Analysis of the time-varying Cox model for the cause-specific hazard functions with missing causes

Fei Heng, Yanqing Sun, Seunggeun Hyun & Peter B. Gilbert

Lifetime Data Analysis
An International Journal Devoted to Statistical Methods and Applications for Time-to-Event Data

ISSN 1380-7870

Lifetime Data Anal
DOI 10.1007/s10985-020-09497-y

ONLINE FIRST

Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Springer

Author's personal copy

Lifetime Data Analysis
https://doi.org/10.1007/s10985-020-09497-y

Analysis of the time-varying Cox model for the cause-specific hazard functions with missing causes

Fei Heng[1] · Yanqing Sun[2] · Seunggeun Hyun[3] · Peter B. Gilbert[4]

Received: 22 October 2018 / Accepted: 27 March 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

This paper studies the Cox model with time-varying coefficients for cause-specific hazard functions when the causes of failure are subject to missingness. Inverse probability weighted and augmented inverse probability weighted estimators are investigated. The latter is considered as a two-stage estimator by directly utilizing the inverse probability weighted estimator and through modeling available auxiliary variables to improve efficiency. The asymptotic properties of the two estimators are investigated. Hypothesis testing procedures are developed to test the null hypotheses that the covariate effects are zero and that the covariate effects are constant. We conduct simulation studies to examine the finite sample properties of the proposed estimation and hypothesis testing procedures under various settings of the auxiliary variables and the percentages of the failure causes that are missing. These simulation results demonstrate that the augmented inverse probability weighted estimators are more efficient than the inverse probability weighted estimators and that the proposed testing procedures have the expected satisfactory results in sizes and powers. The proposed methods are illustrated using the Mashi clinical trial data for investigating the effect of randomization to formula-feeding versus breastfeeding plus extended infant zidovudine prophylaxis on death due to mother-to-child HIV transmission in Botswana.

Keywords Augmented inverse probability weighted estimator · Auxiliary variables · Cause-specific hazard function · Competing risks model · Hypothesis testing procedures · Missing causes of failure · Inverse probability weighted estimator · Cox model with time-dependent coefficients · Two-stage augmented inverse probability weighted estimator

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10985-020-09497-y) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Published online: 09 April 2020

Author's personal copy

1 Introduction

In the analysis of competing risks data, the cause of failure is often missing. For example, the cause of death of a patient cannot be determined due to the lack of necessary medical diagnostic information. Several methods have been proposed to deal with the problem of missing causes of failure. Under the cause-specific Cox proportional hazards model framework, Goetghebeur and Ryan (1995) presented an approach based on partial likelihood, an approach also studied by Lu and Tsiatis (2005). Lu and Tsiatis (2001) used multiple imputation of missing causes of failure in the Cox model. Gao and Tsiatis (2005) developed inverse probability weighting (IPW) and augmented IPW (AIPW) approaches under linear transformation models, while Lu and Liang (2008) applied these techniques for the additive hazards model. Hyun et al. (2012) developed the IPW and AIPW estimators for the Cox model, while Liu et al. (2018) applied IPW in the Cox model to investigate conditional risks of colorectal cancer subtypes by incorporating a biomarker associated with disease subtype in the weight to reduce selection bias. Nevo et al. (2018) investigated an informative likelihood approach in the Cox model, which includes modeling of the conditional distribution of auxiliary variables that are observed for all cases. The literature, including citations here, has focused on modeling time-constant covariate effects.

While time-varying effects are common in failure time data analysis in practice, few have investigated the problem with missing causes of failure. This work is motivated by the Mashi study in Botswana (Thior et al. 2006), for investigating the effect of randomization to formula-feeding (FF) versus breastfeeding plus extended infant zidovudine prophylaxis (BF+AZT) on death due to mother-to-child HIV transmission. The Mashi study shows that formula-feeding increases the risk of death before 7 months of age compared with the breastfeeding plus AZT strategy, and this effect diminishes later on. However, the effect of randomized feeding strategy on death related to HIV infection has not been thoroughly studied. In addition, the causes of death are missing for 61 infants of the 111 live-born infants who died in the Mashi study. In this article we study the Cox model with time-varying coefficients for cause-specific hazard functions, allowing some failure events to be missing the cause of failure. This model assuming complete data on failure causes was investigated by Sun et al. (2008), with application to an oral cholera vaccine efficacy trial in Bangladesh (Clemens et al. 1990), where the cause of failure was the type of infecting cholera strain. Extending these methods to account for missing causes of failure requires solving additional theoretical and computational challenges. Our developed methods are applied to assess the cause-specific treatment effects on HIV-related death and on HIV-unrelated death in the Mashi study.

In situations where only a single cause of failure is of interest, several studies, for example Zucker and Karr (1990), Murphy and Sen (1991), Martinussen et al. (2002), Cai and Sun (2003), Tian et al. (2005), and Sun et al. (2009), have been carried out on the Cox model with time-varying regression coefficients. Cai and Sun (2003) applied a local partial likelihood estimation technique to estimate the time-dependent coefficients. Tian et al. (2005) further studied the local constant partial likelihood estimator and constructed pointwise and simultaneous confidence intervals for the

Author's personal copy

regression parameters. Sun et al. (2009) investigated empirical likelihood inferences for the model.

Modifying the score function of the local linear partial maximum likelihood estimator of Cai and Sun (2003), we propose a two-stage procedure to estimate the cause-specific time-varying Cox model with missing causes. In the first stage, an inverse probability weighted (IPW) complete-case estimator is developed. The augmented inverse probability weighted (AIPW) estimator is constructed by directly utilizing the IPW estimator (termed the first-stage estimator) in the augmentation term of the IPW estimating equation. The proposed two-stage AIPW estimating equation has the same structure as the AIPW approaches proposed in Robins et al. (1994). Normative/regular AIPW approaches such as Gao and Tsiatis (2005) and Lu and Liang (2008) model both the probability of a complete case and the conditional distribution of the cause of failure (depending on auxiliary variables) through parametric models such as logistic regressions. We show that the conditional distribution of the cause of failure depends on the failure time data and the auxiliary variables through the conditional cause-specific hazard functions and the conditional distribution of the auxiliary variables, a connection that is used to derive the proposed two-stage AIPW estimator, a key difference from the previously developed AIPW estimators. The proposed two-stage AIPW method can utilize the auxiliary variables to improve efficiency compared to the IPW estimator. In addition, the AIPW method allows for missingness to depend on auxiliary variables that are associated with the cause of failure. Hypothesis testing procedures are developed to examine whether covariate effects are zero and whether the effects are constant over time. Critical values of the test statistics are approximated using the Gaussian multipliers resampling approach advocated by Lin et al. (1993). This technique has been shown to work well in many papers, among them, Sun and Wu (2005) and Gilbert and Sun (2015). We justify the validity of our proposed tests numerically through simulation studies.

The paper is organized as follows. Notation and assumptions are introduced in Sect. 2.1. In Sects. 2.2 to 2.4, we develop our AIPW estimator in two stages: the first stage involves IPW estimation, and the second stage involves AIPW estimation utilizing auxiliary information and the IPW estimators from the first stage. The asymptotic results for both the IPW and AIPW estimators are presented in Sect. 3. In Sect. 4, we propose two hypothesis testing procedures for covariate effects. A simulation study is conducted to examine the performances of the proposed estimation and hypothesis testing procedures in Sect. 5. The proposed methods are applied to analyze the Mashi clinical trial data in Sect. 6 and concluding remarks are given in Sect. 7.

2 Two-stage estimation via local linear partial likelihood

2.1 Notation and assumptions

Let T be the failure time and $Z(t)$ a possibly time-dependent p-dimensional covariate over the follow-up time period $[0, \tau]$. Let $\bar{Z}(\tau) = \{Z(t), 0 \le t \le \tau\}$ be the covariate history. For a typical right-censored competing risks data set, the observable random variables are $(X, \delta, \delta V, \bar{Z}(\tau))$, where $X = \min\{T, C\}$, $\delta = I(T \le C)$, the cause of

Author's personal copy

failure V is a categorical variable taking L categories, and C is the censoring time that is assumed to be independent of T and V conditional on $\bar{Z}(\tau)$. The cause of failure V is only observable when $\delta = 1$, whereas if T is censored, the cause is unknown. Suppose that the conditional cause-specific hazard function for cause $V = k$ at time t, given the covariate history $\bar{Z}(\tau)$, only depends on the current value $Z(t)$, which is defined as $\lambda_k(t|z) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \leq T \leq t + \Delta t, V = k \mid T \geq t, Z(t) = z\}$, with t ranging over the interval $[0, \tau]$.

We study the following Cox model of conditional cause-specific hazard functions with time-varying coefficients:

$$\lambda_k(t|Z(t)) = \lambda_{k0}(t) \exp\big(\beta_k(t)^{\mathsf{T}} Z(t)\big), \quad k = 1, 2, \ldots, L, \tag{1}$$

where for each k $\lambda_{k0}(t)$ is an unspecified baseline hazard function, and $\beta_k(t) = \big(\beta_{k1}(t), \ldots, \beta_{kp}(t)\big)^{\mathsf{T}}$ is a p-dimensional vector of unspecified time-dependent regression coefficients.

Let R be the indicator of observing the cause V, where, for observed failure events, $R = 1$ ($R = 0$) if the failure cause is observed (not observed). Moreover, $R = 1$ if the failure time is censored. In addition to the covariate $Z(t)$ considered in model (1), our procedures allow for the use of auxiliary variables A measured at the failure time T for subjects with observed failure. These auxiliaries can justify more plausible missing data assumptions and improve efficiency. In particular, our methods rely on the following missing at random (Rubin 1976) assumption:

$$\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}: \quad P(R = 1|\delta = 1, T, Z(T), A, V) = P(R = 1|\delta = 1, T, Z(T), A). \tag{2}$$

$\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ assumes that the missingness probability is independent of the cause of failure when conditioning on T, $Z(T)$ and A. $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ also implies that V is independent of R given $(T, Z(T), A)$: $P(V = k|R = 1, \delta = 1, T, Z(T), A) = P(V = k|\delta = 1, T, Z(T), A)$.

The assumption $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ is similar to the missing at random assumption in Lu and Tsiatis (2001), Gao and Tsiatis (2005), Lu and Liang (2008), Liu et al. (2018), and Nevo et al. (2018). Other commonly used missing at random assumptions include $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$: $P(R = 1|\delta = 1, T, Z(T), V) = P(R = 1|\delta = 1, T, Z(T))$, and $\text{MAR}_{\boldsymbol{T}}$: $P(R = 1|\delta = 1, T, V) = P(R = 1|\delta = 1, T)$. In many applications certain causes of failure are more likely to be missing than others even after controlling for T and Z (Nevo et al. 2018), such that the missingness probability depends on information related to V conditional on T and Z, which means that $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ is not expected to hold. Goetghebeur and Ryan (1995) studied the cause-specific Cox model with missing causes under the assumption $\text{MAR}_{\boldsymbol{T}}$. Under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$, $P(R = 1|\delta = 1, T, Z(T), V) = E_A\{P(R = 1|\delta = 1, T, Z(T), A, V)\}$, which is likely to depend on V, where E_A stands for the conditional expectation with respect to A conditional on $(\delta = 1, T, Z(T), V)$. $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ can be considered to be a weaker assumption than $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ or $\text{MAR}_{\boldsymbol{T}}$, because it allows the missingness probability to depend on

Author's personal copy

A, which captures some information in V. These assumptions are a priori specified based on the data collection process, and cannot be tested based on the data.

Let $\zeta = (T, Z(T))$, $r(\zeta, A) = P(R = 1|\delta = 1, \zeta, A)$, and $\rho_k(\zeta, A) = P(V = k|\delta = 1, \zeta, A)$. In practice, the conditional probability of missingness $r(\zeta, A)$ and the conditional distribution of the cause of failure $\rho_k(\zeta, A)$ may be influenced by different sets of auxiliary variables. Let A_1 and A_2 be subsets of the auxiliary variables A, where A_1 is relevant in predicting whether V is observed, and A_2 is informative about the conditional distribution of V. That is, $P(R = 1|\delta = 1, \zeta, A) = P(R = 1|\delta = 1, \zeta, A_1)$ and $P(V = k|\delta = 1, \zeta, A) = P(V = k|\delta = 1, \zeta, A_2)$, which leads to $r(\zeta, A) = r(\zeta, A_1)$ and $\rho_k(\zeta, A) = \rho_k(\zeta, A_2)$, respectively. In some situations, the two subsets may be the same and equal to A, i.e., $A_1 = A_2 = A$.

The observed data consist of independent identically distributed (i.i.d.) replicates

$$O_i = \{X_i, \delta_i, \bar{Z}_i(\tau), R_i, R_i\delta_i V_i, \delta_i A_i\}, \quad i = 1, \ldots, n,$$

of $O = \{X, \delta, \bar{Z}(\tau), R, R\delta V, \delta A\}$. We define the counting processes $N_{ik}(t) = I(X_i \le t, \delta_i = 1, V_i = k)$, $N_i(t) = I(X_i \le t, \delta_i = 1)$ and the at-risk process $Y_i(t) = I(X_i \ge t)$.

2.2 Full data local linear partial likelihood estimator

For the competing risks failure time data under consideration, we refer to the full data as the data with no missing failure causes, but right-censoring may still be present. For full data, the local linear partial likelihood method of Cai and Sun (2003) can be used to estimate the regression coefficients model (1). For cause $V = k$, by the Taylor expansion for u in a neighborhood of t, we have

$$\beta_{kl}(u) \approx \beta_{kl}(t) + \beta'_{kl}(t)(u - t), \quad l = 1, 2, \ldots, p.$$

Let $\xi_k(t) = \big(\beta_{k1}(t), \ldots, \beta_{kp}(t), \beta'_{k1}(t), \ldots, \beta'_{kp}(t)\big)^\mathsf{T}$ and $\widetilde{Z}_i(u, u - t) = Z_i(u) \circledast (1, u-t)^\mathsf{T}$, where $\circledast$ is the Kronecker product. Let $S_f^{(j)}(u, t, \xi_k) = n^{-1}\sum_{i=1}^n Y_i(u) \exp\big(\xi_k(t)^\mathsf{T}\widetilde{Z}_i(u, u-t)\big)\big(\widetilde{Z}_i(u, u-t)\big)^{\otimes j}$ for $j = 0, 1$ and 2. Here, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^\mathsf{T}$ for a column vector a. Following Cai and Sun (2003), at each time $t \in (0, \tau)$, the score function for $\xi_k(t)$ is

$$U_f(t, \xi_k) = \sum_{i=1}^n \int_0^\tau K_h(u - t)\Big(\widetilde{Z}_i(u, u - t) - \overline{S_f}(u, t, \xi_k)\Big)\, dN_{ik}(u), \tag{3}$$

where $\overline{S_f}(u, t, \xi_k) = S_f^{(1)}(u, t, \xi_k)/S_f^{(0)}(u, t, \xi_k)$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a symmetric kernel function with support $[-1, 1]$, and h is the bandwidth. The local linear partial maximum likelihood estimator of $\beta_k(t)$ is the vector consisting of the first p components of $\widehat{\xi}_{f,k}(t)$ that solves (3) with respect to ξ_k.

Author's personal copy

2.3 Inverse probability weighted estimator

When the cause of failure is subject to missingness, a straightforward estimation method for $\xi_k(t)$ is to fit the complete data, where the cases with missing failure causes are excluded/ignored, using the local linear partial likelihood score Eq. (3). Such a complete-case estimator is inefficient and may lead to bias when the complete cases are not a random sample of all cases. Following the idea of Horvitz and Thompson (1952), the method of inversely weighting the probability of complete-case has been commonly used in missing data problems under the missing at random assumption $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ given in (2). To estimate the probability of complete-case, a parametric model $r(\zeta_i, A_i, \psi)$ is often used for $r(\zeta_i, A_i)$, where $\zeta_i = (T_i, Z_i(T_i))$ and ψ is a q-dimensional vector of parameters. Suppose that $r(\zeta_i, A_i, \psi)$ is a parametric model for $r(\zeta_i, A_i)$, where $\zeta_i = (T_i, Z_i(T_i))$ and ψ is a q-dimensional vector of parameters. The maximum likelihood estimator $\widehat{\psi}$ of ψ is obtained by maximizing the observed data likelihood,

$$\prod_{i=1}^{n}\Big(r(\zeta_i, A_i, \psi)\Big)^{R_i\delta_i}\Big(1 - r(\zeta_i, A_i, \psi)\Big)^{(1-R_i)\delta_i}. \tag{4}$$

Let $Q_i = (\zeta_i, \delta_i A_i, \delta_i)$. The probability of a complete case $\pi(Q_i) = P(R_i = 1|Q_i) = \delta_i r(\zeta_i, A_i) + (1-\delta_i)$ is then modeled by $\pi(Q_i, \psi) = \delta_i r(\zeta_i, A_i, \psi) + (1-\delta_i)$. Define

$$S_I^{(j)}(u, t, \xi_k, \psi) = n^{-1}\sum_{i=1}^{n} q_i Y_i(u) \exp\Big(\xi_k(t)^{\mathsf{T}}\widetilde{Z}_i(u, u-t)\Big)\Big(\widetilde{Z}_i(u, u-t)\Big)^{\otimes j},$$

for $j = 0, 1, 2$, where $q_i = R_i/\pi(Q_i, \psi)$. Denote $\overline{S_I}(u, t, \xi_k, \psi) = S_I^{(1)}(u, t, \xi_k, \psi)/S_I^{(0)}(u, t, \xi_k, \psi)$.

The inverse probability weighted (IPW) complete-case estimating function for $\xi_k(t)$ is given by

$$U_I(t, \xi_k, \widehat{\psi}) = \sum_{i=1}^{n}\int_0^{\tau} K_h(u-t)\Big(\widetilde{Z}_i(u, u-t) - \overline{S_I}(u, t, \xi_k, \widehat{\psi})\Big)\widehat{q}_i\, dN_{ik}(u),$$

where $\widehat{q}_i = R_i/\pi(Q_i, \widehat{\psi})$. The IPW estimator $\widehat{\xi}_{I,k}(t)$ of $\xi_k(t)$ is the solution of the estimating equation $U_I(t, \xi_k, \widehat{\psi}) = 0$. Then $\widehat{\beta}_{I,k}(t)$ is the first p components of $\widehat{\xi}_{I,k}(t)$.

The baseline function $\lambda_{k0}(t)$ can be estimated by kernel smoothing $\widehat{\lambda}_{I,k0}(t) = \int_0^{\tau} K_h(u-t)\, d\widehat{\Lambda}_{I,k0}(u)$, where

$$\widehat{\Lambda}_{I,k0}(t) = \sum_{i=1}^{n}\int_0^{t}\frac{1}{nS_I^{*(0)}(u, \widehat{\beta}_{I,k}, \widehat{\psi})}\widehat{q}_i\, dN_{ik}(u)$$

Author's personal copy

is the estimator of the cumulative baseline function $\Lambda_{k0}(t) = \int_0^t \lambda_{k0}(u)\,du$, and $S_I^{*(0)}(t, \widehat{\beta}_{I,k}, \widehat{\psi}) = n^{-1}\sum_{i=1}^n \hat{q}_i Y_i(t) \exp\left(\widehat{\beta}_{I,k}(t)^\mathsf{T} Z_i(t)\right)$.

2.4 Two-stage augmented inverse probability weighted estimator

Studies have shown that the IPW estimator is inefficient and relies on the correct modeling of the probability $r(\zeta_i, A_i)$, cf. Scharfstein et al. (1999), Gao and Tsiatis (2005), and Lu and Liang (2008). To increase estimation efficiency, we propose the augmented inverse probability weighted complete-case (AIPW) estimating function obtained by including a projection term of the full data estimating function onto the space of the observed data (Robins et al. 1994). Suppose that the full data estimating function is the sum of independent identically distributed (iid) terms of the form $\sum_{i=1}^n g_i$. Robins et al. (1994) introduced a class of estimators with the estimating equations $\sum_{i=1}^n q_i g_i + (1-q_i)h_i$, where $q_i = R_i/\phi(Q_i)$ and h_i is an arbitrary function of observed data, and showed that the optimal estimator in this class is obtained by taking $h_i = E(g_i|\text{observed data})$, which is termed the AIPW estimator. AIPW estimators have been shown to be more efficient than IPW estimators in many situations, cf. Gao and Tsiatis (2005), Lu and Liang (2008), and Sun et al. (2017), among others. In this section, we develop the AIPW estimator for the Cox model with time-varying coefficients for cause-specific hazard functions as in (1). The proposed AIPW estimating equation utilizes available information for individuals with missing causes through a consistent estimator of the conditional distribution of the failure cause. The IPW estimators $\widehat{\lambda}_{I,k0}(t)$ and $\widehat{\xi}_{I,k}(t)$ are used in the construction of this consistent estimator.

Note that $N_{ik}(t) = N_i(t)I(V_i = k)$, and, using the AIPW formulation above, $g_i = \int_0^\tau K_h(u-t)\big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, t, \xi_k)\big)\,dN_i(u)I(V_i = k)$ for the full data estimating function (3). The implementation of the AIPW procedure requires evaluation of $E(g_i|\text{observed data})$ for each k, which equals $\int_0^\tau K_h(u-t)\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, t, \xi_k)\Big)\,dN_i(u)\rho_k(\zeta_i, A_i)$, where $\zeta_i = (T_i, Z_i(T_i))$ and $\rho_k(t, z, a) = P(V_i = k|\delta_i = 1, \zeta_i = (t, z), A_i = a)$ is the conditional distribution of the failure cause. By Lemma 1 given in Appendix A in the Web-based Supplementary Material, we have

$$\rho_k(t, z, a) = \frac{\lambda_k(t|z) f(a|k, t, z)}{\sum_{l=1}^L \lambda_l(t|z) f(a|l, t, z)}, \tag{5}$$

where $f(a|k, t, z) = P(A_i = a|\delta_i = 1, T_i = t, Z_i = z, V_i = k)$ is a conditional density if A_i is continuous, and is a conditional probability mass function if A_i is discrete. If A_i is independent of V_i conditional on $(\delta_i = 1, T_i, Z_i)$, then $f(a|k, t, z)$ does not depend on k, and in this case $\rho_k(t, z, a) = \lambda_k(t|z)/\sum_{l=1}^L \lambda_l(t|z)$. This relationship also holds when no auxiliary variable A_i is used in the data analysis. In the situation that the auxiliary variable A_i correlates with V_i conditional on $(\delta_i = 1, T_i, Z_i)$, $\rho_k(t, z, a)$ depends on $f(a|k, t, z)$ as well as on $\{\lambda_l(t|z), l = 1, 2, \ldots, L\}$. Although nonparametric/semiparametric density estimation methods are available, developments for conditional density estimation is limited, in particular, if the dimension is high (Hall et al. 2004; Efromovich 2010; Izbicki and Lee 2016). To estimate $f(a|k, t, z)$, we

Author's personal copy

assume the following consistent association condition that the conditional distribution of A_i given $(\delta_i = 1, T_i, Z_i, V_i)$ is the same whether V_i is missing or not, that is,

$$\text{ASSOC}_{\boldsymbol{A}}\text{: } P(A_i = a|\delta_i = 1, T_i, Z_i, V_i, R_i = 1) = P(A_i = a|\delta_i = 1, T_i, Z_i, V_i).$$

Nevo et al. (2018) assumed $\text{ASSOC}_{\boldsymbol{A}}$ in the construction of the informative partial likelihood. Here, we posit a parametric model $f(a|k, t, z, \varphi_k)$ for $f(a|k, t, z)$, where φ_k is a vector of unknown parameters. Maximum likelihood methods can be used to obtain the estimator $\hat{\varphi}_k$ of φ_k.

Let $\widehat{\lambda}_{I,k}(t|z) = \widehat{\lambda}_{I,k0}(t)\exp\left(\widehat{\beta}_{I,k}(t)^\top z\right)$ be the IPW estimator of the conditional cause-specific hazard function. Then, $\rho_k(\zeta_i, A_i)$ can be estimated by

$$\widehat{\rho}_k(\zeta_i, A_i) = \frac{\widehat{\lambda}_{I,k}(T_i|Z_i) f(A_i|k, T_i, Z_i, \widehat{\varphi}_k)}{\sum_{l=1}^{L} \widehat{\lambda}_{I,l}(T_i|Z_i) f(A_i|l, T_i, Z_i, \widehat{\varphi}_l)}. \tag{6}$$

Because the estimator $\widehat{\rho}_k(\zeta_i, A_i)$ is based on the (first-stage) IPW estimator, we term the following estimating function the two-stage AIPW estimating function for $\xi_k(t)$:

$$U_A(t, \xi_k, \widehat{\psi}, \widehat{\rho}_k) = \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t)\Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, t, \xi_k)\Big)\Big[\widehat{q}_i\, dN_{ik}(u) + (1-\widehat{q}_i)\widehat{\rho}_k(\zeta_i, A_i)\, dN_i(u)\Big].$$

The two-stage AIPW estimator of $\xi_k(t)$ is the solution to the estimating equation $U_A(t, \xi_k, \widehat{\psi}, \widehat{\rho}_k) = 0$ and is denoted by $\widehat{\xi}_{A,k}(t)$. The AIPW estimator $\widehat{\beta}_{A,k}(t)$ of $\beta_k(t)$ is the first p components of $\widehat{\xi}_{A,k}(t)$.

The baseline function $\lambda_{k0}(t)$ can be estimated by $\widehat{\lambda}_{A,k0}(t) = \int K_h(u - t)\, d\widehat{\Lambda}_{A,k0}(u)$, where

$$\widehat{\Lambda}_{A,k0}(t) = \sum_{i=1}^{n}\int_0^t \frac{1}{nS^{(0)}(u, \widehat{\beta}_{A,k})}\Big[\widehat{q}_i I(V_i = k) + (1-\widehat{q}_i)\widehat{\rho}_k(\zeta_i, A_i)\Big] dN_i(u)$$

is the estimator of the cumulative baseline function of $\Lambda_{k0}(t)$. Here $S^{(0)}(t, \widehat{\beta}_{A,k})) = n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\widehat{\beta}_{A,k}(t)^\top Z_i(t))$.

Studies show that the choice of kernel generally has little effect on the efficiency of the estimator, cf. Fan and Gijbels (1996). However, the bandwidth plays an essential role on the performance of the kernel-based estimator, which can be selected by a widely used M-fold cross-validation procedure, cf. Rice and Silverman (1991) and Tian et al. (2005). Specifically, we randomly divide the sample into M roughly equal-sized groups, say $(G_1, G_2, \cdots, G_M)$. The cross-validation procedure selects the bandwidth that minimizes the total prediction error $PE(h) = \sum_{m=1}^{M} PE_m(h)$ with respect to h, where

Author's personal copy

$$PE_m(h) = -\sum_{k=1}^{L}\sum_{i\in G_m}\int_0^{\tau}\Big[(\widehat{\beta}_{A,k}^{(-m)}(t))^{\mathsf{T}}Z_i(t) - \log\Big\{\sum_{s\in G_m}Y_s(t)\exp\Big(\widehat{\beta}_{A,k}^{(-m)}(t)^{\mathsf{T}}Z_s(t)\Big)\Big\}\Big]dN_{ik}(t).$$

That is, $h_{opt} = \arg\min_h PE(h)$. Here, $\widehat{\beta}_{A,k}^{(-m)}(t)$ is the AIPW estimator based on the data excluding subjects in G_m. And, $PE_m(h)$ is a cross-validation measure of the prediction error based on the minus of the log-partial likelihood function, cf., Tian et al. (2005). In practice, to increase stability one can take h_{opt} to be the average of multiple h_{opt} values obtained from each of multiple divisions of the sample, e.g., 10 divisions.

Our proposed two-stage estimator $\widehat{\beta}_{A,k}(t)$ has the same structure as the AIPW estimator of Robins et al. (1994), as well as those of Gao and Tsiatis (2005) and Lu and Liang (2008), except that we model $\rho_k(\zeta, A)$ in a different way, incorporating information in the auxiliary variables through estimation of $f(A_i|k, T_i, Z_i, \varphi_k)$. Even if auxiliary variables are not available, the first-stage IPW estimator can be utilized further to improve efficiency through $\widehat{\rho}_k(\zeta_i, A_i) = \widehat{\lambda}_{I,k}(T_i|Z_i)/\sum_{l=1}^{L}\widehat{\lambda}_{I,l}(T_i|Z_i)$, since the AIPW estimator has the smallest variance among a class of the estimators that includes the IPW estimator (Robins et al. 1994). The informative likelihood approach for the cause-specific Cox model developed by Nevo et al. (2018) is a related approach that utilizes parametric models for $\rho_k(\zeta, A)$ and $f(A_i|k, T_i, Z_i)$.

Although our proposed two-stage AIPW estimator improves the efficiency over the IPW estimator, as shown in the simulation study, it does not possess the double robustness property enjoyed by regular AIPW estimators that model both the probability of a complete case $r(\zeta_i, A_i)$ and the conditional distribution of the cause of failure $\rho_k(\zeta_i, A_i)$; see Gao and Tsiatis (2005) and Lu and Liang (2008), among others. An AIPW estimator would be considered to be doubly robust if it is consistent for its target parameter if either $r(\zeta_i, A_i)$ or $\rho_k(\zeta_i, A_i)$ is correctly specified. However, for the proposed two-stage AIPW estimator, if $r(\zeta_i, A_i)$ is misspecified, then the estimator of $\rho_k(\zeta_i, A_i)$ is inconsistent as it is based on an inconsistent IPW estimator. Thus the two-stage AIPW estimator is inconsistent if $r(\zeta_i, A_i)$ is misspecified. On the other hand, it is challenging to find the correct model for the conditional distribution of the cause of failure for regular AIPW estimators. The estimator $\widehat{\rho}_k(\zeta_i, A_i)$ given in (6) may provide a better or at least an alternative method, which instead models the conditional cause-specific hazard function and the conditional density or mass function of the auxiliary variable. Our simulation studies show that the proposed two-stage AIPW estimator has better finite-sample performance than the IPW and complete-case estimators even when the models for $r(\zeta_i, A_i)$ and $\rho_k(\zeta_i, A_i)$ are both misspecified.

3 Asymptotic properties

We investigate the asymptotic properties of the IPW estimator $\widehat{\beta}_{I,k}(t)$ and the AIPW estimator $\widehat{\beta}_{A,k}(t)$. The regularity conditions for the asymptotic results are given in

Author's personal copy

(C.1)–(C.5) placed in the "Appendix". Conditions (C.1)–(C.4) are standard assumptions for the local linear method under the Cox model with time-varying coefficients (Cai and Sun 2003). Condition (C.5) is a smoothness condition on the parametric models for $r(\zeta_i, A_i)$ and $f(A_i|k, T_i, Z_i)$, comparable to the assumptions used in Gao and Tsiatis (2005) and Lu and Liang (2008).

3.1 Asymptotic results of the IPW estimator

The consistency and asymptotic normality of $\widehat{\beta}_{I,k}(t),\ k = 1, 2, \ldots, L$, are established in the next two theorems. To avoid the problems at the boundaries $t = 0$ and $t = \tau$, we study the asymptotic properties of $\widehat{\beta}_k(t)$ for interior values of $t \in [t_1, t_2] \subset (0, \tau)$. The proofs of Theorems 1 and 2 are placed in Appendix A in the Web-based Supplementary Material.

Theorem 1 *Assume* $\mathrm{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$. *Under conditions (C.1)–(C.5) given in the "Appendix", if the model for* $r(\zeta_i, A_i)$ *is correctly specified, then* $\widehat{\beta}_{I,k}(t) \xrightarrow{P} \beta_k(t)$ *uniformly in* $t \in [t_1, t_2] \subset (0, \tau)$ *as* $n \to \infty$.

Theorem 2 *Assume* $\mathrm{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$. *Under conditions (C.1)–(C.5) given in the "Appendix", if the model for* $r(\zeta_i, A_i)$ *is correctly specified, then*

$$\sqrt{nh}\Big(\widehat{\beta}_{I,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2 \beta_k''(t)\Big) \xrightarrow{\mathcal{D}} N\Big(0, \nu_0 \Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)\Big),$$

for $t \in [t_1, t_2] \subset (0, \tau)$ *as* $n \to \infty$, *where* $\mu_2 = \int_{-1}^{1} x^2 K(x)dx$, $\nu_0 = \int_{-1}^{1} K^2(x)dx$, *and* $\Sigma_k(t)$ *and* $\Sigma_k^*(t)$ *are defined in the "Appendix".*

Let $\mathcal{I}_{I,k}(t)$ be the upper left $p \times p$ matrix of $n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(u-t)J_I(u, t, \widehat{\xi}_{I,k}(t), \widehat{\psi})\widehat{q}_i\, dN_{ik}(u)$, and $\widetilde{\Sigma}_{I,k}(t)$ be the upper left $p \times p$ matrix of

$$\frac{h}{n}\sum_{i=1}^{n}\int_0^\tau \big(K_h(u-t)\big)^2\Big(\widetilde{Z}_i(u, u-t) - \overline{S_I}(u, t, \widehat{\xi}_{I,k}(t), \widehat{\psi})\Big)^{\otimes 2}\widehat{q}_i^2\, dN_{ik}(u),$$

where $J_I(u, t, \xi_k, \psi) = S_I^{(2)}(u, t, \xi_k, \psi)/S_I^{(0)}(u, t, \xi_k, \psi) - (\overline{S_I}(u, t, \xi_k, \psi))^{\otimes 2}$. Then the asymptotic covariance matrix $\nu_0\Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)$ of $\sqrt{nh}(\widehat{\beta}_{I,k}(t) - \beta_k(t))$ can be consistently estimated by $\mathcal{I}_{I,k}^{-1}(t)\widetilde{\Sigma}_{I,k}(t)\mathcal{I}_{I,k}^{-1}(t)$ as $n \to \infty$.

3.2 Asymptotic results of the AIPW estimator

Next, we present the asymptotic properties of the AIPW estimators $\widehat{\beta}_{A,k}(t),\ k = 1, 2, \ldots, L$. Theorem 3 shows that the AIPW estimators are consistent if $r(\zeta_i, A_i)$ is correctly specified. Theorem 4 shows the asymptotic normality of $\widehat{\beta}_{A,k}(t),\ k = 1, 2, \ldots, L$. The proofs of Theorems 3 and 4 are placed in Appendix A in the Web-based Supplementary Material.

Author's personal copy

Theorem 3 *Assume* $\mathrm{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ *and* $\mathrm{ASSOC}_{\boldsymbol{A}}$. *Under conditions (C.1)–(C.5) given in the "Appendix",* $\widehat{\beta}_{A,k}(t) \xrightarrow{P} \beta_k(t)$ *uniformly in* $t \in [t_1, t_2] \subset (0, \tau)$ *as* $n \to \infty$. *This consistency holds if* $r(\zeta_i, A_i)$ *is correctly specified.*

Theorem 4 *Assume* $\mathrm{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ *and* $\mathrm{ASSOC}_{\boldsymbol{A}}$. *Under conditions (C.1)–(C.5) given in the "Appendix", if both* $r(\zeta_i, A_i)$ *and* $f(A_i|k, T_i, Z_i)$ *are correctly specified, then*

$$\sqrt{nh}\Big(\widehat{\beta}_{A,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2 \beta_k''(t)\Big) \xrightarrow{\mathcal{D}} N\Big(0, \nu_0 \Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)\Big)$$

for $t \in [t_1, t_2] \subset (0, \tau)$ *as* $n \to \infty$, *where* $\Sigma_k(t)$ *and* $\Sigma_k^*(t)$ *are defined in the "Appendix". The estimator* $\widehat{\beta}_{A,k}(t)$ *is more efficient than* $\widehat{\beta}_{I,k}(t)$ *at the order of h, which is shown in the equation below:*

$$\begin{aligned}
&Cov\Big\{\sqrt{nh}\Big(\widehat{\beta}_{I,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2 \beta_k''(t)\Big)\Big\} \\
&\quad = Cov\Big\{\sqrt{nh}\Big(\widehat{\beta}_{A,k}(t) - \beta_k(t) - \frac{1}{2}\mu_2 h^2 \beta_k''(t)\Big)\Big\} \\
&\qquad + h\Sigma_k^{-1}(t) Cov\{\mathcal{B}_1(t, \beta_k) - \mathcal{O}_1(t, \beta_k)\}\Sigma_k^{-1}(t) + o_p(h),
\end{aligned}$$

where $\mathcal{B}_i(t, \beta_k)$ *and* $\mathcal{O}_i(t, \beta_k)$ *are defined in the "Appendix".*

Let $\mathcal{I}_{A,k}(t)$ be the upper left $p \times p$ matrix of

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(u-t) J_f(u, t, \widehat{\xi}_{A,k}(t))\Big[\widehat{q}_i I(V_i = k) + (1 - \widehat{q}_i)\widehat{\rho}_k(\zeta_i, A_i)\Big] dN_i(u),$$

and $\widetilde{\Sigma}_{A,k}(t)$ be the upper left $p \times p$ matrix of

$$\begin{aligned}
&\frac{h}{n}\sum_{i=1}^{n}\int_0^{\tau} \big(K_h(u-t)\big)^2 \Big(\widetilde{Z}_i(u, u-t) - \overline{S_f}(u, t, \widehat{\xi}_{A,k}(t))\Big)^{\otimes 2} \\
&\quad \Big[\widehat{q}_i I(V_i = k) + (1 - \widehat{q}_i)\widehat{\rho}_k(\zeta_i, A_i)\Big]^2 dN_i(u),
\end{aligned}$$

where $J_f(u, t, \xi_k) = S_f^{(2)}(u, t, \xi_k)/S_f^{(0)}(u, t, \xi_k) - (\overline{S_f}(u, t, \xi_k))^{\otimes 2}$. The asymptotic covariance matrix $\nu_0 \Sigma_k^{-1}(t)\Sigma_k^*(t)\Sigma_k^{-1}(t)$ of $\sqrt{nh}(\widehat{\beta}_{A,k}(t) - \beta_k(t))$ can be consistently estimated by $\mathcal{I}_{A,k}^{-1}(t)\widetilde{\Sigma}_{A,k}(t)\mathcal{I}_{A,k}^{-1}(t)$ as $n \to \infty$.

Let $B_k(t) = \int_{t_1}^{t} \beta_k(s)ds$ and $\widehat{B}_{A,k}(t) = \int_{t_1}^{t} \widehat{\beta}_{A,k}(s)ds$. The following theorem presents a weak convergence result for $G_n(t) = n^{1/2}(\widehat{B}_{A,k}(t) - B_k(t))$ over $t \in [t_1, t_2] \subset [0, \tau]$. The result provides a justification for using the Gaussian multipliers resampling method to estimate the critical values of the hypothesis tests developed next, cf. Lin et al. (1993).

Author's personal copy

Theorem 5 *Assume* $\mathrm{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ *and* $\mathrm{ASSOC}_{\boldsymbol{A}}$. *Under conditions (C.1)–(C.5) given in the "Appendix",* $G_n(t) = n^{-1/2}\sum_{i=1}^{n} H_i(t) + o_p(1)$ *uniformly for* $t \in [t_1, t_2]$ *if both* $r(\zeta_i, A_i)$ *and* $f(A_i|k, T_i, Z_i)$ *are correctly specified, where*

$$H_i(t) = \int_{t_1}^{t} \Sigma_k^{-1}(u)\left[Z_i(u) - \frac{s^{(1)}(u, \beta_k)}{s^{(0)}(u, \beta_k)}\right]\left\{\frac{R_i}{\pi(Q_i)} dM_{ik}(u) + \left[1 - \frac{R_i}{\pi(Q_i)}\right] E(dM_{ik}(u)|Q_i)\right\},$$

and $s^{(j)}(u, \beta_k)$, $j = 1, 2$, *are defined in the "Appendix". The process* $G_n(t)$ *converges weakly to a zero-mean Gaussian process on* $[t_1, t_2]$.

4 Hypothesis testing for $\boldsymbol{\beta_k(t)}$

For assessing time-varying covariate effects on a specific cause of failure, we develop hypothesis testing procedures to test two null hypotheses – H_{10}: $\beta_k(t) = 0$ for $t \in [t_1, t_2]$ and H_{20}: $\beta_k(t)$ does not depend on t for $t \in [t_1, t_2]$. Without loss of generality, we assume $\beta_k(t)$ is a one-dimensional function. Accordingly, this procedure applies for testing a certain component of the p-dimensional function vector.

First, we develop tests of H_{10}: $\beta_k(t) = 0$ for $t \in [t_1, t_2]$ against two alternative hypotheses – H_{1a}: $\beta_k(t) \neq 0$ for some $t \in [t_1, t_2]$ and H_{1m}: $\beta_k(t) \leq 0$ with strict inequality for some $t \in [t_1, t_2]$. Based on the test process $D_1(t) = n^{1/2}\widehat{B}_{A,k}(t)$, $t \in [t_1, t_2]$, we propose the following test statistics for testing H_{10}:

$$T_{a1}^{(1)} = \sup_{t\in[t_1,t_2]} |D_1(t)|, \qquad T_{a2}^{(1)} = \int_{t_1}^{t_2} (D_1(t))^2\, du,$$

$$T_{m1}^{(1)} = \inf_{t\in[t_1,t_2]} D_1(t), \qquad T_{m2}^{(1)} = \int_{t_1}^{t_2} D_1(t)\, du.$$

General departures under H_{1a} are captured by $T_{a1}^{(1)}$ and $T_{a2}^{(1)}$, whereas $T_{m1}^{(1)}$ and $T_{m2}^{(1)}$ are sensitive to monotone departures under H_{1m}. From Theorem 5, the distribution of $D_1(t)$ can be approximated using the Gaussian multipliers resampling method on $t \in [t_1, t_2]$ under H_{10}. Let $S^{(j)}(t, \beta_k) = n^{-1}\sum_{i=1}^{n} Y_i(t)\exp\left(\beta_k(t)^{\mathsf{T}} Z_i(t)\right) Z_i(t)^{\otimes j}$, for $j = 0, 1, 2$, and let $\overline{S}(t, \beta_k) = S^{(1)}(t, \beta_k)/S^{(0)}(t, \beta_k)$. Let $\phi_1, \cdots, \phi_n$ be independent standard normal random variables. By Theorem 5 and the Gaussian multipliers resampling method, cf. Lin et al. (1993), the distribution of the process $D_1(t)$ under H_{10} can be approximated by the conditional distribution of the process $D_1^*(t) = n^{-1/2}\sum_{i=1}^{n} \hat{H}_i(u)\phi_i$ given the observed data, where

$$\widehat{H}_i(t) = \int_{t_1}^{t} \mathcal{I}_{A,k}^{-1}(s)\left\{\int_0^{\tau} K_h(u - s)\left(Z_i(u) - \overline{S}(u, \widehat{\beta}_{A,k})\right)\left[\widehat{q}_i I(V_i = k) + (1 - \widehat{q}_i)\widehat{\rho}_k(\zeta_i, A_i)\right] dN_i(u)\right\} ds.$$

Author's personal copy

It can be shown that by Lemma 1 of Sun and Wu (2005), $D_1(t)$ and $D_1^*(t)$ have the same asymptotic distribution under H_{10}. Thus, the distributions of $T_{a1}^{(1)}$, $T_{a2}^{(1)}$, $T_{m1}^{(1)}$ and $T_{m2}^{(1)}$ can be approximated by the empirical distributions of

$$T_{a1}^{(1)*} = \sup_{t\in[t_1,t_2]} |D_1^*(t)|, \qquad T_{a2}^{(1)*} = \int_{t_1}^{t_2} (D_1^*(t))^2\, du,$$
$$T_{m1}^{(1)*} = \inf_{t\in[t_1,t_2]} D_1^*(t), \qquad T_{m2}^{(1)*} = \int_{t_1}^{t_2} D_1^*(t)\, du, \tag{7}$$

respectively, obtained through sampling a large number of, say 1000, variable sets $(\phi_1, \cdots, \phi_n)$. The p-values of the test statistics $T_{a1}^{(1)}$ and $T_{a2}^{(1)}$ for testing H_{10} against H_{1a} are the conditional probabilities that $T_{a1}^{(1)*} > T_{a1}^{(1)}$ and $T_{a2}^{(1)*} > T_{a2}^{(1)}$ given the observed data, respectively. Similarly, the p-values of the test statistics $T_{m1}^{(1)}$ and $T_{m2}^{(1)}$ for testing H_{10} against H_{1m} are the conditional probabilities that $T_{m1}^{(1)*} < T_{m1}^{(1)}$ and $T_{m2}^{(1)*} < T_{m2}^{(1)}$ given the observed data, respectively.

To test the null hypothesis H_{20} that the covariate effect is a constant, we also consider two alternative hypotheses – H_{2a}: $\beta_k(t)$ changes over $t \in [t_1, t_2]$ and H_{2m}: $\beta_k(t)$ increases with $t \in [t_1, t_2]$. Let

$$D_2(t) = n^{1/2}\left\{\frac{\widehat{B}_{A,k}(t) - \widehat{B}_{A,k}(t_1)}{t - t_1} - \frac{\widehat{B}_{A,k}(t_2) - \widehat{B}_{A,k}(t_1)}{t_2 - t_1}\right\}.$$

Based on the test process $D_2(t)$, we propose the following test statistics for H_{20}:

$$T_{a1}^{(2)} = \sup_{t\in[t_1^*,t_2]} |D_2(t)|, \qquad T_{a2}^{(2)} = \int_{t_1^*}^{t_2} (D_2(t))^2\, du,$$
$$T_{m1}^{(2)} = \inf_{t\in[t_1^*,t_2]} D_2(t), \qquad T_{m2}^{(2)} = \int_{t_1^*}^{t_2} D_2(t)\, du,$$

where t_1^* is a number between t_1 and t_2 to ensure the value of the denominator in $D_2(t)$ is not zero. We can choose t_1^* close to t_1 to make use of more data, and to make the tests more stable. The test statistics $T_{a1}^{(2)}$ and $T_{a2}^{(2)}$ can capture variation of $\beta_k(t)$ over time, whereas $T_{m1}^{(2)}$ and $T_{m2}^{(2)}$ can capture an increasing trend in $\beta_k(t)$ over time.

For testing H_{20}, we compute

$$D_2^*(t) = n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\hat{H}_i(t) - \hat{H}_i(t_1)}{t - t_1} - \frac{\hat{H}_i(t_2) - \hat{H}_i(t_1)}{t_2 - t_1}\right\}\phi_i$$

to approximate the null distribution of $D_2(t)$ by the Gaussian multipliers technique. The distributions of $T_{a1}^{(2)}$, $T_{a2}^{(2)}$, $T_{m1}^{(2)}$ and $T_{m2}^{(2)}$ can be approximated by the empirical distributions of

Author's personal copy

$$T_{a1}^{(2)*} = \sup_{t\in[t_1^*,t_2]} |D_2^*(t)|, \qquad T_{a2}^{(2)*} = \int_{t_1^*}^{t_2} (D_2^*(t))^2\, du,$$

$$T_{m1}^{(2)*} = \inf_{t\in[t_1^*,t_2]} D_2^*(t), \qquad T_{m2}^{(2)*} = \int_{t_1^*}^{t_2} D_2^*(t)\, du, \tag{8}$$

respectively, obtained through sampling a large number of, say 1000, variable sets $(\phi_1, \cdots, \phi_n)$. The p-values of the test statistics $T_{a1}^{(2)}$ and $T_{a2}^{(2)}$ for testing H_{20} against H_{2a} are the conditional probabilities that $T_{a1}^{(2)*} > T_{a1}^{(2)}$ and $T_{a2}^{(2)*} > T_{a2}^{(2)}$ given the observed data, respectively. Similarly, the p-values of the test statistics $T_{m1}^{(1)}$ and $T_{m2}^{(1)}$ for testing H_{20} against H_{2m} are the conditional probabilities that $T_{m1}^{(2)*} < T_{m1}^{(2)}$ and $T_{m2}^{(2)*} < T_{m2}^{(2)}$ given the observed data, respectively.

5 Numerical results

We present a simulation study conducted to evaluate the performance of the proposed estimation and hypothesis testing procedures. We compare the two-stage AIPW estimator to the IPW estimator and to the complete-case data estimator (CC) that deletes the observations with missing causes from the analysis. These estimators are also compared to the full data likelihood estimator (FULL), which analyzes the simulated dataset without missing causes.

We consider a competing risks model with two failure causes with the cause-specific hazard functions equal to

$$\lambda_1(t|Z) = 0.2\exp((\theta_1 t + \theta_2)Z), \quad \lambda_2(t|Z) = 0.1(t+0.1)^{-1/2}\exp((t+0.1)^{1/2}Z), \tag{9}$$

where Z has a uniform distribution on [0, 1]. Different values of the parameters (θ_1, θ_2) are chosen to examine the sizes and powers of proposed tests, as well as the performances of the proposed estimation procedures. All failure times are censored at the administration time $\tau = 2$. The random-censoring time C is generated from a uniform distribution on [0, 10] which yields about 50% censoring. We simulate the missing causes under two missing at random assumption scenarios. First, we consider a logistic regression model $\text{logit}\{r(Z_i, A_i, \psi)\} = \psi_0 + \psi_1 Z_i + \psi_2 A_i$ for $r(T_i, Z_i(T_i), A_i)$ under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$, where $\psi = (\psi_0, \psi_1, \psi_2)$. The percentages of missing causes are approximately 30%, 40% and 50% for $\psi = (1.4, -0.5, -0.5)$, $\psi = (1, -0.5, -0.5)$ and $\psi = (0.5, -0.5, -0.5)$, respectively. Then, a logistic regression model $\text{logit}\{r(Z_i, \psi)\} = \psi_0 + \psi_1 Z_i$ is considered under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$, where $\psi = (\psi_0, \psi_1)$. We set the parameters to $\psi = (1.4, -1)$, $\psi = (1, -1)$ and $\psi = (0.5, -1)$, which yield 30%, 40% and 50% missing causes, respectively. We use ($r30$), ($r40$) and ($r50$) to denote the percentages of missing causes, 30%, 40% and 50%, respectively.

Author's personal copy

We also generate a binary auxiliary covariate A_i for the failure cause V_i from the following models for $f(a|k,t,z)$:

$$P(A_i = 1|V_i = k) = \frac{e^{a_k}}{1+e^{a_k}}, \quad k = 1, 2. \tag{10}$$

The models allow A_i to depend on V_i, but not T_i and Z_i conditional on V_i; thus $f(a|k,t,z) = f(a|k)$. We examine the performance of the estimators under four different levels of association between A_i and V_i, by considering the settings $(a_1, a_2) = (1, 1)$, $(a_1, a_2) = (-1, 1)$, $(a_1, a_2) = (-2, 2)$, and $(a_1, a_2) = (-3, 3)$, which result in approximate Kendall's tau values of 0, 0.45, 0.75 and 0.90, respectively. We denote these four auxiliary association level settings by (A0), (A1), (A2), and (A3), respectively. We note that A_i is independent of V_i for the setting (A0), and the association between A_i and V_i increases from (A1) to (A3). Let $\widehat{f}(a|k)$ be the estimator of the conditional mass function $f(a|k)$ of A_i given $V_i = k$. If follows from (5) that $\rho_k(t,z,a)$ is estimated by

$$\widehat{\rho}_k(t,z,a) = \frac{\widehat{\lambda}_{I,k}(t|z)\widehat{f}(a|k)}{\widehat{\lambda}_{I,1}(t|z)\widehat{f}(a|1) + \widehat{\lambda}_{I,2}(t|z)\widehat{f}(a|2)}, \quad \text{for } k = 1, 2,$$

where $\widehat{\lambda}_{I,k}(t|z)$, $k = 1, 2$, are the first stage IPW estimators.

To study the performance of the proposed IPW and AIPW estimators under misspecifications of the models $r(t,z,a,\psi)$ and/or $f(a|k)$, the simulations are conducted by positing a misspecified constant model $r_0 \in (0, 1)$ for $r(t,z,a,\psi)$ and/or by positing (A0) while the true setting is (A2). The estimators based on the correctly specified models of $r(t,z,a,\psi)$ and $f(a|k)$ are compared to those obtained when at least one of the two models is misspecified. We use IPW-c to denote the IPW estimator with the correctly specified model $r(t,z,a,\psi)$ for missing causes, and IPW-m for the IPW estimator with misspecified model for missing causes. AIPW-A2-c stands for the AIPW estimator under the setting (A2) with the correctly specified model for $r(t,z,a,\psi)$, and AIPW-A2-m stands for the AIPW estimator under the setting (A2) for the misspecified model for missing causes. AIPW-mA2-c stands for the AIPW estimator with misspecified $f(a|k)$ by assuming (A0) while the true setting is (A2) but correctly specified model $r(t,z,a,\psi)$, while AIPW-mA2-m is the AIPW estimator where both $f(a|k)$ and $r(t,z,a,\psi)$ are misspecified.

The finite sample performances of the proposed test procedures are evaluated through simulations under model (9) with the following parameter settings:

(1) For testing H_{10}, M_1: $(\theta_1, \theta_2) = (0, 0)$, M_2: $(\theta_1, \theta_2) = (0, -0.6)$, M_3: $(\theta_1, \theta_2) = (0, -0.9)$, and M_4: $(\theta_1, \theta_2) = (0, -1.2)$;
(2) For testing H_{20}, N_1: $(\theta_1, \theta_2) = (0, -0.5)$, N_2: $(\theta_1, \theta_2) = (0.9, 0)$, N_3: $(\theta_1, \theta_2) = (1.2, 0)$, and N_4: $(\theta_1, \theta_2) = (1.5, 0)$.

The estimation procedures are examined under the setting N_1 of model (9). The observed sizes and powers of the tests for testing $\beta_1(t)$ are examined under model (9) under the settings M_1 to M_4 for testing H_{10}, and N_1 to N_4 for testing H_{20}, where

Author's personal copy

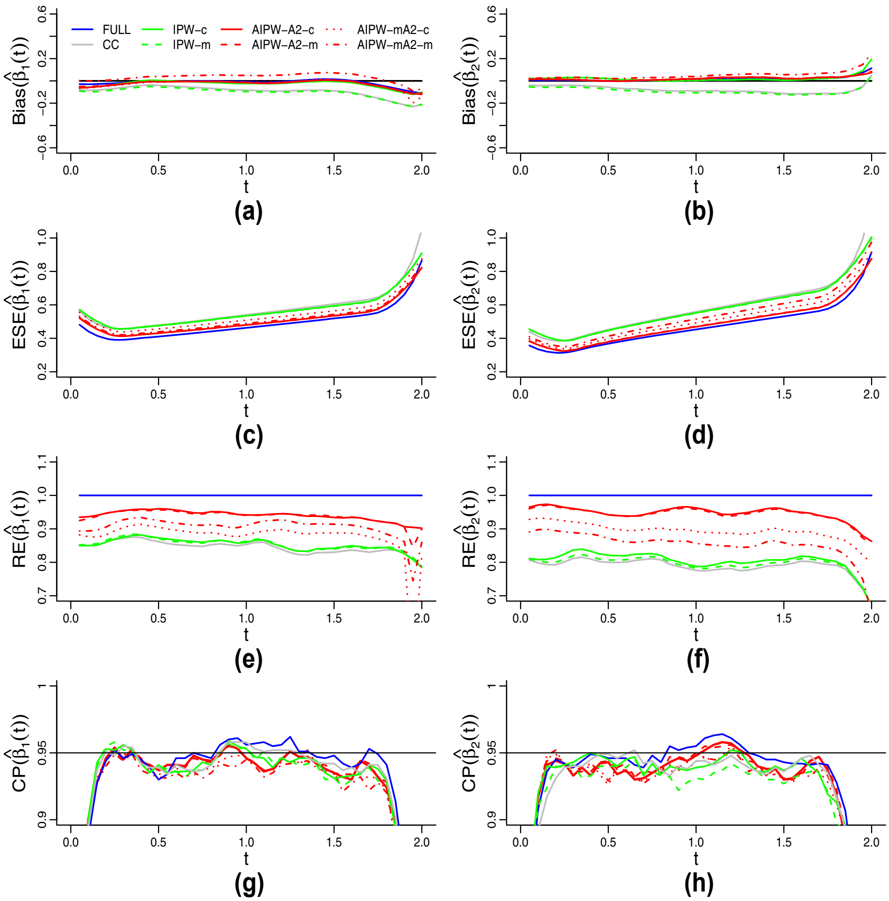

Fig. 1 Bias, mean of the estimated standard errors (ESE), relative efficiency (RE) and 95% coverage probability (CP) of the IPW and AIPW estimators for the setting N_1 : $\beta_1(t) = -0.5$ and $\beta_2(t) = (t + 0.1)^{1/2}$, with 30% of causes missing under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ based on 1000 simulations for $n = 1200$ and $h = 0.3$. The legends AIPW-A2 and AIPW-mA2 refer to the AIPW estimators using the correctly specified (A2) and misspecified (A2), respectively, while -c and -m indicate the estimators using the correctly specified and misspecified model for $r(t, z, a)$, respectively. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only

M_1 and N_1 are the null hypotheses under H_{10} and H_{20}, respectively, M_2 to M_4 are the alternatives to H_{10}, and N_2 to N_4 are the alternatives to H_{20}.

Throughout the simulations, we use the Epanechnikov kernel $K(x) = 3/4(1 - x^2)I\{|x| \leq 1\}$. The simulation study uses 1000 iterations.

Under setting N_1 of model (9), with sample size $n = 1200$ under various correctly specified and misspecified models with 30% of causes missing for each of the two missing at random scenarios $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$, simulation results for $\widehat{\beta}_{I,k}(t)$, $\widehat{\beta}_{A,k}(t)$, and the CC and the FULL estimators are reported in Figs. 1 to 4. The 5-fold cross-validation bandwidth selection procedure of Sect. 2.4 indicates $h = 0.3$ as a reasonable bandwidth, based on the average of total prediction error

Author's personal copy

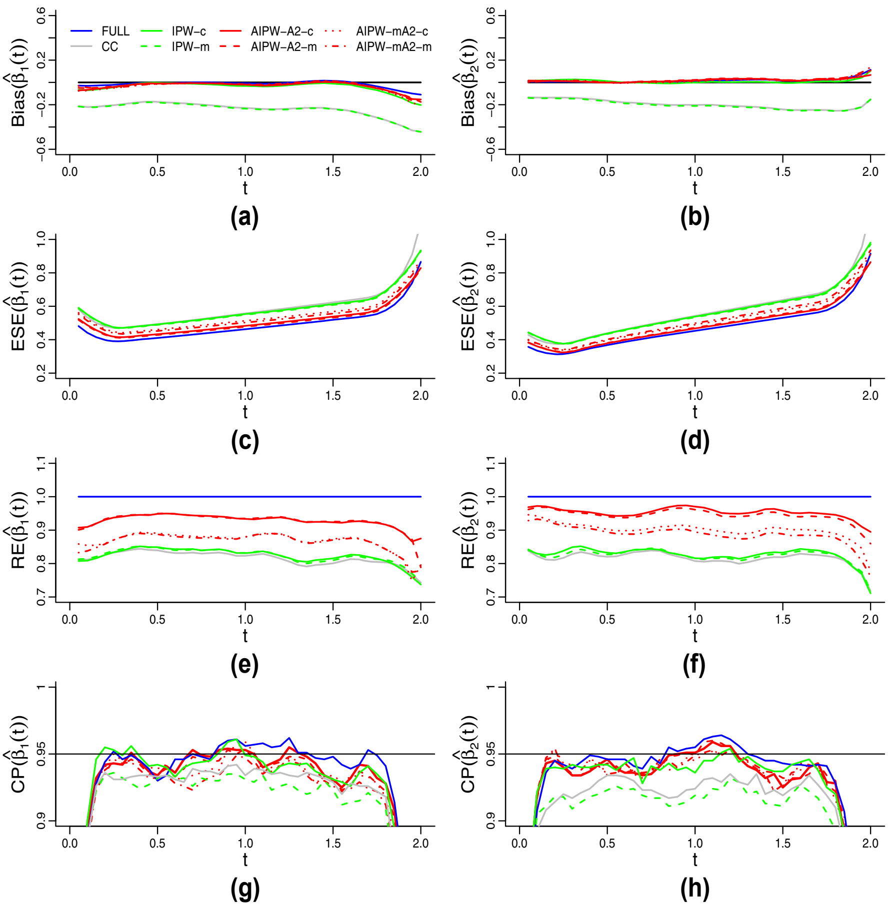

Fig. 2 Bias, mean of the estimated standard errors (ESE), relative efficiency (RE) and 95% coverage probability (CP) of the IPW and AIPW estimators for the setting N_1 : $\beta_1(t) = -0.5$ and $\beta_2(t) = (t + 0.1)^{1/2}$, with 30% of causes missing under MAR$_{\boldsymbol{T},\boldsymbol{Z}}$ based on 1000 simulations for $n = 1200$ and $h = 0.3$. The legends AIPW-A2 and AIPW-mA2 refer to the AIPW estimators using the correctly specified (A2) and misspecified (A2), respectively, while -c and -m indicate the estimators using the correctly specified and misspecified model for $r(t, z, a)$, respectively. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only

$PE(h) = \sum_{l=1}^{5} PE_l(h)$ across 10 simulations for $h \in [0.1, 0.5]$ decreasing as h declines from 0.1 to 0.3 and stabilizing for $h \geq 0.3$ (Web Figs. 5 and 6). The finite sample performances of the estimators are assessed through the bias (Bias), the sample standard error (SSE) of the estimator and mean of the estimated standard errors (ESE) in 1000 simulations. The standard errors of the IPW estimators are estimated using the formula following Theorem 2 in Sect. 3.1, the standard errors of the AIPW estimators are estimated using the formula following Theorem 4 in Sect. 3.2, and the standard errors of the CC and the FULL estimators are obtained by letting $\hat{q}_i = 1$ in the estimated standard errors for the IPW estimators. The 95% confidence interval of $\beta(v)$ at each v is constructed using the estimated $\beta(v)$ plus/minus 1.96 times the

Author's personal copy

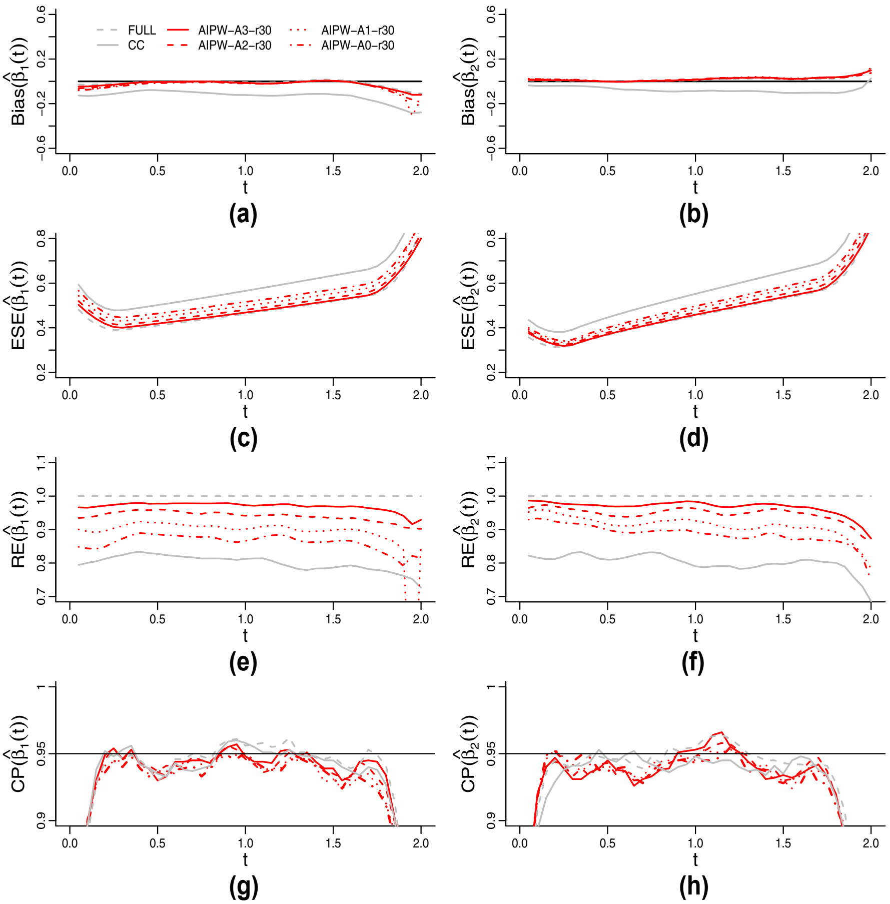

Fig. 3 Bias, mean of the estimated standard errors (ESE), relative efficiency (RE) and 95% coverage probability (CP) of the AIPW estimators for the setting N_1 : $\beta_1(t) = -0.5$ and $\beta_2(t) = (t + 0.1)^{1/2}$, with 30% of causes missing under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and the settings (A0), (A1), (A2) and (A3), based on 1000 simulations for $n = 1200$ and $h = 0.3$. The legends AIPW-A0-r30, AIPW-A1-r30, AIPW-A2-r30 and AIPW-A3-r30 stand for the AIPW estimators under the settings (A0), (A1), (A2) and (A3), respectively, with 30% of causes missing. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only

estimated standard error of the estimator for each simulated data set, and its 95% coverage probability is the percentage of the times of the 95% confidence intervals include the true $\beta(v)$ in 1000 simulations. The efficiency of each estimator is evaluated through the relative efficiency to the FULL estimator, defined by SSE of FULL estimator divided by SSE of the estimator under evaluation. Additional simulation results for sample size $n = 800$ and different percentages of the missing causes and bandwidths are reported in Appendix B of the Web-based Supplementary Material.

Figures 1 and 2 show the biases, relative efficiencies, ESE, and 95% empirical coverage probabilities of the estimators of $\beta_1(t)$ and $\beta_2(t)$ based on 1000 simulations under the $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ scenarios, respectively. The complete-case (CC)

Author's personal copy

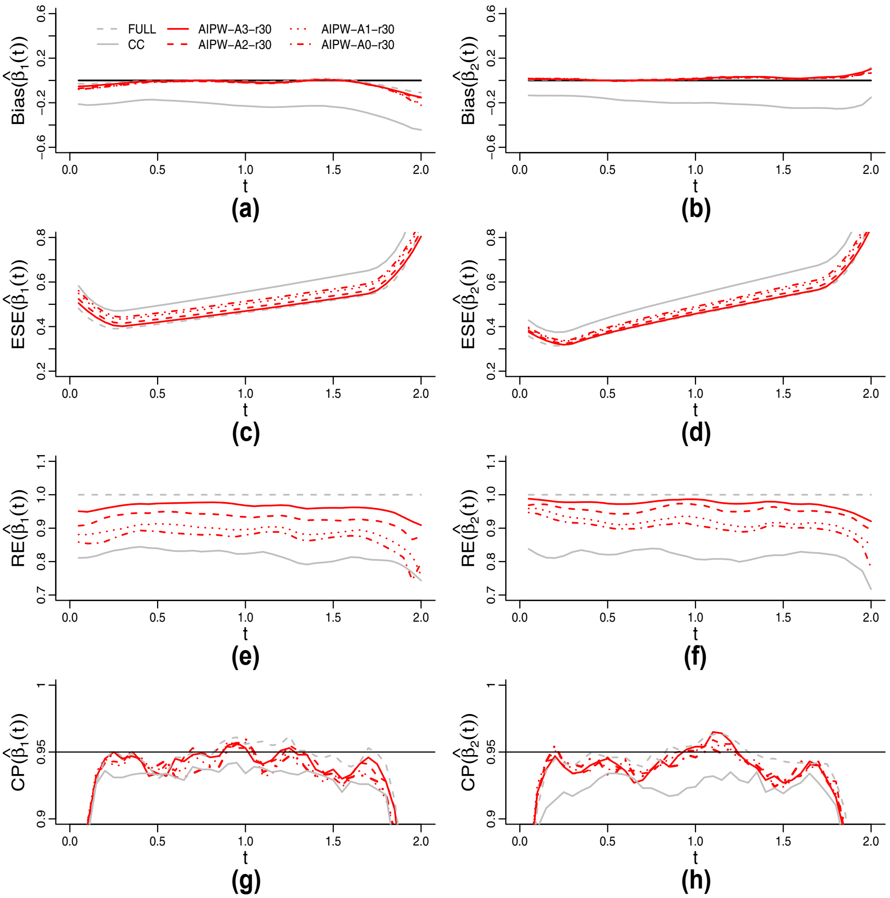

Fig. 4 Bias, mean of the estimated standard errors (ESE), relative efficiency (RE) and 95% coverage probability (CP) of the AIPW estimators for the setting $N_1 : \beta_1(t) = -0.5$ and $\beta_2(t) = (t + 0.1)^{1/2}$, with 30% of causes missing under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ and the settings (A0), (A1), (A2) and (A3), based on 1000 simulations for $n = 1200$ and $h = 0.3$. The legends AIPW-A0-r30, AIPW-A1-r30, AIPW-A2-r30 and AIPW-A3-r30 stand for the AIPW estimators under the settings (A0), (A1), (A2) and (A3), respectively, 30% of causes missing. FULL is for the estimator based on the full data and CC is for the estimator based on the complete data only

estimator has large biases. The biases of the IPW estimator are small when the parametric model for $r(t, z, a)$ is correctly specified but large when it is misspecified. The biases of the AIPW estimators are smaller than those of the IPW estimators for all settings, even in the case that the parametric models are both misspecified. Furthermore, the AIPW estimators are more efficient than the IPW estimators even when the auxiliary variable A_i is independent of V_i. The 95% confidence intervals have reasonably accurate empirical coverage probabilities, slightly lower than the expected 95%, with greater deviation for the CC estimator and the IPW-m estimator.

Author's personal copy

Table 1 Empirical sizes and powers of the test statistics $T_{a1}^{(1)}, T_{a2}^{(1)}, T_{m1}^{(1)}$ and $T_{m2}^{(1)}$ for testing H_{10}: $\beta_1(t) = 0$ for $t \in [t_1, t_2]$ at the nominal level 0.05 for models M_1 to M_4, with 30% of causes missing under two different missing at random scenarios $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ and four auxiliary association level settings (A0), (A1), (A2) and (A3), for sample size $n = 1200$ and bandwidth $h = 0.3$ based on 1000 Gaussian multiplier samples and 1000 simulation iterations

Model	(θ_1, θ_2)	Test	Under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$				Under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$			
			$T_{a1}^{(1)}$	$T_{a2}^{(1)}$	$T_{m1}^{(1)}$	$T_{m2}^{(1)}$	$T_{a1}^{(1)}$	$T_{a2}^{(1)}$	$T_{m1}^{(1)}$	$T_{m2}^{(1)}$
			Auxiliary association level setting (A0): Kendall's tau = 0							
M_1	(0, 0)	Size	0.051	0.054	0.060	0.052	0.055	0.059	0.058	0.053
M_2	(0, −0.6)	Power	0.565	0.534	0.673	0.650	0.575	0.539	0.681	0.658
M_3	(0, −0.9)		0.820	0.794	0.890	0.869	0.827	0.806	0.891	0.875
M_4	(0, −1.2)		0.936	0.936	0.964	0.961	0.940	0.939	0.965	0.964
			Auxiliary association level setting (A1): Kendall's tau = 0.45							
M_1	(0, 0)	Size	0.058	0.054	0.058	0.055	0.065	0.059	0.053	0.053
M_2	(0, −0.6)	Power	0.582	0.562	0.691	0.663	0.585	0.562	0.705	0.672
M_3	(0, −0.9)		0.842	0.818	0.909	0.882	0.838	0.817	0.899	0.877
M_4	(0, −1.2)		0.947	0.947	0.974	0.970	0.949	0.946	0.975	0.967
			Auxiliary association level setting (A2): Kendall's tau = 0.75							
M_1	(0, 0)	Size	0.053	0.051	0.058	0.057	0.057	0.055	0.059	0.052
M_2	(0, −0.6)	Power	0.608	0.585	0.720	0.693	0.618	0.582	0.725	0.700
M_3	(0, −0.9)		0.876	0.859	0.926	0.906	0.873	0.862	0.920	0.902
M_4	(0, −1.2)		0.969	0.960	0.984	0.978	0.966	0.964	0.984	0.978
			Auxiliary association level setting (A3): Kendall's tau = 0.9							
M_1	(0, 0)	Size	0.048	0.052	0.055	0.052	0.050	0.054	0.055	0.048
M_2	(0, −0.6)	Power	0.620	0.597	0.735	0.710	0.623	0.598	0.731	0.710
M_3	(0, −0.9)		0.885	0.873	0.938	0.918	0.891	0.869	0.936	0.922
M_4	(0, −1.2)		0.969	0.970	0.990	0.988	0.972	0.970	0.988	0.986

To further investigate our proposed method, we assess the performance of the AIPW estimators for four auxiliary association level settings under the $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ scenarios, shown in Figs. 3 and 4, respectively. The simulation results indicate that performance improves as the association strengthens.

We report the results for testing $H_{10} : \beta_1(t) = 0$ and $H_{20} : \beta_1(t)$ does not depend on t in Tables 1 and 2, respectively, where we take $t_1 = 0.3, t_2 = 1.7, t_1^* = 0.35$ for the test statistics. The results confirm that the empirical sizes are all close to their nominal level 0.05 for the four auxiliary association level settings under both $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ scenarios. The powers of the tests increase with sample size and with the strength of association between the auxiliary variable A_i and the cause V_i. The powers of the supremum type tests are comparable to the powers of the integrated tests for testing H_{10}, whereas the powers of the integrated tests are slightly higher than those of the supremum type tests for testing H_{20}.

Author's personal copy

Table 2 Empirical sizes and powers of the test statistics $T_{a1}^{(2)}$, $T_{a2}^{(2)}$, $T_{m1}^{(2)}$ and $T_{m2}^{(2)}$ for testing H_{20}: $\beta_1(t)$ does not depend on t for $t \in [t_1, t_2]$ at the nominal level 0.05 for models N_1 to N_4, with 30% of causes missing under two different missing at random scenarios $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$ and $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$ and four auxiliary association level settings (A0), (A1), (A2) and (A3), for sample size $n = 1200$ and bandwidth $h = 0.3$ based on 1000 Gaussian multiplier samples and 1000 simulation iterations

Model	(θ_1, θ_2)	Test	Under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z},\boldsymbol{A}}$				Under $\text{MAR}_{\boldsymbol{T},\boldsymbol{Z}}$			
			$T_{a1}^{(2)}$	$T_{a2}^{(2)}$	$T_{m1}^{(2)}$	$T_{m2}^{(2)}$	$T_{a1}^{(2)}$	$T_{a2}^{(2)}$	$T_{m1}^{(2)}$	$T_{m2}^{(2)}$
			Auxiliary association level setting (A0): Kendall's tau = 0							
N_1	(0, −0.5)	Size	0.049	0.051	0.057	0.053	0.049	0.047	0.054	0.048
N_2	(0.9, 0)	Power	0.436	0.504	0.560	0.655	0.435	0.501	0.560	0.655
N_3	(1.2, 0)		0.670	0.758	0.767	0.874	0.666	0.764	0.779	0.879
N_4	(1.5, 0)		0.866	0.930	0.928	0.967	0.876	0.938	0.935	0.973
			Auxiliary association level setting (A1): Kendall's tau = 0.45							
N_1	(0, −0.5)	Size	0.052	0.053	0.053	0.054	0.052	0.054	0.049	0.054
N_2	(0.9, 0)	Power	0.443	0.514	0.574	0.674	0.449	0.527	0.574	0.673
N_3	(1.2, 0)		0.690	0.776	0.783	0.886	0.685	0.778	0.789	0.887
N_4	(1.5, 0)		0.882	0.936	0.942	0.980	0.883	0.936	0.940	0.978
			Auxiliary association level setting (A2): Kendall's tau = 0.75							
N_1	(0, −0.5)	Size	0.047	0.051	0.052	0.051	0.050	0.049	0.051	0.050
N_2	(0.9, 0)	Power	0.469	0.537	0.581	0.693	0.465	0.548	0.588	0.697
N_3	(1.2, 0)		0.726	0.801	0.804	0.900	0.718	0.799	0.808	0.897
N_4	(1.5, 0)		0.900	0.948	0.958	0.981	0.904	0.947	0.954	0.978
			Auxiliary association level setting (A3): Kendall's tau = 0.9							
N_1	(0, −0.5)	Size	0.051	0.057	0.049	0.055	0.055	0.056	0.058	0.058
N_2	(0.9, 0)	Power	0.472	0.546	0.587	0.698	0.471	0.548	0.594	0.698
N_3	(1.2, 0)		0.724	0.811	0.830	0.906	0.732	0.807	0.823	0.908
N_4	(1.5, 0)		0.919	0.961	0.962	0.985	0.918	0.964	0.962	0.982

6 Analysis of the Mashi data

We apply the proposed methods to the Mashi clinical trial data. The Mashi trial was conducted among HIV-infected women and their infants to compare the effect of infant feeding strategy on two outcomes in live-born infants: HIV infection (through postnatal mother-to-child HIV transmission) and death (Thior et al. 2006). Twelve hundred HIV positive pregnant mothers were randomized to two infant feeding strategies: 6 months of breastfeeding and zidovudine for the infant (BF+AZT, 588 live-born infants) versus 12 months of formula feeding with zidovudine for the infant for the first month of life (FF, 591 live-born infants). All mothers were instructed to wean their infants between 5 and 6 months of age and were supplied free formula from 5 through 12 months of age to facilitate safe weaning. Infants were tested for HIV infection at birth, monthly until age 7 months, at age 9 months, and then every 3 months through age 18 months. We include in the analysis the subset of live-born infants with complete covariate information at delivery, which totals 1123 live-born infants out of the 1179 total live-

Author's personal copy

births. Of the 107 infants who died over the first 18 months of life, 28 infants died of an HIV-related cause, 21 infants died of an HIV-unrelated cause, and the cause of death was missing for 58 infants. A death is considered to be HIV-related if either the study clinicians deemed the death HIV-related, or the infant had at least one positive test result from the PCR assay used to test for HIV infection prior to death. On the other hand, a death is considered to be HIV-unrelated if the study clinician deemed the death unrelated to HIV/AIDS.

The Mashi study showed that the treatment effect of the randomized feeding strategy BF+AZT vs. FF on the risk of all-cause death varies over time. It is our interest to assess the treatment effect on HIV-related death with HIV-unrelated death as a competing risk. We consider the following cause-specific Cox models with time-varying coefficients

$$\lambda_k(t|Z) = \lambda_{k0}(t)\exp\{\beta_k(t)Z\}, \tag{11}$$

where Z is the feeding strategy (1 for BF+AZT and 0 for FF defined in the first section). Here we use $k = 1$ for HIV-related death and $k = 2$ for HIV-unrelated death. The Epanechnikov kernel is used in the analysis.

We use a model selection procedure to select the variables (A_1) that are relevant in predicting whether the cause of death (V) is observed, and what auxiliary variables (A_2) are informative about the distribution of V. Considering 20 covariates collected for babies or their mothers, we use logistic regression and all-subsets model selection (with criterion Mallows C_p) to select a model for predicting $r(\zeta, A_1)$ as in Sun et al. (2012). The chosen model includes the following covariates: the infant had birth weight < 2.5 kilograms, the second randomization assignments of mom/baby was switched from Placebo/Placebo to Placebo/Nevirapine during the trial due to the DSMB recommendation, log 10 plasma viral load level of the mom at delivery, the infant had AZT toxicity, and whether the baby was hospitalized with a severe adverse event. In addition, we use the binary covariate of whether the infant received HAART (highly active ART therapy) as an auxiliary variable A_2 for the failure cause V and estimate $\rho_k(\zeta, A_2)$ using a logistic regression model for $f(A_2|V = k)$.

Figure 5a and b shows the IPW and AIPW estimates and 95% pointwise confidence bands of $\beta_k(t)$ for $k = 1$ and 2. Figure 5c shows the AIPW estimates of the cumulative baseline hazard functions, $\widehat{\Lambda}_{A,k0}(t)$, for $k = 1$ and 2. Figure 5d shows estimates of the log hazard ratio for all-cause death using the method of Cai and Sun (2003). The estimations are evaluated over 40 evenly distributed grid points between 0 and 365 days, where the bandwidth $h = 365 \times 0.6 = 219$ days is chosen using the 5-fold cross-validation procedure. The IPW and the AIPW estimates of $\beta_k(t)$ are close to one another but more different in the early and later times of follow-up, especially for $k = 1$. The confidence bands around the IPW estimates are slightly wider as expected.

Figure 5 supports that BF+AZT had an effect on reducing HIV-related deaths compared to FF until about 6 months or 183 days, and BF+AZT also had an effect, albeit weaker, on reducing HIV-unrelated deaths compared to FF. However, after about 6 months, the data suggest that the risk of HIV-related death may have been higher for the BF+AZT arm compared to the FF arm, whereas this is not the case for HIV-unrelated death. Figure 5d supports that BF+AZT reduces the risk of all-cause death compared to FF before 6 months. However, it appears to elevate the risk after 6 months.

Author's personal copy

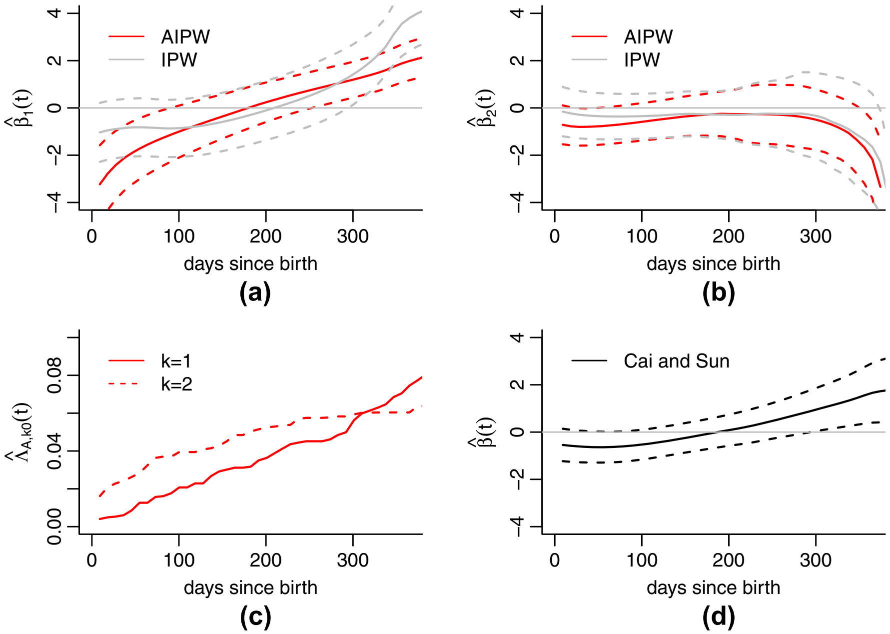

Fig. 5 Estimation of $\beta_1(t)$ and $\beta_2(t)$ with 95% pointwise confidence bands, and the cumulative baseline hazard functions $\Lambda_{A,k0}(t)$ $(k = 1, 2)$ for the Mashi randomized clinical trial with bandwidth $h = 219$. The estimates $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ of the log hazard ratio (BF+AZT / FF) are given in (a) for HIV-related death and in (b) for HIV-unrelated death. The estimated cumulative baseline hazard functions $\hat{\Lambda}_{A,k0}(t)$ are given in (c) for $k = 1$ and 2. The estimator $\hat{\beta}(t)$ of the log hazard ratio for all-cause death is given in (d) using the method of Cai and Sun (2003)

Figure 5a and d also indicate a lack of fit of the Cox model with constant coefficients for the first year of follow-up. The results are consistent with the original study results that showed that infants assigned to formula-feeding had a higher rate of all-cause mortality by the age of 7 months compared to those assigned to BF+AZT (Thior et al. 2006).

Next, we conduct formal hypothesis tests to examine the effects of the feeding strategy on the risks of HIV-related death and HIV-unrelated death. We test the null hypothesis H_{10}: $\beta_k(t) = 0$ for $t \in [t_1, t_2]$ against H_{1a}: $\beta_k(t) \neq 0$ for some $t \in [t_1, t_2]$ and H_{1m}: $\beta_k(t) \leq 0$ with strict inequality for some $t \in [t_1, t_2]$. We also test the null hypothesis H_{20} that $\beta_k(t)$ does not depend on t within $[t_1, t_2]$ against the alternatives H_{2a} that $\beta_k(t)$ changes over $t \in [t_1, t_2]$ and H_{2m} that $\beta_k(t)$ increases with $t \in [t_1, t_2]$, $k = 1, 2$. The tests are conducted for the time interval days since birth $[t_1, t_2] = [9, 365]$. Because infants in the BF+AZT group were given 6 months of breastfeeding and the mothers were instructed to wean their infants between 5 and 6 months of age and were supplied formula from 5 to 12 month of age, we also conducted some hypothesis tests for early days [9, 183] before the feeding change and for later days [183, 365] after the feeding change. The p-values given in Table 3 of the test statistics $T_{a1}^{(1)}$, $T_{a2}^{(1)}$, $T_{m1}^{(1)}$ and $T_{m2}^{(1)}$ for testing H_{10} and the test statistics $T_{a1}^{(2)}$,

Author's personal copy

Table 3 Observed p-values of the test statistics $T_{a1}^{(1)}, T_{a2}^{(1)}, T_{m1}^{(1)}, T_{m2}^{(1)}, T_{a1}^{(2)}, T_{a2}^{(2)}, T_{m1}^{(2)}$ and $T_{m2}^{(2)}$ for testing the log hazard ratio (BF+AZT/FF) for HIV-related death, $\beta_1(t)$, and for HIV-unrelated death, $\beta_2(t)$, over the time interval $t \in [t_1, t_2]$ under model (11) based on 1000 Gaussian multiplier samples with $h = 219$ and $t_1^* - t_1 = 9$ days. The tests are conducted for the time intervals $[t_1, t_2] = [9, 365]$, [9, 183] and [183, 365] (days since birth)

For HIV related death in $t \in [9, 365]$ *days since birth*							
$H_{10} : \beta_1(t) = 0$				$H_{20} : \beta_1(t)$ does not depend on t			
$H_{1a} : \beta_1(t) \neq 0$		$H_{1m} : \beta_1(t) \leq 0$		$H_{2a} : \beta_1(t)$ changes in t		$H_{2m} : \beta_1(t)$ increases in t	
$T_{a1}^{(1)}$	$T_{a2}^{(1)}$	$T_{m1}^{(1)}$	$T_{m2}^{(1)}$	$T_{a1}^{(2)}$	$T_{a2}^{(2)}$	$T_{m1}^{(2)}$	$T_{m2}^{(2)}$
0.160	0.113	0.081	0.050	< 0.001	< 0.001	< 0.001	< 0.001
For HIV unrelated death in $t \in [9, 365]$ *days since birth*							
$H_{10} : \beta_2(t) = 0$				$H_{20} : \beta_2(t)$ does not depend on t			
$H_{1a} : \beta_2(t) \neq 0$		$H_{1m} : \beta_2(t) \leq 0$		$H_{2a} : \beta_2(t)$ changes in t		$H_{2m} : \beta_2(t)$ increases in t	
$T_{a1}^{(1)}$	$T_{a2}^{(1)}$	$T_{m1}^{(1)}$	$T_{m2}^{(1)}$	$T_{a1}^{(2)}$	$T_{a2}^{(2)}$	$T_{m1}^{(2)}$	$T_{m2}^{(2)}$
0.211	0.211	0.106	0.085	0.554	0.621	0.261	0.391
For HIV related death in two different time intervals							
$H_{10} : \beta_1(t) = 0$, for $t \in [9, 183]$				$H_{10} : \beta_1(t) = 0$, for $t \in [183, 365]$			
$H_{1a} : \beta_1(t) \neq 0$		$H_{1m} : \beta_1(t) \leq 0$		$H_{1a} : \beta_1(t) \neq 0$		$H_{1m} : \beta_1(t) \geq 0$	
$T_{a1}^{(1)}$	$T_{a2}^{(1)}$	$T_{m1}^{(1)}$	$T_{m2}^{(1)}$	$T_{a1}^{(2)}$	$T_{a2}^{(2)}$	$T_{m1}^{(2)}$	$T_{m2}^{(2)}$
0.020	0.010	0.010	0.003	0.016	0.062	0.010	0.050

$T_{a2}^{(2)}, T_{m1}^{(2)}$ and $T_{m2}^{(2)}$ for testing H_{20} are calculated using $h = 219$, $t_1^* - t_1 = 9$ days and 1000 Gaussian multiplier samples. The first block of Table 3 shows the p-values of the tests for HIV-related death over the time interval [9, 365] days since birth. While the evidence against H_{10}: $\beta_1(t) = 0$ is weak, the data strongly suggests that the log hazard ratio (BF+AZT versus FF) for HIV-related death, $\beta_1(t)$, changes with time and that it increases with time since birth. Further hypothesis tests for two different time intervals shown in the third block of Table 3 support $\beta_1(t) < 0$ at some t before 6 months and $\beta_1(t) > 0$ for some t after 6 months. These conclusions from the testing results are further supported by the estimated log hazard ratio (BF+AZT versus FF) for HIV-related death, $\beta_1(t)$, given in Fig. 5a. The second block of Table 3 shows the p-values of the tests for HIV-unrelated death over the time interval [9, 365] days since birth. The data do not suggest that the log hazard ratio for HIV-unrelated death, $\beta_2(t)$, differs from zero, nor does it support that $\beta_2(t)$ varies with time.

The diagnostic plots Fig. 6a and c indicate that the observed test processes $D_1(t)$ and $D_2(t)$ for testing H_{10} and H_{20} for $\beta_1(t)$ deviate significantly from the 1000 random realizations from $D_1^*(t)$ and $D_2^*(t)$, respectively. On the other hand, Fig. 6b and d show that the deviations of the observed test processes for testing $\beta_2(t)$ from the null models are not significant. These diagnostic plots are consistent with the p-value results reported in Table 3.

The results of the analysis using the Gaussian kernel are similar, which are given in Appendix C of the Web-based Supplementary Material. We use the timecox() function

Author's personal copy

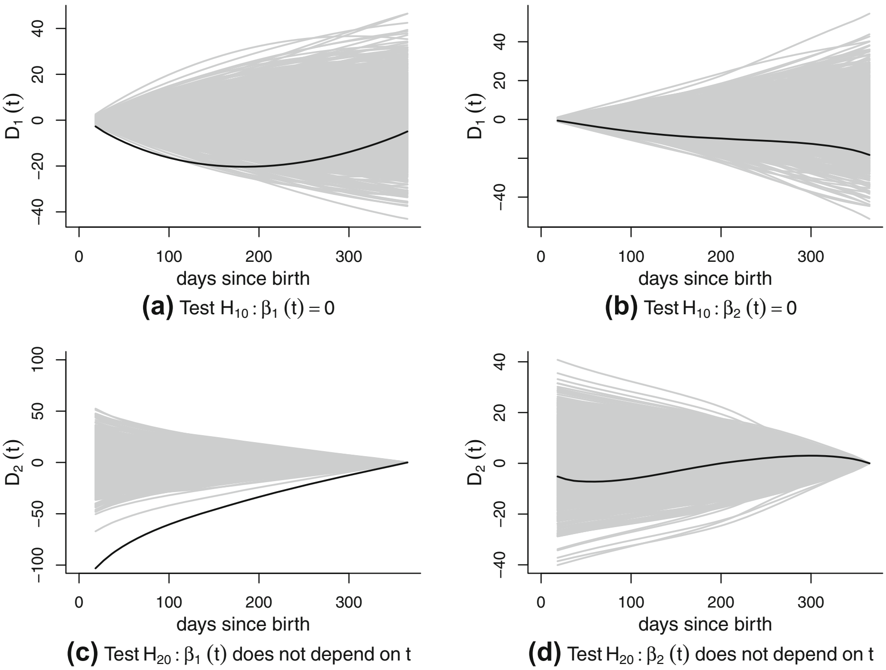

Fig. 6 Test processes $D_1(t)$ and $D_2(t)$ (black solid lines) for testing H_{10} and H_{20} for the log hazard ratio (BF+AZT/FF) for HIV-related death and HIV-unrelated death are plotted against 1000 random realizations from $D_1^*(t)$ and $D_2^*(t)$ (grey lines), respectively. **a** and **c** show the plots for testing H_{10} and H_{20} for $\beta_1(t)$, respectively. **b** and **d** show the plots for testing H_{10} and H_{20} for $\beta_2(t)$, respectively

in the timereg package in R to test for the time-invariant effect of all-cause death, cf. Martinussen and Scheike (2006). The p-values for the Kolmogorov-Smirnov test and Cramer von Mises test are 0.004 and 0.010, respectively, which support that the effect of feeding strategy on all-cause death varies with time.

7 Concluding remarks

In this paper, we have developed IPW and AIPW estimation methods for cause-specific hazard regression models with missing failure causes, where the Cox model with time-varying coefficients are utilized to examine cause-specific covariate effects. The AIPW estimator is a two-stage estimator by utilizing the IPW estimator and through modeling available auxiliary variables to improve efficiency. Both the IPW and AIPW estimation methods allow for the missingness of failure cause to depend on auxiliary variables that correlate with the cause of failure. Simulation studies demonstrate that the performance of the AIPW estimators improves as the association between the auxiliary variable and the cause of failure strengthens. Furthermore, the AIPW estimators are more efficient than the IPW estimators even when the auxiliary variables are not available due to the more efficient construction of the AIPW estimating equation. Most existing works model the conditional distribution of the cause of failure $\rho_k(\zeta_i, A_i)$ using a logistic

Author's personal copy

regression model, cf. Gao and Tsiatis (2005) and Lu and Liang (2008). However, correct modeling of $\rho_k(\zeta_i, A_i)$ is intricate. The proposed AIPW estimators utilize the expression (5) for $\rho_k(\zeta_i, A_i)$, which describes how the distribution of the cause of failure depends on the variables (T_i, Z_i, A_i) through the conditional cause-specific hazard functions and the conditional density function of the auxiliary variables. The proposed AIPW estimators do not possess the double robustness property because they use the first stage IPW estimators in the implementation. Nevertheless, since it is often dubious to model $\rho_k(\zeta_i, A_i)$ accurately, such a property is also unachievable for the regular AIPW estimators that model both the probability of a complete case $r(\zeta_i, A_i)$ and the conditional distribution of the cause of failure $\rho_k(\zeta_i, A_i)$.

There would be loss of efficiency by fitting the Cox model with time-varying effects when the Cox model with constant effects holds, as we show in the asymptotic results that the rate of convergence for $\widehat{\beta}_{A,k}(t)$ is at the order of $(nh)^{-1/2}$ while the rate of convergence of the estimated $\hat{\beta}$ under the Cox model is at the order of $n^{-1/2}$. One of the two hypothesis testing procedures developed in Sect. 4 for the two hypotheses H_{10}: $\beta_k(t) \equiv 0$ and H_{20}: $\beta_k(t) \equiv \beta_k$ can be used to evaluate whether the Cox model with constant effects is appropriate.

The proposed estimation and hypothesis testing procedures contribute to the analysis of the Mashi clinical trial data, for examining the randomized treatment effects on HIV-related and HIV-unrelated infant death, where the cause of death is frequently missing, and the treatment effects are demonstrated to vary over time. This manuscript, however, does not provide a formal testing procedure for whether the treatment effect is different against cause 1 failure than against cause 2 failure. The difference in the effects of feeding strategy on HIV-related death and on HIV-unrelated death is shown in Fig. 5a, b, where the fact that the estimated curves are quite different, and there are time periods during which the two sets of pointwise confidence bands do not overlap, suggests potential differences between the failure causes. However, as the reviewer notes, this does not formally imply a difference; for that one would need simultaneous confidence bands about the difference $\beta_1(t) - \beta_2(t)$ excluding 0 for at least one t, and this manuscript does not study such simultaneous confidence bands. Therefore, we only have an informal suggestion of a potential difference, without a formal inference backing it up. An alternative and complementary approach to assessing the effect of feeding strategy on HIV infection and death would use an illness-death model, where HIV infection, as the “illness”, is subject to interval censoring. It would be worthwhile to study the illness-death model using the Cox model with time-varying coefficients for this situation, which would provide a different approach to investigating the scientific question in the Mashi study.

Acknowledgements We thank the reviewers for their constructive comments that have improved the contents and the presentation of the paper. The authors also thank the Mashi study team (led by Dr. Max Essex) and the Mashi study participants. Research reported in this publication was partially supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R37 AI054165. The research of Yanqing Sun was also partially supported by National Science Foundation grant DMS-1513072 and DMS-1915829. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author's personal copy

Appendix

This Appendix introduces the notations and presents the conditions for the asymptotic results presented in Theorems 1–5.

Let $\mathcal{F}_t$ be the right continuous filtration generated by the data processes $\{N_{ik}(s), Y_i(s), Z_i(s); i = 1, \ldots, n, k = 1, 2, \ldots, L, 0 \le s \le t\}$. Assume $E(dN_{ik}(t) = 1|\mathcal{F}_{t-}) = E(dN_{ik}(t) = 1|Y_i(t), Z_i(t)) = Y_i(t)\lambda_{ik}(t|Z_i(t))dt$. It follows that $M_{ik}(t) = N_{ik}(t) - \int_0^t Y_i(u)\lambda_k(u|Z_i(u))du$, $i = 1, \ldots, n, k = 1, 2, \ldots, L$, are multivariate orthogonal martingales with respect to $\mathcal{F}_t$ (Aalen and Johansen 1978). To accommodate additional information introduced due to missing data, we define the augmented filtration $\mathcal{F}_t^*$ generated by the data processes $\{N_{ik}(s), Y_i(s), Z_i(s), R_i, \delta_i A_i; i = 1, \ldots, n, k = 1, 2, \ldots, L, 0 \le s \le t\}$. Let $\lambda_{ik}^*(t)\, dt = P\{T_i \in [t, t + dt), V_i = k|X_i \ge t, Z_i(t), R_i, \delta_i A_i\}$. Then $Y_i(t)\lambda_{ik}^*(t)$ is the intensity of $N_{ik}(t)$ with respect to $\mathcal{F}_t^*$, and $M_{ik}^*(t) = N_{ik}(t) - \int_0^t Y_i(u)\lambda_{ik}^*(u)du$, $i = 1, \ldots, n, k = 1, 2, \ldots, L$, are multivariate orthgonal martingales with respect to $\mathcal{F}_t^*$.

Let $S^{(j)}(t, \beta_k) = n^{-1}\sum_{i=1}^n Y_i(t)\exp\left(\beta_k(t)^\top Z_i(t)\right) Z_i(t)^{\otimes j}$, and $S_I^{*(j)}(t, \beta_k, \psi) = n^{-1}\sum_{i=1}^n q_i\ Y_i(t)\exp\left(\beta_k(t)^\top Z_i(t)\right) Z_i(t)^{\otimes j}$, for $k = 1, \ldots, L$ and $j = 0, 1, 2$. Let $s^{(j)}(t, \beta_k) = ES^{(j)}(t, \beta_k)$ and $s_I^{*(j)}(t, \beta_k, \psi) = ES_I^{*(j)}(t, \beta_k, \psi)$. If the model $r(\zeta_i, A_i, \psi)$ is correctly specified, then $s^{(j)}(t, \beta_k) = s_I^{*(j)}(t, \beta_k, \psi)$. Define $\Sigma_k(t) = \left[s^{(2)}(t, \beta_k) - \left(s^{(1)}(t, \beta_k)\right)^{\otimes 2}/\ s^{(0)}(t, \beta_k)\right]\lambda_{k0}(t)$ and $\Sigma_k^*(t) = E\left[\left(Z_i(t) - s^{(1)}(t, \beta_k)/s^{(0)}(t, \beta_k)\right)^{\otimes 2} R_i\pi^{-2}(Q_i)Y_i(t)\lambda_{ik}^*(t)\right]$.

Let S_i^ψ and I^ψ be the score vector and information matrix for $\widehat{\psi}$ under (4). Then,

$$S_i^\psi = \frac{\delta_i(R_i - r(\zeta_i, A_i, \psi_0))}{r(\zeta_i, A_i, \psi_0)(1 - r(\zeta_i, A_i, \psi_0))}\frac{\partial r(\zeta_i, A_i, \psi_0)}{\partial\psi},$$

$$I^\psi = E\left\{\frac{\delta_i}{r(\zeta_i, A_i, \psi_0)(1 - r(\zeta_i, A_i, \psi_0))}\frac{\partial r(\zeta_i, A_i, \psi_0)}{\partial\psi}\left(\frac{\partial r(\zeta_i, A_i, \psi_0)}{\partial\psi}\right)^\top\right\},$$

and $\widehat{\psi} - \psi = n^{-1}\sum_{i=1}^n (I^\psi)^{-1}S_i^\psi + o_p(n^{-1/2})$, where ψ_0 is the true value of ψ. We also define the following notations:

$$\mathcal{A}_i(t, \beta_k) = \int_0^\tau K_h(u - t)H^{-1}\left(Z_i(u) - \frac{s^{(1)}(u, \beta_k)}{s^{(0)}(u, \beta_k)}\right)q_{i0}\, dM_{ik}(u),$$

$$\mathcal{B}_i(t, \beta_k) = \int_0^\tau K_h(u - t)H^{-1}\left(Z_i(u) - \frac{s^{(1)}(u, \beta_k)}{s^{(0)}(u, \beta_k)}\right)(1 - q_{i0})\ E(dM_{ik}(u)|Q_i),$$

$$\mathcal{D}^n(t, \beta_k) = n^{-1}\sum_{i=1}^n \int_0^\tau K_h(u - t)\left(Z_i(u) - \frac{s^{(1)}(u, \beta_k)}{s^{(0)}(u, \beta_k)}\right)\frac{-R_i}{(\pi(Q_i, \psi_0))^2}$$

Author's personal copy

$$\left(\frac{\partial \pi(Q_i, \psi_0)}{\partial \psi}\right)^{\mathsf{T}} dM_{ik}(u),$$

$$\mathcal{O}_i(t, \beta_k) = \mathcal{D}^n(t, \beta_k)(I^{\psi})^{-1} S_i^{\psi}.$$

The following conditions are assumptions we use to prove the theorems:

(C.1) For $k = 1, \ldots, L$, $\beta_k(t)$ has componentwise second derivatives on $[0, \tau]$. The sample path of the covariate process $Z_i(t)$ is left continuous and of bounded variation, and satisfies the moment condition $E[||Z_i(t)||^4 \exp(2M||Z_i(t)||)] < \infty$, where M is a constant such that $(t, \beta_k(t)) \in [0, \tau] \times [-M, M]^p$ for all t and $||A|| = \max_{k,l} |a_{kl}|$ for a matrix $A = (a_{kl})$.

(C.2) The kernel function $K(\cdot)$ is bounded and symmetric with bounded support $[-1, 1]$. The bandwidth h satisfies $nh^2 \to \infty$ and nh^5 is bounded as $n \to \infty$.

(C.3) The matrix $\Sigma_k(t)$ is positive definite for all $t \in [0, \tau]$.

(C.4) For $k = 1, \ldots, L$ and for $j = 0, 1, 2$, the functions $s^{(j)}(t, \beta_k)$ and $s_I^{*(j)}(t, \beta_k, \psi)$ are componentwise continuous on $t \in [0, \tau]$, $\beta_k \in [-M, M]^p$, $\psi \in \Theta_\psi$, where Θ_ψ is a compact set. $\sup_{t\in[0,\tau],\beta_k\in[-M,M]^p} ||S^{(j)}(t, \beta_k) - s^{(j)}(t, \beta_k)|| = O_p(n^{-1/2})$, and $\sup_{t\in[0,\tau],\beta_k\in[-M,M]^p,\psi\in\Theta_\psi} ||S_I^{*(j)}(t, \beta_k, \psi) - s_I^{*(j)}(t, \beta_k, \psi)|| = O_p(n^{-1/2})$.

(C.5) The function $r(\zeta_i, A_i, \psi)$ is twice differentiable with respect to ψ on a compact set Θ_ψ, $r'(\zeta_i, A_i, \psi) = \partial r(\zeta_i, A_i, \psi)/\partial\psi$ is uniformly bounded, and there is an $\varepsilon > 0$ such that $r(\zeta_i, A_i, \psi) \geq \varepsilon$ for all i. The function $f(A_i|k, T_i, Z_i, \varphi_k)$ is also twice differentiable with respect to φ_k on a compact set Θ_{φ_k} for $k = 1, \ldots, L$.

Supplementary materials

The Web-based Supplementary Materials referenced in the manuscript are available with this paper at the journal's online website.

References

Aalen OO, Johansen S (1978) An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand J Stat 5:141–150

Cai Z, Sun Y (2003) Local linear estimation for time-dependent coefficients in Cox's regression models. Scand. J Stat 30:93–111

Clemens JD, Sack DA, Harris JR et al (1990) Field trial of oral cholera vaccines in Bangladesh: results from three-year follow-up. Lancet 335:270–273

Efromovich S (2010) Dimension reduction and adaptation in conditional density estimation. J Am Stat Assoc 105:761–774

Fan J, Gijbels I (1996) Local polynomial modelling and its applications: monographs on statistics and applied probability 66, 1st edn. Chapman and Hall/CRC, New York

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

Author's personal copy

Gilbert P, McKeague I, Sun Y (2008) The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy. Biostatistics 9(2):263–276

Gilbert P, Sun Y (2015) Inferences on relative failure rates in stratified mark-specific proportional hazards models with missing marks, with application to human immunodeficiency virus vaccine efficacy trials. J R Stat Soc Ser C (Appl Stat) 64(1):49–73

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

Hall P, Racine JS, Li Q (2004) Cross-validation and the estimation of conditional probability densities. J Am Stat Assoc 99:1015–1026

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

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

Izbicki R, Lee AB (2016) Nonparametric conditional density estimation in a high-dimensional regression setting. J Comput Gr Stat 25(4):1297–1316

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

Liu L, Nevo D, Nishihara R, Cao Y, Song M, Twombly T, Chan A, Giovannucci E, VanderWeele T, Wang M, Ogino S (2018) Utility of inverse probability weighting in molecular pathological epidemiology. Eur J Epidemiol 33(4):381–392

Lu W, Liang Y (2008) Analysis of competing risks data with missing cause of failure under additive hazards model. Stat Sin 19:219–234

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

Lu K, Tsiatis A (2005) Comparison between two partial likelihood approaches for the competing risks model with missing cause of failure. Lifetime Data Anal 11:29–40

Martinussen T, Scheike TH, Skovgaard IM (2002) Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. Scand J Stat 29:59–77

Martinussen T, Scheike T (2006) Dynamic regression models for survival data. Springer, New York

Murphy SA, Sen PK (1991) Time-dependent coefficients in a Cox-type regression model. Stoch Process Appl 39:153–180

Nevo D, Nishihara R, Ogino S, Wang M (2018) The competing risks Cox model with auxiliary case covariates under weaker missing-at-random cause of failure. Lifetime Data Anal 24:425–442

Rice JA, Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. J R Stat Soc Ser B 53:233–243

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

Rubin D (1976) Inference and missing data. Biometrika 63:581–592

Scharfstein DO, Rotnitzky A, Robins JM (1999) Adjusting for nonignorable drop-out using semiparametric nonresponse models: rejoinder. J Am Stat Assoc 94:1135–1146

Sun Y, Gilbert PB (2012) Estimation of stratified mark-specific proportional hazards models with missing marks. Scand J Stat 39:34–52

Sun Y, Hyun S, Gilbert PB (2008) Testing and estimation of time-varying cause-specific hazard ratios with covariate adjustment. Biometrics 64:1070–1079

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

Sun Y, Sundaram R, Zhao Y (2009) Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood. Scand J Stat 36:444–462

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

Sun Y, Wu H (2005) Semiparametric time-varying coefficients regression model for longitudinal data. Scand J Stat 32:21–47

Thior, I., Lockman, S., Smeaton, L.M., Shapiro, R.L., Wester, C., Heymann, S.J., Gilbert, P.B., Stevens, L., Peter, T., Kim, S., van Widenfelt, E., Moffat, C., Ndase, P., Arimi, P., Kebaabetswe, P., Mazonde, P., Makhema, J., McIntosh, K., Novitsky, V., Lee, T.H., Marlink, R., Lagakos, S., Essex M. and the Mashi Study Team (2006) Breastfeeding plus infant zidovudine prophylaxis for 6 months vsformula

Author's personal copy

feeding plus infant zidovudine for 1 month to reducemother-to-child HIV transmission in Botswana: a randomized trial:the Mashi study. J. Am. Stat Medical Assoc 296: 794–805
Tian L, Zucker D, Wei LJ (2005) On the Cox model with time-varying regression coefficients. J Am Stat Assoc 100:172–183
van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge
Zucker DM, Karr AF (1990) Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. Ann Stat 18:329–353

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Fei Heng[1] · Yanqing Sun[2] · Seunggeun Hyun[3] · Peter B. Gilbert[4]

✉ Yanqing Sun
yasun@uncc.edu

1 Department of Mathematics and Statistics, University of North Florida, Jacksonville, FL 32224, USA

2 Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

3 Division of Mathematics and Computer Science, University of South Carolina Upstate, Spartanburg, SC 29303, USA

4 University of Washington and Fred Hutchinson Cancer Research Center Seattle, Seattle, WA 98109, USA