 2}Y_i(t)dt\right]^{-1}\left[\sum_{i=1}^{n}\frac{V_i}{\pi_i}\int_0^{\tau}\{Z_i-\bar{Z}_{sw}(\pi,t)\}dN_i(t)\right]. \quad (3)$$

In a case-cohort study (Prentice, 1986) with Bernoulli sampling (Kulich & Lin, 2000), i.e., $V=1$ for all failure events (cases) and $V=1$ with probability π for censored observations (controls), the SWE is the same as the estimator proposed by Kulich & Lin (2000).

To study rigorously the asymptotic properties of $\hat{\beta}_{sw}(\pi)$, we impose regularity conditions (a1) to (a5) given in the Appendix. Since π is a function of X, δ, Z^c and may not be predictable, the techniques of Andersen and Gill (1982) cannot be directly applied to $U_{sw}(\beta,\pi)$. We employ modern empirical process theory to establish the consistency and asymptotic normality of $\hat{\beta}_{sw}(\pi)$, and further introduce the following notation to present the asymptotic results in Theorem 1. Define, for $k=0, 1$, $S_{sw}^{(k)}(\pi,t)=n^{-1}\sum_{j=1}^{n}(V_j/\pi_j)Z_j^{\otimes k}Y_j(t)$, $s^{(k)}(t)=E\{Z^{\otimes k}Y(t)\}$, $a^{\otimes 0}=I$, and $a^{\otimes 1}=a$. Let $M_i(t)=N_i(t)-\int_0^t Y_i(s)\lambda_0(s)ds-\int_0^t\beta^T Z_i Y_i(s)ds$ be the counting process martingale for the failure process, and $e(t)=s^{(1)}(t)/s^{(0)}(t)$. Let $M_{\tilde{Z}}=\int_0^{\tau}\{Z-e(t)\}dM(t)$ be the martingale transformation with mean $E\{M_{\tilde{Z}}\}=0$ and variance $\Sigma_A=E\{M_{\tilde{Z}}^{\otimes 2}\}$, where $a^{\otimes 2}=aa^T$.

Theorem 1 *Under the regularity conditions (a1) to (a5) given in the* Appendix, $\hat{\beta}_{sw}(\pi)$ *is consistent for the true parameter* β, *and* $n^{1/2}(\hat{\beta}_{sw}(\pi)-\beta)$ *converges to* $N(0,\Sigma^{-1}\Sigma_{sw}(\pi)\Sigma^{-1})$ *in distribution, where* $\Sigma=E\left[\int_0^{\tau}\{Z-e(t)\}^{\otimes 2}Y(t)dt\right]$ and $\Sigma_{sw}(\pi)=E\{\pi^{-1}M_{\tilde{Z}}^{\otimes 2}\}=\Sigma_A+E\{(\pi^{-1}-1)M_{\tilde{Z}}^{\otimes 2}\}$.

The first term Σ_A of $\Sigma_{sw}(\pi)$ is the asymptotic variance of the full-cohort pseudo-score estimator, while the second term quantifies the efficiency loss due to the missing covariates. The variance $\Sigma_{sw}(\pi)$ can be estimated consistently by $\hat{\Sigma}_{sw}(\pi)=n^{-1}\sum_{i=1}^{n}(V_i/\pi_i^{-2})\hat{M}_{\tilde{Z},i}^{\otimes 2}$, where

$$\widehat{M}_{\tilde{Z},i} = \int_0^\tau \left\{Z_i - \bar{Z}_{sw}(\pi,t)\right\}\left\{dN_i(t) - Y_i(t)Z^T\hat{\beta}_{sw}(\pi)dt - Y_i(t)d\hat{\Lambda}_0^{sw}(t)\right\}\hat{\Lambda}_0^{sw}(t) = \frac{1}{n}\int_0^t \frac{\sum_{i=1}^n \left(V_i/\pi_i\right)dN_i(s)}{S_{sw}^{(0)}(\pi,s)}$$

$$- \int_0^t \hat{\beta}_{sw}(\pi)^T \bar{Z}_{sw}(\pi,s)ds\,.$$

and Σ can be estimated consistently by $\hat{\Sigma}(\pi) = n^{-1}\sum_{i=1}^n \left(V_i/\pi_i\right)\int_0^\tau \left\{Z_i - \bar{Z}_{sw}(\pi,t)\right\}^{\otimes 2} Y_i(t)dt$.

When the true selection probability $\boldsymbol{\pi}$ is used, $\hat{\beta}_{sw}(\pi)$ is obtained excluding all incomplete observations, so it may not be efficient. To improve efficiency, an estimate of $\boldsymbol{\pi}$ can be used in the estimating function. We implement nonparametric methods to estimate $\boldsymbol{\pi}$ using all observed data, allowing incomplete observations to also contribute to the calculation of $\hat{\beta}_{sw}(\pi)$. Let W denote the variables used to estimate $\boldsymbol{\pi}$. When W is discrete, $\boldsymbol{\pi}$ can be estimated by the empirical proportion based on the observed data,

$$\hat{\pi}(w) = \frac{\sum_{i=1}^n V_i I\left(W_i = w\right)}{\sum_{j=1}^n I\left(W_j = w\right)}. \quad (4)$$

If W has d number of continuous components, $\boldsymbol{\pi}$ can be estimated consistently by nonparametric kernel smoothers. Let K be a rth-order ($r > d$) kernel function with bounded support, with $\int K(u)du = 1$, $\int u^m K(u)du = 0$ for $m = 1,\ldots,(r-1)$, $\int u^r K(u)du \neq 0$, and $\int K(u)^2\,du < \infty$. Let $K_h(\cdot) = K(\cdot/h)$, and h is the smoothing parameter, also called the bandwidth. We estimate $\boldsymbol{\pi}(w)$ using the Nadaraya-Watson (Nadaraya 1964; Watson 1964) estimator

$$\hat{\pi}(w) = \frac{\sum_{i=1}^n V_i K_h\left(w - W_i\right)}{\sum_{j=1}^n K_h\left(w - W_j\right)}. \quad (5)$$

The kernel function K usually has little effect on $\hat{\pi}$, and thus on the estimator of β, while h affects the behavior of the estimator both theoretically and practically. We assume h satisfies $nh^{2d} \to \infty$ and $nh^{2r} \to 0$, as $n \to \infty$. Similar to Qi et al. (2005) and Wang & Wang (2001), we may choose $h = O(n^{-1/p})$ for some integer $p > 2d$, and the smallest even integer for r such that $r \geq p - d$. For example, when $d = 2$, p and r can take values of 5 and 4, respectively.

Plugging in $\hat{\pi}$ in the simple weighted estimating function (2) results in the following estimating function:

$$U_{sw}(\beta,\hat{\pi}) = \sum_{i=1}^n \frac{V_i}{\hat{\pi}_i}\int_0^\tau \left\{Z_i - \bar{Z}_{sw}(\hat{\pi},t)\right\}\left\{dN_i(t) - Y_i(t)\beta^T Z_i dt\right\}.$$

Solving $U_{sw}(\beta, \hat{\pi}) = 0$, we obtain $\hat{\beta}_{sw}(\hat{\pi})$ with the same closed form of (3) replacing π with $\hat{\pi}$. The consistency and asymptotic normality of $\hat{\beta}_{sw}(\hat{\pi})$ were established in the online supplementary material, with more imposed regularity conditions (a6) to (a10) in the Appendix, and the results are stated in the following Theorem 2.

Theorem 2 *Under the regularity conditions (a1) to (a10), given in the* Appendix, $\hat{\beta}_{sw}(\hat{\pi})$ *is consistent for the true parameter* β, *and* $n^{1/2}\left(\hat{\beta}_{sw}(\hat{\pi}) - \beta\right)$ *converges to* $N\left(0, \Sigma^{-1}\Sigma_{sw}^{\hat{\pi}}(\pi)\Sigma^{-1}\right)$ *in probability, with* $\Sigma_{sw}^{\hat{\pi}}(\pi) = \Sigma_A + \Sigma_{sw}^*(\pi)$, *where*

$$\Sigma_{sw}^*(\pi) = E\left[\left(\pi^{-1} - 1\right)\left\{M_{\tilde{Z}} - E\left(M_{\tilde{Z}}|W\right)\right\}^{\otimes 2}\right] = E\left\{\left(\pi^{-1} - 1\right)var\left(M_{\tilde{Z}}|W\right)\right\}.$$

Theorem 2 shows that with the consistent estimator $\hat{\pi}$ in (4) or (5), $\hat{\beta}_{sw}(\hat{\pi})$ is consistent for β, and has a smaller asymptotic variance than $\hat{\beta}_{sw}(\pi)$ because their asymptotic variances have the same first term Σ_A and the second term in $\sum_{sw}(\pi)$ is $E\left\{\left(\pi^{-1} - 1\right)M_{\tilde{Z}}^{\otimes 2}\right\}$, greater than $E\left[\left(\pi^{-1} - 1\right)\left\{M_{\tilde{Z}} - E\left(M_{\tilde{Z}}|W\right)\right\}^{\otimes 2}\right]$, the second term in $\Sigma_{sw}^{\hat{\pi}}(\pi)$. Hence using the nonparametrically estimated selection probability in the simple weighted estimating equation allows more effective use of the available data and improves the efficiency of the SWE with π.

Also $var\left(M_{\tilde{Z}}|W\right)$ is non-increasing over the dimension of W, so the more variables are used to estimate π, the smaller is $var\left(M_{\tilde{Z}}|W\right)$ and the more efficient is $\hat{\beta}_{sw}(\hat{\pi})$. This suggests that estimating selection probabilities using additional variables besides the variables on which they depend may lead to further efficiency gains for the SWEs. The variance $\sum_{sw}^{\hat{\pi}}(\pi)$ can be estimated consistently by estimating Σ_A and $\Sigma_{sw}^*(\pi)$ consistently. Let $M_{\tilde{Z}}^0 = E\left(M_{\tilde{Z}}|W\right)$. Estimate $d\Lambda_0(t)$ and $M_{\tilde{Z}}$, respectively, by

$$d\hat{\Lambda}_0^{sw}(\hat{\pi}, t) = \frac{1}{n}\frac{\sum_{i=1}^n \left(V_i/\hat{\pi}_i\right)\left\{dN_i(t) - Y_i(t)\hat{\beta}_{sw}^T(\hat{\pi})Z_i dt\right\}}{S_{sw}^{(0)}(\hat{\pi}, t)},$$

$$\widehat{M}_{\tilde{Z}} = \int_0^\tau \left\{Z - \bar{Z}_{sw}(\hat{\pi}, t)\right\}\left\{dN(t) - Y(t)d\hat{\Lambda}_0^{sw}(t, \hat{\pi}) - \hat{\beta}_{sw}(\hat{\pi})^T ZY(t)dt\right\}.$$

We estimate Σ_A by

$$\hat{\Sigma}_A = n^{-1}\sum_{i=1}^n \frac{V_i}{\hat{\pi}_i}\int_0^\tau \left\{Z_i - \bar{Z}_{sw}(\hat{\pi}, t)\right\}^{\otimes 2} dN_i(t)$$

and $\Sigma_{sw}^*(\pi)$ by

$$\hat{\Sigma}^{*}_{sw}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \frac{V_i(1-\hat{\pi}_i)}{\hat{\pi}_i^2}\left(\hat{M}_{\tilde{Z},i} - \hat{M}^0_{\tilde{Z},i}\right)^{\otimes 2}, \quad (6)$$

where $\hat{M}^0_{\tilde{Z},i}$ is obtained by the Nadaraya-Watson estimator (12) of Section 2.2.

2.2 Fully Augmented Weighted Estimating Equations

We propose the following fully augmented estimating function for the additive hazards model,

$$U_{faw}(\beta,\pi) = \sum_{i=1}^{n} \frac{V_i}{\pi_i}\int_0^{\tau}\left\{Z_i - \bar{Z}_{faw}(\pi,E,t)\right\}\left\{dN_i(t) - \beta^T Z_i Y_i(t)dt\right\} + \sum_{i=1}^{n} A_i^F(\beta,\pi), \quad (7)$$

where $\bar{Z}_{faw}(\pi,E,t) = S^{(1)}_{faw}(\pi,E,t)/S^{(0)}_{faw}(\pi,E,t)$, and let $W = (X, \delta, Z^c)$,

$$A_i^F(\beta,\pi) = \left(1-\frac{V_i}{\pi_i}\right)\int_0^{\tau}\left(E\left[Z_i\left\{dN_i(t) - \beta^T Z_i Y_i(t)dt\right\}\middle| W_i\right] - \bar{Z}_{faw}(\pi,E,t)E\left[\left\{dN_i(t) - \beta^T Z_i Y_i(t)dt\right\}\middle| W_i\right]\right),$$

and for $k = 0, 1$,

$$S^{(k)}_{faw}(\pi,E,t) = \frac{1}{n}\sum_{j=1}^{n}\frac{V_j}{\pi_j}Y_j(t)Z_j^{\otimes k} + \frac{1}{n}\sum_{j=1}^{n}\left(1-\frac{V_j}{\pi_j}\right)Y_j(t)E\left(Z_j^{\otimes k}\middle| W_j\right). \quad (8)$$

When $k = 0$, $S^{(0)}_{faw}(\pi,E,t) = \sum_{j=1}^{n} Y_j(t)/n$.

The fully augmented weighted estimating function $U_{faw}(\beta, \pi)$ uses incomplete observations through the augmented averages $S^{(1)}_{faw}(\pi,E,t)$ and $S^{(0)}_{faw}(\pi,E,t)$, and the augmentation term $A_i^F(\beta,\pi)$. The resulting fully augmented weighted estimator (FAWE) possesses the so-called double-robust property, ie., the estimator is consistent if either the missing-data mechanism (i.e. the selection probability) or the distribution of the missing covariates given the observed data is modeled correctly (Wang & Chen 2001, Qi et al. 2005). Solving $U_{faw}(\beta, \pi) = 0$, we can obtain the FAWE explicitly:

$$\hat{\beta}_{faw}(\pi,E) = \left\{\sum_{i=1}^{n}\left[\frac{V_i}{\pi_i}\int_0^{\tau} Y_i(t)\bar{Z}_i^{\otimes 2}dt + \left(1-\frac{V_i}{\pi_i}\right)\int_0^{\tau} Y_i(t)E\left(\bar{Z}_i^{\otimes 2}\middle|\ \widetilde{W}\right)dt\right]\right\}^{-1}\sum_{i=1}^{n} \quad (9)$$
$$\left[\frac{V_i}{\pi_i}\int_0^{\tau}\bar{Z}_i dN_i(t) + \left(1-\frac{V_i}{\pi_i}\right)\int_0^{\tau} E\left(\bar{Z}_i\middle|\ \widetilde{W}\right)dN_i(t)\right],$$

where $\bar{Z}_i = Z_i - \bar{Z}_{faw}(\pi, E, t)$ and $\widetilde{W} = (W_1, \ldots, W_n)$.

The conditional expectations $E(\bar{Z}_i^{\otimes 2} | \widetilde{W})$ and $E(\bar{Z}_i | \widetilde{W})$ in equation (9) are unknown since they contain unknown quantities $E(Z_i^{\otimes 2} | W_i)$ and $E(Z_i | W_i)$. Specifically, $E(\bar{Z}_i^{\otimes 2} | \widetilde{W}) = E(Z_i^{\otimes 2} | W_i) - 2E(Z_i | W_i)\bar{Z}_{faw}(\pi, E, t)^T + \bar{Z}_{faw}(\pi, E, t)^{\otimes 2}$, and $E(\bar{Z}_i | \widetilde{W}) = E(Z_i | W_i) - \bar{Z}_{faw}(\pi, E, t)$. The denominator of $\bar{Z}_{faw}(\pi, E, t)$, $S^{(0)}_{faw}(\pi, E, t)$ does not contain unknown quantities, while the nominator of $\bar{Z}_{faw}(\pi, E, t)$, $S^{(1)}_{faw}(\pi, E, t)$ involves $E(Z_i | W_i)$, $i = 1, \ldots, n$, as seen in equation (8). So once $E(Z_i^{\otimes 2} | W_i)$ and $E(Z_i | W_i)$ are estimated, we can estimate $E(\bar{Z}_i^{\otimes 2} | \widetilde{W})$ and $E(\bar{Z}_i | \widetilde{W})$ as well as $\bar{Z}_{faw}(\pi, E, t)$ in equation (9), and generate a FAWE $\hat{\beta}_{faw}(\pi, \hat{E})$. Furthermore, since $Z = (Z^m, Z^c)^T$ and Z^c is known for all subjects, we have

$$E(Z | W) = \begin{pmatrix} E(Z^m | W) \\ Z^c \end{pmatrix}, \quad (10)$$

and

$$E(Z^{\otimes 2} | W) = \begin{pmatrix} E\left[(Z^m)^{\otimes 2} | W\right] & E(Z^m | W)(Z^c)^T \\ Z^c(E(Z^m | W))^T & (Z^c)^{\otimes 2} \end{pmatrix}. \quad (11)$$

Hence we can use the observed values for Z^c and only need to estimate the unknown quantities $E[(Z^m)^{\otimes 2} | W]$ and $E(Z^m | W)$ in equations (10) and (11). We propose to use a nonparametric kernel-assisted method to estimate $E[(Z^m)^{\otimes 2} | W]$ and $E(Z^m | W)$. Specifically, let ζ denote Z^m or $(Z^m)^{\otimes 2}$, then we can estimate $E(\zeta | W)$ using the Nadaraya-Watson (Nadaraya 1964; Watson 1964) estimator based on the complete observations.

Let $\phi(w) = E(\zeta | w)$. Assuming $\phi(w)$ is a smooth function with r continuous and bounded partial derivatives with respect to the continuous components of W a.e., then a Nadaraya-Watson estimator of $\phi(w)$ is

$$\hat{\phi}(w) = \frac{\sum_{i=1}^n V_i \zeta_i K_h(w - W_i)}{\sum_{j=1}^n V_j K_h(w - W_j)}, \quad (12)$$

where K is an rth-order kernel function as defined in Section 2.1 and h is a smoothing parameter.

A step by step algorithm for obtaining the FAWEs using equation (9) can be summarized as below:

1. Apply the kernel smoother (12) to obtain the estimated $E[(Z^m)^{\otimes 2}|W)]$ and $E(Z^m|W)$, i.e., $\hat{E}[(Z^m)^{\otimes 2}|W)]$ and $\hat{E}(Z^m|W)$, respectively. When there are both discrete and continuous components in W, first stratify by the discrete components, then at each level, implement the kernel smoothers.

2. Calculate $S^{(1)}_{faw}(\pi, E, t)$ using equation (8). For the augmented term $\sum_{j=1}^{n}\left(1 - V_j/\pi_j\right)Y_j(t)E\left(Z_j|W_j\right)/n$, plug in the observed values of Z^c and $\hat{E}(Z^m|W)$. Calculate $S^{(0)}_{faw}(\pi, E, t) = \sum_{j=1}^{n} Y_j(t)/n$ and obtain $\bar{Z}_{faw}(\pi, E, t) = S^{(1)}_{faw}(\pi, E, t)/S^{(0)}_{faw}(\pi, E, t)$.

3. Obtain an FAWE $\hat{\beta}_{faw}(\pi, \hat{E})$ using equation (9). For the augmented terms in both the nominator and the denominator, plug in the observed values of Z^c, and $\hat{E}[(Z^m)^{\otimes 2}|W]$, $\hat{E}(Z^m|W)$ and the estimated $\bar{Z}_{faw}(\pi, E, t)$.

When the selection probability π is unknown, $\hat{\pi}(w)$ in (5) with $w = (x, \delta, z^c)$ can substitute for π in equation (7), resulting in the FAWE $\hat{\beta}_{faw}(\hat{\pi}, \hat{E})$. When both selection probabilities and conditional expectations are estimated nonparametrically, different kernel functions may be used respectively. For simplicity, we used the same kernel function for theoretical derivations and simulation studies. The following theorems present the asymptotical properties of $\hat{\beta}_{faw}(\pi, E)$, $\hat{\beta}_{faw}(\hat{\pi}, E)$, $\hat{\beta}_{faw}(\pi, \hat{E})$ and $\hat{\beta}_{faw}(\hat{\pi}, \hat{E})$.

Theorem 3 *Under the regularity conditions (a1) to (a5) given in the* Appendix, $\hat{\beta}_{faw}(\pi, E)$ *is consistent for the true parameter* β, *and* $n^{1/2}\left(\hat{\beta}_{faw}(\pi, E) - \beta\right)$ *converges to* $N(0, \Sigma^{-1}\Sigma_{faw}(\pi)\Sigma^{-1})$ *in distribution, with* $\Sigma_{faw}(\pi) = \Sigma_A + \Sigma^*_{faw}(\pi)$, where $\Sigma^*_{faw}(\pi) = E\left\{\left(\pi^{-1} - 1\right)var\left(M_{\tilde{Z}}|W\right)\right\}$.

Theorem 4 *Under the regularity conditions (a1) to (a10) given in the* Appendix, $\hat{\beta}_{faw}(\hat{\pi}, E)$, $\hat{\beta}_{faw}(\pi, \hat{E})$, *and* $\hat{\beta}_{faw}(\hat{\pi}, \hat{E})$ *are consistent for the true parameter* β. *Moreover,* $n^{1/2}\left(\hat{\beta}_{faw}(\hat{\pi}, E) - \beta\right)$, $n^{1/2}\left(\hat{\beta}_{faw}(\pi, \hat{E}) - \beta\right)$ *and* $n^{1/2}\left(\hat{\beta}_{faw}(\hat{\pi}, \hat{E}) - \beta\right)$ *are asymptotically normal with mean 0 and variance matrix* $\Sigma^{-1}\Sigma_{faw}(\pi)\Sigma^{-1}$.

All the FAWEs have the same asymptotic distribution, indicating that the asymptotic properties of the FAWEs are not affected by the nonparametric estimation of the selection probabilities and the conditional expectations. The FAWEs are more efficient than the SWE with true π and the SWEs with $\hat{\pi}$ except that the SWE with nonparametric $\hat{\pi}(X, \delta, Z^c)$ has the same asymptotic distribution as all the FAWEs. When Z^m can be exactly specified by $W = (X, \delta, Z^c)$, then the martingale transformation $M_{\tilde{Z}}$ is constant given W, so $var\left(M_{\tilde{Z}}|W\right) = 0$, and $\Sigma_{faw}(\pi) = \Sigma^{\hat{\pi}}_{sw}(\pi) = \Sigma_A$. Therefore the SWE with $\hat{\pi}(X, \delta, Z^c)$ and the kernel-assisted FAWEs achieve the efficiency of the estimator based on the full cohort data in this special situation.

Consistent estimators of $\Sigma_{faw}(\pi)$ can be obtained similarly to those of the $\sum_{sw}^{\hat{\pi}}(\pi)$ For illustration, we demonstrate how to estimate the variance $\Sigma_{faw}(\pi)$ for $\hat{\beta}_{faw}(\hat{\pi}, \hat{E})$. Set

$$d\hat{\Lambda}_0^{faw}(t, \hat{\pi}) = \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{j=1}^{n} Y_j(t)} - \hat{\beta}_{faw}(\hat{\pi}, \hat{E})^T \bar{Z}_{faw}(\hat{\pi}, \hat{E}, t)dt$$

and

$$\hat{M}_{\tilde{Z}} = \int_0^{\tau} \left\{ Z - \bar{Z}_{faw}(\hat{\pi}, \hat{E}, t)dt \right\} \left\{ dN(t) - Y(t)\hat{\beta}_{faw}^T(\hat{\pi}, \hat{E})Zdt - Y_i(t)d\hat{\Lambda}_0^{faw}(t, \hat{\pi}) \right\}$$

to be the estimators of $d\Lambda_0(t)$ and $M_{\tilde{Z}}$, respectively. Then Σ_A and $\Sigma_{faw}^*(\pi)$ are estimated respectively by

$$\hat{\Sigma}_A = n^{-1} \sum_{i=1}^{n} \frac{V_i}{\hat{\pi}_i} \int_0^{\tau} \left\{ Z_i - \bar{Z}_{\text{faw}}(\hat{\pi}, \hat{E}, t) \right\}^{\otimes 2} dN_i(t)$$

and

$$\hat{\Sigma}_{faw}^*(\pi) = \frac{1}{n} \sum_{i=1} \frac{V_i(1 - \hat{\pi}_i)}{\hat{\pi}_i^2} \left(\hat{M}_{\tilde{Z}, i} - \hat{M}_{\tilde{Z}, i}^0 \right)^{\otimes 2},$$

where $\hat{M}_{\tilde{Z}, i}^0$ is obtained using the Nadaraya-Watson estimator in (12).

3 Numerical Studies

3.1 Simulation Study

A comprehensive simulation study was conducted to examine the moderate sample size performance of the SWEs and the kernel-assisted FAWEs and to compare their performance with that of the full-cohort (i.e. Lin & Ying's estimator, 1994) and complete-case analyses. In all simulations, 1000 datasets were generated and either $n = 250$ or 500 subjects were used. In the first three simulation settings, we considered two independent covariates, a missing covariate Z^m and an observed covariate Z^c. In the first setting, a binary variable Z^c was generated from the Bernoulli distribution with probability 0.5, and Z^m followed a standard normal distribution. We generated the failure time using $\lambda(t; Z^m, Z^c) = 1.5 - 0.5Z^m + 1.0Z^c$, and the censoring time based on the exponential distribution with mean 1.5, resulting in about 45% censored observations. The first setting mimicked the case-cohort (Prentice, 1986) sampling scheme used in Kulich & Lin (2000), with $\pi(\delta) = \delta + 0.5(1 - \delta)$, so that we can compare our estimators with theirs.

In the second setting, $\pi(\delta) = 0.7\delta + 0.5(1 - \delta)$, resulting in an overall missing rate of 39%. We considered correlated Z^m and Z^c in this setting. Specifically, $Z^c \sim$ Bernoulli(0.5), $Z^m =$

$Z^c - 0.5 + \epsilon$, where $\epsilon \sim N(0, 1)$, and the correlation between Z^m and Z^c is about 0.44. In the third setting, both Z^m and Z^c were generated from a Bernoulli distribution with probability 0.5. The hazard function $\lambda(t; Z^m, Z^c) = 2t + 0.8Z^m + 0.4Z^c$ was used to generate the failure time and the censoring time was generated from the exponential distribution with mean 1, yielding about 44% censored observations. The selection probability was $\pi(\delta, X) = 1/\{1+\exp(1.5 - 2.5\delta - X)\}$, and about 44% observations had missing Z^m. In the fourth setting, we considered two missing covariates, $Z^{m1} \sim$ N(0,1), $Z^{m2} \sim$ Bernoulli(0.5), $Z^c \sim$ Bernoulli(0.5), $\beta = (0.5, 1.0, -0.5)$ and the baseline hazard function equal to 1.0. The censoring time was generated from exponential distribution with mean 0.5, yielding about 49% censored observations. The selection probability $\pi(\delta) = 0.6\delta + 0.4(1 - \delta)$, resulting in 50% observations with Z^{m1} and Z^{m2} missing.

The theoretical standard errors for the SWEs, the kernel-assisted FAWEs were obtained from the corresponding variance estimators discussed in Sections 2. The conditional expectations in the FAWEs were estimated by the Nadaraya-Watson estimators with the smoothing parameter $h = 4\sigma_W n^{-1/3}$, where σ_W was the standard deviation of observed times stratified by δ and Z^c.

We obtain the following measures for all the estimators. Bias is the average difference between a parameter and its estimate. Relative bias is the result of bias divided by the true β value. Sample standard error (SE) equals the square root of the sample variance of the 1000 parameter estimates. Mean theoretical standard error (SE) is the average of the one thousand standard error estimates. And 95% confidence interval coverage probabilities (CP) were calculated using the theoretical standard error estimates.

Table 1 presents the results from the first simulation setting. All estimated selection probabilities are consistent for the true π. There is no evidence of bias in any of the estimates $\hat{\beta}$ except for the complete-case analysis, due to the strong association between the selection probability and the outcome variable δ. All FAWEs had smaller bias than other estimators for both β_1 and β_2. The sample standard errors are generally in good agreement with the corresponding mean theoretical standard errors. As the cohort size increased from 250 to 500, the sample and mean theoretical standard errors for $\hat{\beta}_1$ became closer for all the weighted estimators, and the standard errors of SWE ($\hat{\pi}(X, \delta, Z^c)$) became closer to those of the FAWEs.

In addition, the estimator of Kulich & Lin (2000) and SWE ($\pi(\delta)$) have the same point estimates because they share the same estimating equations. All weighted estimators have similar standard errors for $\hat{\beta}_1$, indicating that compared to the SWE with true π, using partially incomplete data in estimation did not improve the efficiency of estimates of β_1. However, compared to SWE($\pi(\delta)$) the sample standard error for $\hat{\beta}_2$ was reduced by using the SWEs with $\hat{\pi}(\delta, Z^c)$ (e.g. about 9% when n = 250) and $\hat{\pi}(X, \delta, Z^c)$ (about 11% when n = 250) and as well as all the kernel-assisted FAWEs (about 16% when n = 250). So the efficiency of $\hat{\beta}_2$ was improved by including the incomplete data in estimation. Especially, all kernel-assisted FAWEs almost achieved the full-cohort efficiency for estimating β_2. For $n =$

250, the ratios of sample and theoretical standard errors between the full-cohort estimator and the FAWEs were 96%. When cohort size was 500, the ratio was about 98% if calculated from sample standard errors, and about 97% if from theoretical standard errors.

The results from the second simulation setting, where Z^m and Z^c are correlated, show similar patterns as those from the first setting and are provided in Table 2. The bias of complete-case method was smaller in this setting because the missing data mechanism was less dependent on the censoring indicator δ (selection probability was 70% for cases and 50% for controls).

Table 3 displays the results from the third simulation setting where both Z^m and Z^c were binary and the selection probability depended on both the censoring indicator and the survival time. The results from this setting show similar patterns as those from the first two settings, except that the sample and mean theoretical standard errors of SWE($\hat{\pi}(\delta)$) are not in good agreement with each other and its 95% confidence interval coverage probability is low when the sample size was 250. This occurred because the $\hat{\pi}$ based on δ alone was inconsistent for the true $\pi(\delta, X)$. These issues are reduced when the sample size increases to 500. The FAWE($\hat{\pi}(\delta)$), however, does not have these issues and still performs well under both sample sizes. Because the selection probability depended heavily on both survival time and censoring indicator, the bias for estimates from the complete-case analysis was elevated to about 15% of the true parameter value.

The results from the fourth setting with a mixture of continuous and binary missing covariates, Z^{m1} and Z^{m2}, and a binary Z^c, showed similar patterns as those from the first three settings (Table 4).

In summary, the results from the simulation studies suggest that (1) the kernel-assisted FAWEs and most of the SWEs with nonparametric $\hat{\pi}$ are more efficient than the SWE with true π; (2) the SWEs with $\hat{\pi}$ are not as efficient as the kernel-assisted FAWEs most of the time, and the efficiency of SWE($\hat{\pi}(X, \delta, Z^c)$), the most efficient among all SWEs, approaches to those of the FAWEs when sample size increases; (3) the complete-case analysis generates inconsistent estimates when true π depends on outcome variables; (4) all FAWEs and the SWEs correct such bias with true π or consistent $\hat{\pi}$; (5) an inconsistent $\hat{\pi}$ may affect the variance estimation of the SWEs but not of the kernel-assisted FAWEs.

3.2 Application to the Alzheimers Disease Neuroimaging Initiative Data

We illustrate our methods using a data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, a large depository of clinical and demographic data of Alzheimer's disease (AD) patients, as well as their longitudinal outcome and imaging measurements (adni.loni.usc.edu). The ADNI, led by Dr. Michael W. Weiner at VA Medical Center and University of California San Francisco, was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, pharmaceutical companies and non-profit organizations as a public-private partnership. The ADNI had several goals including AD pathophysiology investigation, diagnostic tool improvement and biomarker development.

Our example involves the progression from mild cognitive impairment (MCI) to dementia data from the first phase of ADNI (ADNI-1), where 800 patients were enrolled from over 50 sites across the U.S. and Canada in 2005–2007. Patients were advised to have office visit very six months for a period of two years, followed by a visit after one-year interval. Totally 382 patients had MCI diagnosis at the initial visit and had at least one follow-up visit. The middle point between two visits was used as the time-to-event, as in conventional analyses of follow-up survival outcomes in practice. Among the 382 patients, 159 developed dementia and 223 were censored. The median follow-up time for patients without dementia diagnosis at subsequent visits was 36 months. Baseline covariates were obtained during the initial visit to address various questions. Here we consider the association between the risk of developing dementia from MCI and two covariates: the APOE-e4 status, a binary variable with 1 indicating the presence of APOE-e4 and 0 otherwise, and Abeta, a biomarker with continuous expression levels. All patients had APOE-e4 status while the Abeta expression levels were available only for 192 patients who consented to get the lumbar punctures done. The standardized Abeta expression level was included in the analysis and each unit of the standardized variable represented 50 in the original expression level. We applied the complete-case analysis, the SWEs and the kernel-assisted FAWEs with estimated selection probabilities to the data. Selection probabilities were estimated based on censoring indicator δ only, observed time X and δ, and X, δ and APOE-e4 status for the SWEs and the FAWEs. When including X to estimate the selection probabilities, the Nadaraya-Watson estimator was used with the band width $h = 4\sigma_W n^{-1/3}$. Conditional expectations were estimated by the Nadaraya-Watson estimator (12) based on X, δ and the APOE-e4 status.

Results in Table 5 show that the Abeta level was significantly associated with the risk of dementia after adjusting for APOE-e4 status. For patients with the same APOE-e4 status, a high Abeta expression level was associated with a lower risk of dementia. However, APOE-e4, was not statistically significant regardless of the methods used.

3.3 Application to the Mouse Leukemia Study

We also illustrate our methods using the data set from the mouse leukemia study (Kalbfleisch and Prentice, 1980). This study was conducted in the laboratories of Dr. Robert Nowinski of the Fred Hutchinson Cancer Research Center, Seattle, Washington, investigating genetic and viral factors in the development of spontaneous leukemia in mice. Totally 204 mice were followed for 2 years for mortality due to thymic or nonthymic leukemia, or other natural causes. Two covariates, the Gpd-1 phenotype and the level of endogenous murine leukemia virus, were of interest. Almost all mice had the level of endogenous murine leukemia virus measured. The Gpd-1 phenotype was obtained for 100 mice that survived 400 days, indicating whether the Gpd-1 phenotype was observed on a mouse depends on its follow-up time. The MAR assumption seems proper here since the missingness was caused by design.

Following previous publications (e.g. Wang and Chen, 2001, and Qi et. al. 2005), we excluded the animals with missing endogenous murine leukemia virus for computational simplicity. A total of 175 mice were analyzed in the data analysis. The virus level was classified into two categories, with $Z^c = 0$ if a virus level $< 10^4$ PFU/ml and 1 otherwise. We

conducted separate analyses for the death of thymic leukemia and the death of thymic or nonthymic leukemia as the endpoint, respectively. We obtained the estimates of regression coefficients in the AFT model using the complete case analysis, the SWEs and the kernel-assisted FAWEs. To estimate the selection probabilities for the SWEs and the FAWEs, we applied the Nadaraya-Watson estimator in (5) with bandwidth $h = 4\sigma_W n^{-1/3}$. Conditional expectations were estimated by the Nadaraya-Watson estimator (12).

Results in Table 6 show that only the complete case analysis indicated a significant association between the Gpd-1 phenotype and the death of thymic leukemia adjusting for the virus load, while for the death of thymic or nonthymic leukemia, none of the methods had significant results. For the virus load, all methods except for the complete case analysis resulted in significant associations with the death of thymic leukemia and with the death of thymic or nonthymic leukemia, respectively. The FAWEs sometimes have slightly smaller SE than the SWEs, especially for the coefficient of the observed covariate, virus load. This trend is consistent with what we have seen in the simulation results.

4 Discussion

Missing covariates complicate analysis of survival data. Inconsistent and inefficient estimates can be generated by naively discarding subjects with missing covariates. The additive hazards model is a useful alternative to the commonly used Cox PH model, especially when the primary interest is to estimate the difference in disease risk for the covariates or when the proportional hazards assumption is violated. Assuming the missingness is MAR, we proposed the SWEs and kernel-assisted FAWEs for the additive hazards model. By using the nonparametric smoothing techniques, the proposed SWEs and FAWEs are robust against model misspecifications for the selection probability and the conditional expectation of missing covariates, which is an advantage over the existing methods in the literature. The proposed weighted estimators are consistent and asymptotically normal and can improve the efficiency of the estimates from the SWE with true $\boldsymbol{\pi}$ as well as from the complete-data analysis. All the weighted estimators possess an explicit expression, an advantage of using the additive hazards model over the Cox PH model.

The proposed SWEs and FAWEs expand the SWE of Kulich & Lin (2000) for the case-cohort studies (Prentice, 1986) to general missing-data mechanisms. The asymptotic distribution theory of the SWEs with nonparametric $\hat{\pi}$ suggests that the more variables are used in obtaining $\hat{\pi}$, the more efficiency may be gained potentially. So the SWE with $\hat{\pi}(X, \delta, Z^c)$ has the best efficiency among all the SWEs. Although the SWE with $\hat{\pi}(X, \delta, Z^c)$ has the same asymptotic distribution as the kernel-assisted FAWEs, our simulation studies and the data analysis example suggest that the FAWEs tend to perform better than the SWE with $\hat{\pi}(X, \delta, Z^c)$ when sample sizes are moderate. In addition, the FAWEs are robust toward misspecifications of the selection probabilities due to their double robustness property.

The proposed methods can utilize surrogate variables to predict the missing covariates and the selection probability for increased efficiency. In this case, the surrogate variables should considered as a part of the observed data under the MAR and included as elements in W.

These weighted methods can be applied to situations where missing covariates occur by happenstance or by design, such as two-phase studies where selection probabilities are known. For two-phase studies with moderate sample sizes, we suggest the use of the FAWE with true selection probabilities. When the sample sizes are large and missingness rates are not extreme, one can employ either the SWE with $\hat{\pi}(X, \delta, Z^c)$ or the kernel-assisted FAWE with π, and the estimators will generate consistent estimates with similar efficiency.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The authors wish to thank the editor, associate editor and the referee for their constructive comments and suggestions which have greatly improved the paper. The authors thank the ADNI study and database for allowing us to use their data (the full acknowledgements section of the ADNI study are in the online supplementary material). Many thanks to Dr. Danielle Harvey for her valuable help with obtaining the ADNI data set and suggestions regarding to the data analysis. Thanks to Wei Ran, Yueheng An, Yiming Hu, and Nan Bi for their help. Dr. Yichuan Zhao was partially supported by the NSF grant DMS-1406163 and NSA grant H98230-12-1-0209. Dr. Yanqing Sun was partially supported by the National Science Foundation grant DMS-1208978, DMS-1513072, and the National Institute of Health NIAID grant R37 AI054165.

Appendix

The following regularity conditions are needed in the proofs of Theorems 1 to 4.

(a1) $\Lambda_0(\tau) < \infty$.

(a2) $P\{Y(\tau) = 1\} > 0$.

(a3) Z is time-independent and bounded.

(a4) The matrix $\Sigma_A = E\left[\int_0^\tau \{Z - e(t)\}^{\otimes 2} dN(t)\right]$ is positive definite.

(a5) W has bounded support $\mathscr{W}$. There exists a constant $\pi_0 > 0$ such that $\pi(w) > \pi_0$ for $w \in \mathscr{W}$.

(a6) The selection probability $\pi(w)$ has r continuous and bounded partial derivatives with respect to the continuous components of W a.e.

(a7) The probability density/mass function $f(w)$ of w and the conditional probability density/mass function $f_{W|V}(w)$ of $W|V$ have r continuous and bounded partial derivatives with respect to the continuous components of W a.s.

(a8) Conditional distributions $f_{W|V=0}(w)$ and $f_{W|V=1}(w)$ have the same support, and $c(w) = f_{W|V=0}(w)/f_{W|V=1}(w)$ is bounded over the support.

(a9) The conditional expectations $E(Z^k | W = w)$, $E\{(Z^k)^{\otimes 2} | W = w\}$, $k = 0,1$, have r continuous and bounded partial derivatives with respect to the continuous components of W a.e.

(a10) $nh^{2d} \to \infty$ and $nh^{2r} \to 0$, as $n \to \infty$.

References

Aalen OO (1980). A model for nonparametric regression analysis of counting processes In Lecture Notes in Statistics, 2 Ed Klonecki N, Kosek A & Rosinski J, pp. 1–25. New York: Springer.

Andersen PK and Gill RD (1982) Cox's regression model for counting processes: A large sample study. Annals of Statistics, 10, 1100–1120.

Breslow NE and Day NE (1987) Statistical Methods in Cancer Research Vol II The Design and Analysis of Cohort Studies. Lyon: IARC.

Breslow NE, Lumley T, Ballantyne CM, Chambless LE, and Kulich M (2009). Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: Applications in epidemiology. Statistics in Biosciences, 1, 32–49. [PubMed: 20174455]

Cox DR (1972). Regression models and life tables (with Discussion). Journal of the Royal Statistical Society: Series B, 34, 187–220.

Cox DR and Oaks D (1984). Analysis of Survival Data. London: Chapman and Hall.

Kalbfleisch JD and Prentice RL (1980). The Statistical Analysis of Failure Time Data. New York: Wiley.

Kulich M and Lin DY (2000). Additive hazards regression for case-cohort studies. Biometrika, 87, 73–87.

Lin DY and Ying Z (1994). Semiparametric analysis of the additive risk model. Biometrika, 81, 61–71.

Lin DY and Ying Z (1997). Additive regression models for survival data In: Lin DY, Fleming TR(eds). In Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis, 185–198. Springer: New York.

Little RJA and Rubin DB (1987). Statistical Analysis with Missing Data. New York: Wiley.

Mark SD and Katki HA (2006). Specifying and implementing nonparametric and semiparametric survival estimators in two-Stage (nested) cohort studies with missing case data. Journal of the American Statistical Association, 101, 460–471.

Nadaraya EA (1964). On estimating regression. Theory of Probability and Its Applications, 9, 141–142.

Prentice RL (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika, 73, 1–11.

Qi L, Wang CY, and Prentice RL (2005). Weighted estimators for proportional hazards regression with missing covariates. Journal of the American Statistical Association, 472, 1250–1263.

Sun Y, Qian X, Shou Q and Gilbert P (2017). Analysis of two-phase sampling data with semiparametric additive hazards models. Lifetime Data Analysis, 23, 377–399. [PubMed: 26995733]

Thomas DC (1986). Use of auxiliary information in fitting nonproportional hazards models Modern Statistical Methods in Chronic Disease Epidemiology. Ed. Moolgavkar SH and Prentice RL, pp. 197–210. New York: Wiley.

Wang CY, Chen HY (2001) Augmented inverse probability weighted estimator for Cox missing covariate regression. Biometrics, 57(2), 414–419. [PubMed: 11414564]

Wang CY, Wang SJ, Zhao LP, and Ou ST (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. Journal of the American Statistical Association, 92, 512–525.

Wang S and Wang CY (2001). A note on kernel assisted estimators in missing covariate regression. Statistics & Probability Letters, 55, 439–449.

Watson GS (1964). Smooth regression analysis. Sankhyā A, 26, 359–372.

Table 1.

Simulation comparison of various estimators of the additive hazards model parameter $\beta = (-0{:}5;\ 1{:}0)$.

		Bias		Relative bias		Sample SE		Theoretical SE		95% CP	
n	**Method**	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
250	Full cohort	−0.010	0.021	−0.019	0.021	0.169	0.340	0.168	0.344	0.948	0.955
	Complete case	−0.059	0.110	−0.118	0.110	0.219	0.446	0.222	0.444	0.945	0.949
	Kulich & Lin	−0.009	0.026	−0.017	0.026	0.200	0.421	0.201	0.414	0.949	0.950
	SWE - $\pi(\delta)^a$	−0.009	0.026	−0.017	0.026	0.200	0.421	0.195	0.406	0.941	0.941
	SWE - $\hat{\pi}(\delta)^b$	−0.009	0.026	−0.019	0.026	0.200	0.421	0.201	0.414	0.952	0.949
	SWE - $\hat{\pi}(\delta, Z^c)$	−0.009	0.026	−0.019	0.026	0.200	0.383	0.201	0.383	0.954	0.948
	SWE - $\hat{\pi}(\delta, X)$	−0.004	0.021	−0.008	0.021	0.205	0.427	0.199	0.407	0.947	0.941
	SWE - $\hat{\pi}(\delta, X, Z^c)$	−0.005	0.059	−0.009	0.059	0.206	0.375	0.197	0.366	0.943	0.947
	FAWE - $\pi(\delta)^a$	−0.003	0.009	0.006	0.009	0.210	0.356	0.199	0.360	0.943	0.961
	FAWE - $\hat{\pi}(\delta)^b$	−0.001	0.007	0.002	0.007	0.210	0.356	0.199	0.360	0.942	0.963
	FAWE - $\hat{\pi}(\delta, Z^c)$	−0.002	0.008	0.004	0.008	0.210	0.356	0.200	0.360	0.942	0.962
	FAWE - $\hat{\pi}(\delta, X)$	−0.004	0.008	0.008	0.008	0.210	0.355	0.200	0.361	0.946	0.962
	FAWE - $\hat{\pi}(\delta, X, Z^c)$	−0.005	0.008	0.010	0.008	0.213	0.355	0.200	0.361	0.933	0.961
500	Full cohort	−0.001	0.008	−0.002	0.008	0.114	0.238	0.115	0.240	0.956	0.947
	Complete case	−0.049	0.104	−0.099	0.104	0.154	0.304	0.152	0.307	0.939	0.948
	Kulich & Lin	−0.001	0.016	−0.003	0.016	0.138	0.285	0.139	0.289	0.945	0.957
	SWE - $\pi(\delta)^a$	−0.001	0.016	−0.003	0.016	0.138	0.285	0.137	0.286	0.945	0.954
	SWE - $\hat{\pi}(\delta)^b$	−0.001	0.016	−0.003	0.016	0.137	0.285	0.139	0.289	0.945	0.955
	SWE - $\hat{\pi}(\delta, Z^c)$	−0.002	0.015	−0.003	0.015	0.138	0.267	0.139	0.267	0.945	0.948
	SWE - $\hat{\pi}(\delta, X)$	0.000	0.014	0.000	0.014	0.140	0.288	0.138	0.285	0.943	0.951
	SWE - $\hat{\pi}(\delta, X, Z^c)$	−0.000	0.036	−0.000	0.036	0.138	0.259	0.136	0.250	0.940	0.934
	FAWE - $\pi(\delta)^a$	0.002	0.003	−0.005	0.003	0.141	0.244	0.137	0.248	0.936	0.950

		Bias		Relative bias		Sample SE		Theoretical SE		95% CP	
n	Method	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
	FAWE - $\hat{\pi}(\delta)$[b]	0.003	0.001	−0.007	0.001	0.141	0.245	0.137	0.248	0.936	0.950
	FAWE - $\hat{\pi}(\delta, Z^C)$	0.002	0.002	−0.005	0.002	0.142	0.245	0.137	0.248	0.938	0.950
	FAWE - $\hat{\pi}(\delta, X)$	−0.001	0.002	0.001	0.002	0.142	0.245	0.138	0.248	0.942	0.950
	FAWE - $\hat{\pi}(\delta, X, Z^C)$	−0.001	0.002	0.002	0.002	0.141	0.245	0.138	0.248	0.945	0.948

The baseline hazard was a constant, $\lambda_0(t) = 1.5$ and $Z^M \sim N(0, 1)$, $Z^C \sim$ Bernoulli(0.5). The censoring time was generated from exponential distribution. Censoring rate 45% and selection probability $\pi(\delta) = \delta + 0.5(1 - \delta)$, a mocked case-cohort sampling.

[a] True π was used.

[b] $\hat{\pi}$ was estimated nonparametrically based on the variable in the bracket. Other $\hat{\pi}$ was obtained on the variables in the bracket using the Nadaraya-Watson estimator with uniform kernel and bandwidth $h = 4\sigma_W n^{-1/3}$.

Table 2.

Simulation comparison of various estimators of the additive hazards model parameter $\beta = (-0.5, 1.0)$. The baseline hazard was a constant, $\lambda_0(t) = 1.5$ and $Z^c \sim$ Bernoulli(0.5); $Z^m = Z^c - 0.5 + \epsilon$, where $\epsilon \sim N(0, 1)$, and the correlation between Z^m and Z^c is about 0.44. The censoring time was generated from exponential distribution. Censoring rate 45% and selection probability $\pi(\delta) = 0.7\delta + 0.5(1 - \delta)$.

		Bias		Relative bias		Sample SE		Theoretical SE		95% CP	
n	**Method**	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
n=250	Full cohort	−0.012	0.017	0.024	0.017	0.170	0.348	0.167	0.342	0.949	0.948
	Complete case	−0.050	0.080	0.100	0.080	0.242	0.497	0.236	0.475	0.948	0.941
	SWE - $\pi(\delta)^a$	−0.019	0.025	0.038	0.025	0.227	0.467	0.212	0.442	0.930	0.943
	SWE - $\hat{\pi}(\delta)^b$	−0.019	0.026	0.037	0.026	0.225	0.467	0.218	0.448	0.946	0.946
	SWE - $\hat{\pi}(\delta, Z^c)$	−0.020	0.017	0.039	0.017	0.226	0.429	0.218	0.409	0.946	0.956
	SWE - $\hat{\pi}(\delta, X)$	−0.019	0.018	0.038	0.018	0.231	0.472	0.216	0.440	0.929	0.935
	SWE - $\hat{\pi}(\delta, X, Z^c)$	−0.019	0.061	0.038	0.061	0.229	0.400	0.215	0.376	0.932	0.940
	FAWE - $\pi(\delta)^a$	−0.020	0.007	0.040	0.007	0.230	0.363	0.215	0.367	0.935	0.961
	FAWE - $\hat{\pi}(\delta)^b$	−0.021	0.008	0.041	0.008	0.230	0.364	0.215	0.367	0.935	0.962
	FAWE - $\hat{\pi}(\delta, Z^c)$	−0.020	0.008	0.041	0.008	0.230	0.364	0.216	0.368	0.935	0.962
	FAWE - $\hat{\pi}(\delta, X)$	−0.021	0.007	0.041	0.007	0.232	0.363	0.216	0.368	0.932	0.959
	FAWE - $\hat{\pi}(\delta, X, Z^c)$	−0.020	0.007	0.039	0.007	0.231	0.362	0.216	0.368	0.928	0.961
n=500	Full cohort	−0.010	−0.007	0.019	−0.007	0.114	0.235	0.116	0.239	0.952	0.952
	Complete case	−0.047	0.059	0.093	0.059	0.161	0.331	0.162	0.329	0.953	0.959
	SWE - $\pi(\delta)^a$	−0.016	0.004	0.032	0.004	0.152	0.316	0.148	0.309	0.945	0.951
	SWE - $\hat{\pi}(\delta)^b$	−0.016	0.005	0.032	0.005	0.151	0.315	0.151	0.311	0.954	0.952
	SWE - $\hat{\pi}(\delta, Z^c)$	−0.016	0.003	0.032	0.003	0.151	0.287	0.151	0.284	0.950	0.948
	SWE - $\hat{\pi}(\delta, X)$	−0.015	0.003	0.029	0.003	0.155	0.323	0.150	0.308	0.943	0.952
	SWE - $\hat{\pi}(\delta, X, Z^c)$	−0.014	0.026	0.027	0.026	0.155	0.265	0.149	0.259	0.934	0.945
	FAWE - $\pi(\delta)^a$	−0.015	−0.013	0.030	−0.013	0.155	0.246	0.148	0.254	0.937	0.959

		Bias		Relative bias		Sample SE		Theoretical SE		95% CP	
n	**Method**	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
	FAWE - $\hat{\pi}(\delta)$[b]	−0.015	−0.013	0.030	−0.013	0.155	0.246	0.148	0.254	0.941	0.959
	FAWE - $\hat{\pi}(\delta, Z^c)$	−0.015	−0.013	0.030	−0.013	0.155	0.246	0.149	0.254	0.940	0.959
	FAWE - $\hat{\pi}(\delta, X)$	−0.015	−0.013	0.030	−0.013	0.157	0.245	0.149	0.255	0.941	0.959
	FAWE - $\hat{\pi}(\delta, X, Z^c)$	−0.015	−0.013	0.030	−0.013	0.157	0.245	0.149	0.255	0.937	0.959

[a] True π was used.

[b] $\hat{\pi}$ was estimated nonparametrically based on the variable in the bracket. Other $\hat{\pi}$ was obtained on the variables in the bracket using the Nadaraya-Watson estimator with uniform kernel and bandwidth $h = 4\sigma_W n^{-1/3}$.

Table 3.

Simulation comparison of various estimators of the additive hazards model parameter $\beta = (0.8, 0.4)$. The baseline hazard was a linear function of time, $\lambda_0(t) = 2t$ and $Z^m \sim$ Bernoulli(0.5), $Z^c \sim$ Bernoulli(0.5). The censoring time was generated from exponential distribution. Censoring rate 44% and selection probability $\pi(\delta, X) = 1/\{1 + \exp(1.5 - 2{:}5\delta - X\}$.

		Bias		Relative bias		Sample SE		Theoretical SE		95% CP	
n	**Method**	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
250	Full cohort	0.010	0.003	0.013	0.008	0.218	0.223	0.226	0.218	0.957	0.951
	Complete case	0.125	0.059	0.156	0.149	0.315	0.314	0.314	0.305	0.942	0.946
	SWE - $\pi(\delta,X)^a$	0.027	0.012	0.034	0.008	0.303	0.303	0.292	0.281	0.940	0.925
	SWE - $\hat{\pi}(\delta)^b$	−0.016	−0.009	−0.020	−0.006	0.320	0.312	0.279	0.267	0.908	0.906
	SWE - $\hat{\pi}(\delta, X)$	0.023	0.007	0.029	0.005	0.307	0.304	0.287	0.277	0.936	0.921
	SWE - $\hat{\pi}(\delta, X, Z^c)$	0.023	0.024	0.029	0.017	0.301	0.256	0.282	0.235	0.935	0.938
	FAWE - $\pi(\delta,X)^a$	0.026	0.016	0.032	0.012	0.301	0.247	0.285	0.233	0.938	0.943
	FAWE - $\hat{\pi}(\delta)^b$	0.029	0.016	0.037	0.011	0.308	0.247	0.317	0.238	0.946	0.947
	FAWE - $\hat{\pi}(\delta, X)$	0.028	0.016	0.035	0.012	0.306	0.247	0.286	0.234	0.937	0.941
	FAWE - $\hat{\pi}(\delta, X, Z^c)$	0.027	0.016	0.034	0.012	0.307	0.247	0.283	0.233	0.938	0.942
500	Full cohort	0.011	0.004	0.014	0.011	0.153	0.149	0.159	0.153	0.956	0.969
	Complete case	0.120	0.063	0.149	0.157	0.216	0.211	0.219	0.213	0.929	0.945
	SWE - $\pi(\delta,X)^a$	0.017	0.009	0.021	0.007	0.208	0.201	0.208	0.200	0.955	0.950
	SWE - $\hat{\pi}(\delta)^b$	−0.032	−0.016	−0.040	−0.012	0.219	0.210	0.203	0.194	0.922	0.932
	SWE - $\hat{\pi}(\delta, X)$	0.012	0.008	0.015	0.006	0.212	0.208	0.204	0.197	0.936	0.939
	SWE - $\hat{\pi}(\delta, X, Z^c)$	0.011	0.015	0.014	0.011	0.211	0.167	0.201	0.163	0.937	0.953
	FAWE - $\pi(\delta,X)^a$	0.014	0.011	0.018	0.008	0.212	0.160	0.201	0.163	0.942	0.961
	FAWE - $\hat{\pi}(\delta)^b$	0.016	0.010	0.019	0.007	0.216	0.160	0.230	0.167	0.964	0.963
	FAWE - $\hat{\pi}(\delta, X)$	0.014	0.010	0.018	0.007	0.214	0.160	0.203	0.163	0.934	0.964
	FAWE - $\hat{\pi}(\delta, X, Z^c)$	0.015	0.010	0.018	0.007	0.215	0.160	0.202	0.163	0.934	0.961

[a] True π was used.

[b] $\hat{\pi}$ was estimated nonparametrically based on the variable in the bracket. Other $\hat{\pi}$ was obtained on the variables in the bracket using the Nadaraya-Watson estimator with uniform kernel and bandwidth $h = 4\sigma_W n^{-1/3}$.

Table 4.

Simulation comparison of various estimators of the additive hazards model parameter $\beta = (0.5, 1.0, -0.5)$. The baseline hazard was a linear function of time, $\lambda_0(t) = 1.0$ and $Z_1^m \sim N(0, 1)$, $Z_2^m \sim$ Bernoulli (0.5), $Z^c \sim$ Bernoulli(0.5). The censoring time was generated from exponential distribution. Censoring rate 49% and selection probability $\pi(\delta) = 0.6\delta + 0.4(1 - \delta)$.

		Bias			Relative bias			Sample SE			Theoretical SE			95% CP		
n	**Method**	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
500	Full cohort	−0.001	0.009	0.003	−0.003	0.009	−0.006	0.138	0.280	0.269	0.136	0.279	0.271	0.943	0.944	0.950
	Complete case	0.044	0.096	−0.052	0.087	0.096	0.103	0.220	0.433	0.411	0.216	0.434	0.426	0.938	0.943	0.960
	SWE - $\pi(\delta,X,Z^c)$[a]	0.003	0.016	−0.014	0.006	0.016	0.028	0.202	0.403	0.381	0.192	0.396	0.386	0.939	0.948	0.957
	SWE - $\hat{\pi}(\delta)$[b]	0.003	0.018	−0.015	0.006	0.018	0.030	0.201	0.403	0.383	0.190	0.387	0.378	0.938	0.945	0.948
	SWE - $\hat{\pi}(\delta, X)$	0.000	0.010	−0.009	0.001	0.010	0.017	0.204	0.410	0.442	0.194	0.393	0.389	0.940	0.943	0.919
	SWE - $\hat{\pi}(\delta, X, Z^c)$	0.004	0.008	−0.030	0.008	0.008	0.059	0.208	0.410	0.318	0.193	0.391	0.306	0.928	0.941	0.945
	FAWE - $\pi(\delta,X,Z^c)$[a]	0.005	0.010	−0.026	0.010	0.010	0.051	0.208	0.413	0.307	0.232	0.471	0.326	0.969	0.974	0.967
	FAWE - $\hat{\pi}(\delta)$[b]	0.005	0.010	−0.026	0.010	0.010	0.051	0.208	0.413	0.307	0.194	0.392	0.304	0.926	0.939	0.949
	FAWE - $\hat{\pi}(\delta, X)$	0.005	0.012	−0.025	0.011	0.012	0.051	0.210	0.417	0.307	0.194	0.394	0.305	0.927	0.945	0.946
	FAWE - $\hat{\pi}(\delta, X, Z^c)$	0.005	0.011	−0.026	0.011	0.011	0.051	0.211	0.418	0.308	0.193	0.391	0.305	0.922	0.939	0.949

[a] True π was used.

[b] $\hat{\pi}$ was estimated nonparametrically based on the variable in the bracket. Other $\hat{\pi}$ was obtained on the variables in the bracket using the Nadaraya-Watson estimator with uniform kernel and bandwidth $h = 4\sigma_W n^{-1/3}$.

Table 5.

Analysis of the ANDI dementia data using the additive hazards model and various estimators.

	Abeta			APOE-e4		
Method	**Estimate**	**SE**	***p*–value**	**Estimate**	**SE**	***p*–value**
Complete-case	−0.0055	0.0016	0.0009	0.0035	0.0047	0.46
SWE - $\hat{\pi}(\delta)$	−0.0052	0.0016	0.0009	0.0044	0.0041	0.28
SWE - $\hat{\pi}(\delta, X)$	−0.0054	0.0012	< 0.0001	0.0037	0.0034	0.27
SWE - $\hat{\pi}\left(\delta, X, Z^c\right)$	−0.0052	0.0012	< 0.0001	0.0063	0.0041	0.13
FAWE - $\hat{\pi}(\delta)$	−0.0053	0.0015	0.0004	0.0057	0.0037	0.12
FAWE - $\hat{\pi}(\delta, X)$	−0.0052	0.0016	0.0011	0.0058	0.0039	0.14
FAWE - $\hat{\pi}\left(\delta, X, Z^c\right)$	−0.0053	0.0016	0.0008	0.0057	0.0039	0.15

$\hat{\pi}$ was obtained on the variables in the bracket using the Nadaraya-Watson estimator with uniform kernel and bandwidth $h = 4\sigma_W n^{-1/3}$.

Table 6.

Analysis of the Mouse Leukemia data using the additive hazards model and various estimators.

		Gpd-1			Virus load	
Method	**Estimate**	**SE**	***p*–value**	**Estimate**	**SE**	***p*–value**
Thymic leukemia death						
Complete-case	−0.0167	0.0069	0.0153	0.0072	0.0047	0.1288
SWE - $\hat{\pi}(\delta, X)$	−0.0145	0.0081	0.0723	0.0173	0.0068	0.0107
SWE - $\hat{\pi}(\delta, X, Z^C)$	−0.0149	0.0078	0.0578	0.0207	0.0063	0.0010
FAWE - $\hat{\pi}(\delta, X)$	−0.0144	0.0084	0.0857	0.0235	0.0065	0.0003
FAWE - $\hat{\pi}(\delta, X, Z^C)$	−0.0150	0.0080	0.0597	0.0233	0.0063	0.0002
Thymic or nonthymic leukemia death						
Complete-case	−0.0144	0.0075	0.0561	0.0107	0.0061	0.0781
SWE - $\hat{\pi}(\delta, X)$	−0.0137	0.0086	0.1090	0.0230	0.0074	0.0018
SWE - $\hat{\pi}(\delta, X, Z^C)$	−0.0131	0.0084	0.1170	0.0233	0.0070	0.0009
FAWE - $\hat{\pi}(\delta, X)$	−0.0148	0.0088	0.0925	0.0262	0.0073	0.0003
FAWE - $\hat{\pi}(\delta, X, Z^C)$	−0.0152	0.0085	0.0743	0.0259	0.0072	0.0003

$\hat{\pi}$ was obtained on the variables in the bracket using the Nadaraya-Watson estimator with uniform kernel and bandwidth $h = 4\sigma_W n^{-1/3}$.