Semiparametric Estimation in the Proportional Subdistribution Hazards Model with Missing Cause of Failure

Jonathan Yabes* Chung-Chou H. Chang†

Abstract

In analyses involving competing risks, the proportional subdistribution hazards regression model of Fine and Gray is commonly used to estimate covariate effects of specific risk factors for disease. In some situations however, the actual cause of failure may be unknown or missing. To avoid bias, we develop two semiparametric estimators of covariate effects: the inverse probability weighted complete-case estimator and the augmented inverse probability weighted estimator. We study the properties of these estimators analytically and use simulations to compare their finite sample size performance to that of estimators obtained via a multiple imputation method, a naïve complete-case analysis, and a method in which missing cases are treated as an extra failure type. We employ the proposed methods to estimate the effects that several risk factors have on the development of coronary heart disease or the occurrence of death related to this disease among individuals infected with the human immunodeficiency virus (HIV).

KEY WORDS : Competing risks; doubly-robust; inverse probability weighting; missing cause of failure; proportional subdistribution hazards.

*Department of Biostatistics, University of Pittsburgh
†Departments of Medicine and Biostatistics, University of Pittsburgh

1. Introduction

In biomedical studies involving competing risks, it is not uncommon to have the cause of failure be unknown or missing for some individuals. This occurs when the information about the cause responsible for the failure was poorly documented or handled, or when the diagnosis of the cause of death is difficult for some patients (Andersen et al., 1996). Heuristic and nonheuristic approaches have been used to deal with this problem. Heuristic approaches involve either a complete-case analysis (CC) or an analysis in which an extra category (EC) is formed for the missing cases. Albeit convenient, these methods could lead to substantial bias (Bakoyannis et al., 2010).

Many of the nonheuristic approaches have been developed under the framework of cause-specific hazards (CSH) regression. Under a proportional hazards assumption, Goetghebeur and Ryan (1995) proposed a model relating the CSH of the event of interest to that of the competing events, while Lu and Tsiatis (2001) suggested a multiple imputation (MI) approach. Using the inverse probability weighted (IPW) and augmented IPW (AIPW) estimators, Gao and Tsiatis (2005) focused on linear transformation models, whereas Lu and Liang (2008) worked on the additive hazards model. All of these strategies assume that the cause of failure is missing at random (MAR). Unfortunately, these methods are not directly applicable to modeling the subdistribution, which is also known as the cumulative incidence function.

The subdistribution is frequently used by clinical researchers because its interpretation is straightforward. However, if we wish to model it directly in settings where the cause of failure is not observed for all subjects, the available methods are relatively limited. Recently, Bakoyannis et al. (2010) adopted an MI approach under the proportional subdistribution hazards model of Fine and Gray (1999); their simulations showed that this technique outperforms the heuristic methods with respect to bias under different missingness assumptions.

In this paper, our objective is to apply semiparametric theory to derive two additional estimators for proportional subdistribution hazards regression with missing cause of failure. By using inverse weighting and augmented inverse weighting, the proposed estimators yield theoretically valid estimates under the MAR assumption. We study their properties and conduct simulations to compare their performance with that of estimators obtained via the MI method of Bakoyannis et al. (2010), via a naïve complete-case analysis, and via a method in which an extra category is formed for the missing cases.

The motivation of this work ensues from data derived from a national cohort study on the development of coronary heart disease (CHD) among individuals infected with the human immunodeficiency virus (HIV). Interest lies in examining the effects that several risk factors have on the development of CHD or the occurrence of death related to CHD. In this setting, CHD-unrelated death constitutes a competing event. However for some subjects who died, the cause of death cannot be confirmed to be CHD-related or not.

This article is organized as follows. In Section 2, we introduce some notations and revisit the proportional subdistribution hazards model. In Sections 3 and 4, we present the proposed IPW estimator and AIPW estimator, respectively, and discuss their asymptotic properties. In Section 5, we assess and compare through simulations the performance of the two estimators with that of the current approaches. In Section 6, we apply the proposed methods to analyze the CHD data, then conclude in the final section.

2. Notations and Model

Suppose we have n independent individuals in a study. Let T be the failure time and ε be the failure type. Without loss of generality, consider two failure types where $\varepsilon = 1$ represents the cause of interest, and $\varepsilon = 2$ the competing event. When there are no missing cases, for each subject i we observe $\{X_i, \varepsilon_i, \boldsymbol{Z}_i\}$, where $X_i = T_i \wedge C_i$; C_i is the censoring time assumed to be independent of T_i; and $\boldsymbol{Z}_i$ is a p-dimensional vector covariates. For simplicity, we restrict $\boldsymbol{Z}$ to be fixed covariates although the results can be extended to external time-dependent covariate processes (Kalbfleisch and Prentice, 2002). Here we define $\varepsilon_i = 0$ whenever $X_i = C_i$. The proportional sudistribution hazards model of Fine and Gray (1999) has the form

$$\lambda_1(t|\boldsymbol{Z}) = \lambda_{10}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}), \tag{1}$$

where the baseline subdistribution hazards $\lambda_{10}(\cdot)$ is left unspecified. This reminds us of the regular Cox model except that it models the event-1 subdistribution hazards $\lambda_1(\cdot)$ rather than the marginal hazards. The corresponding subdistribution can be calculated as $F_1(t|\boldsymbol{Z}) = 1 - \exp\left\{-\int_0^t \lambda_{10}(u)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\, du\right\}$.

The regression coefficients are estimated through a partial likelihood approach with modified risk sets. In particular, when there is no censoring, the risk set at time t includes both individuals who have yet to fail from the event of interest and those that have already failed from the competing cause prior to time t. When only administrative censoring is present, that is, the potential censoring time is known for all individuals, the risk sets are redefined such that an individual who failed from the competing event remains at risk only up to his potential censoring time. In the presence of random right censoring, the inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky, 1992) is utilized to reweight subjects who experienced the competing event. More formally, by defining at time t the counting process $N_i(t) = I(T_i \leqslant t, \varepsilon_i = 1)$, and the at risk process $Y_i(t) = I(T_i < t, \varepsilon_i = 2) + I(T_i \geqslant t)$, the reweighted at risk process for subject i is given by

$$\begin{aligned} \omega_i(t) &= I(C_i \geqslant T_i \wedge t)\frac{\hat{G}(t)}{\hat{G}(T_i \wedge t)}Y_i(t) \\ &= \frac{\hat{G}(t)}{\hat{G}(X_i)}I(X_i < t, \varepsilon_i = 2) + I(X_i \geqslant t), \end{aligned} \tag{2}$$

where $\hat{G}(t)$ is the Kaplan-Meier estimator of the censoring survival distribution. The weighted partial likelihood score equation takes the form

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\tau \left\{ \boldsymbol{Z}_i - \frac{\sum_j \omega_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j) \boldsymbol{Z}_j}{\sum_j \omega_j(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)} \right\} \omega_i(t) dN_i(t), \tag{3}$$

where $\tau = sup\{t : \Pr(\omega(t) \geqslant \epsilon > 0) > 0\}$. The estimator $\hat{\boldsymbol{\beta}}$ can be obtained by finding the root of $\boldsymbol{U}(\boldsymbol{\beta})$. It has been shown that $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal for the true parameter value $\boldsymbol{\beta}_0$ (Fine and Gray, 1999; Geskus, 2011).

3. Inverse Probability Weighted Estimator

3.1. *Estimating Equations*

The estimation procedure discussed above becomes inadequate when the cause of failure is not observed for all subjects. Analogous to the Horvitz-Thompson inverse selection probability technique (Horvitz and Thompson, 1952), we propose an IPW score equations for such settings. The basic idea is to recreate a random sample of the population to correct the selection bias that might have been induced by the missingness process. This is accomplished by upweighting subjects that are less likely to be observed based on some background characteristics. To make this approach plausible, we assume that the failure type is MAR, that is, the missingness mechanism depends only on fully observed quantities and not on the unobserved ε. This assumption has been widely employed in the missing data literature and is also the basis of many other missing data methods such as MI.

Let R_i be the complete-data indicator, that is, $R_i = 1$ if ε_i is observed and $R_i = 0$ if it is missing. We implicitly assume that censoring status is always observed so that we take $R_i = 1$ whenever $\varepsilon_i = 0$. Auxiliary covariates, denoted by $\boldsymbol{A}_i$, may also have been collected for each subject. These variables are not used to model the subdistribution but may be needed to fully account for the missingness process. The MAR assumption gives

$$\Pr(R_i = 1 | \varepsilon_i, \varepsilon_i > 0, \boldsymbol{W}_i) = \Pr(R_i = 1 | \varepsilon_i > 0, \boldsymbol{W}_i) \stackrel{set}{=} \pi_0(\boldsymbol{W}_i),$$

where $\boldsymbol{W}_i = (X_i, \boldsymbol{Z}_i, \boldsymbol{A}_i)$ are the fully observed variables, that is, R and ε are independent conditional on $\boldsymbol{W}$. This coupled with (3) leads us to consider the following IPW partial likelihood equations:

$$\boldsymbol{U}_w(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i)} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t) \right\} \omega_i(t) dN_i(t), \tag{4}$$

where $\pi(\boldsymbol{W}_i) = \Pr(R_i = 1|\boldsymbol{W}_i) = \pi_0(\boldsymbol{W}_i)I(\varepsilon_i > 0) + I(\varepsilon_i = 0)$; $\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t) = S_w^{(1)}(\boldsymbol{\beta}, t)/S_w^{(0)}(\boldsymbol{\beta}, t)$; and for $k = 0, 1, 2$, $S_w^{(k)}(\boldsymbol{\beta}, t) = n^{-1}\sum_j R_j \pi^{-1}(\boldsymbol{W}_j)\omega_j(t)\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}$, with $\boldsymbol{a}^{\otimes 0} = 1, \boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. By finding the root of $\boldsymbol{U}_w(\boldsymbol{\beta})$, we obtain the IPW estimator $\hat{\boldsymbol{\beta}}^w$ under the true $\pi(\boldsymbol{W})$.

Before we present the asymptotic properties of $\hat{\boldsymbol{\beta}}^w$, a few more notations are needed. Let $N^c(t) = \omega(t)N(t)$ be the counting process that incorporates the censoring information, and let $M^c(t) = N^c(t) - \int_0^t \omega(u)\exp(\boldsymbol{\beta}^T \boldsymbol{Z})d\Lambda_{10}(u)$. In addition, let $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t) = E\{\omega(t)\exp(\boldsymbol{\beta}^T \boldsymbol{Z})\boldsymbol{Z}^{\otimes k}\}$ for $k = 0, 1, 2$; $\boldsymbol{e}(\boldsymbol{\beta}, t) = \boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)/s^{(0)}(\boldsymbol{\beta}, t)$;

$$\boldsymbol{v}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} \right\}^{\otimes 2}; \quad \text{and}$$

$$\mathbf{M} = \int_0^{\tau} \{\boldsymbol{Z} - \boldsymbol{e}(\boldsymbol{\beta}, t)\} dM^c(t).$$

Geskus (2011) showed that $M^c(t)$ is a martingale. Note that $dN_i^c(t)$ can replace $\omega_i(t)dN_i(t)$ in (4) because $\omega_i(t) = 1$ whenever $N_i(t)$ has a jump.

THEOREM 1. *Under the regularity conditions given in the Appendix, $\hat{\boldsymbol{\beta}}^w$ is consistent for the true parameter $\boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}^w - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1}\Xi_w\boldsymbol{\mathcal{I}}_\beta^{-1})$, where $\Xi_w = E\{\pi^{-1}(\boldsymbol{W})\mathbf{M}^{\otimes 2}\}$ and $\boldsymbol{\mathcal{I}}_\beta = \int_0^\infty \boldsymbol{v}(\boldsymbol{\beta}, t)s^{(0)}(\boldsymbol{\beta}, t)d\Lambda_{10}(t)$.*

$\boldsymbol{\mathcal{I}}_\beta$ and Ξ_w can be consistently estimated by, respectively, $\hat{\boldsymbol{\mathcal{I}}}_\beta = -\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_w(\hat{\boldsymbol{\beta}}^w)$ and

$$\hat{\Xi}_w = \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(\boldsymbol{W}_i)} \hat{\mathbf{M}}_{w,i}(\hat{\boldsymbol{\beta}}^w) \right\}^{\otimes 2},$$

where $\hat{\mathbf{M}}_{w,i}(\boldsymbol{\beta}) = \int_0^\tau \{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, t)\} d\hat{M}_i^c(t)$, $d\hat{M}_i^c(t) = dN_i^c(t) - \omega_i(t)\exp(\hat{\boldsymbol{\beta}}^{wT}\boldsymbol{Z}_i)d\hat{\Lambda}_{10}(t)$, and $d\hat{\Lambda}_{10}(t) = [nS_w^{(0)}(\hat{\boldsymbol{\beta}}^w, t)]^{-1}\sum_j R_j \pi^{-1}(\boldsymbol{W}_j)dN_j^c(t)$.

In most situations, $\pi_0(\boldsymbol{W})$ is unknown and must be estimated from the data that are observed for everyone. To avoid the curse of dimensionality, we posit a parametric model $\pi_0(\boldsymbol{W};\boldsymbol{\gamma})$ for $\pi_0(\boldsymbol{W})$, where $\boldsymbol{\gamma}$ is a finite dimensional parameter. Although other parametric models can be employed, a logistic regression model of the form $\pi_0(\boldsymbol{W};\boldsymbol{\gamma}) = 1/(1 + \exp\{-\boldsymbol{\gamma}^T\widetilde{\boldsymbol{W}}\})$ with $\widetilde{\boldsymbol{W}} = (1, \boldsymbol{W}^T)^T$ is often adopted since R is binary. Under correct model specification for $\pi_0(\boldsymbol{W};\boldsymbol{\gamma})$, a consistent estimator $\hat{\boldsymbol{\gamma}}$ of the true value $\boldsymbol{\gamma}_0$ of $\boldsymbol{\gamma}$ may be obtained via maximum likelihood estimation (MLE). Equation (4) then becomes

$$\boldsymbol{U}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^n \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \left\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)\right\} dN_i^c(t), \tag{5}$$

where $\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t) = S_w^{(1)}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)/S_w^{(0)}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)$ and for $k = 0, 1, 2$, $S_w^{(k)}(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}, t)$ has the same form as $S_w^{(k)}(\boldsymbol{\beta}, t)$ but with $\pi(\boldsymbol{W}_j)$ replaced by $\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})$. The solution to $\boldsymbol{U}_w(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}) = 0$ gives the IPW estimator $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ when π is estimated.

THEOREM 2. *Under the regularity conditions (see Appendix) and if $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ is correctly specified, $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ is consistent for $\boldsymbol{\beta}_0$ and has the same asymptotic distribution as $\hat{\boldsymbol{\beta}}^w$.*

Theorem 2 implies that, unlike other settings in which IPW is used, we gain no efficiency by using the estimated rather than the true (if known) complete-data probabilities as inverse weights. The reason behind this is that the MAR assumption renders $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ predictable, leaving the covariance between the terms in the score equations for estimating $\boldsymbol{\beta}$ and that of $\boldsymbol{\gamma}$ equal to 0. Thus, to obtain a consistent estimator of the asymptotic variance of $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$, we simply replace $\pi(\boldsymbol{W}_i)$ with its estimate $\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})$ and $\hat{\boldsymbol{\beta}}^w$ with $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ in computing $\hat{\boldsymbol{\mathcal{I}}}_\beta$ and $\hat{\boldsymbol{\Xi}}_w$. This result conveniently enables us to carry out the IPW method in standard software, which normally accepts fixed weights as input, and still get valid standard errors even when the weights were estimated. Details on the implementation are described next.

3.2. *Implementation of the Inverse Weighting Method*

Model (1) for data with complete information on the cause of failure can be fitted several ways using standard software. In R, `crr` in the `cmprsk` library (Gray, 2010) can conveniently be used. However, it does not accommodate weighted analysis. More flexible approaches that involve some data preprocessing to permit the use of standard Cox regression commands (e.g., `coxph` in `R` or `proc phreg` in `SAS`) have been suggested. One method involves multiply-imputing censoring times for competing events using the Kaplan-Meier (Ruan and Gray, 2008). Another suggests explicitly setting up the data to include time-varying IPCW weights for those that experienced any of the competing events (Geskus, 2011). The advantage of the latter is that it is amenable to left-censored data.

We adopt the second approach in the discussion below. The basic idea is to run a Cox regression on the event of interest with case weights equal to the product of the IPCW and IPW, but with the standard errors computed using the robust sandwich estimator. The following steps can be followed:

1. Set the IPW weights for the censored subjects to `w.ipw` $= 1$. For the uncensored subjects, fit a logistic model to get $\pi_0(\boldsymbol{W}_i; \hat{\boldsymbol{\gamma}})$ and set `w.ipw` $= 1/\pi_0(\boldsymbol{W}_i; \hat{\boldsymbol{\gamma}})$.
2. Obtain the Kaplan-Meier estimate $\hat{G}(t)$ of the censoring distribution.
3. Restrict the data to the complete cases. Table 1(a) shows an example of an artificial dataset for a subset of 5 subjects after performing the previous steps. Suppose the unique event-1 failure times are $t_1 < t_2 < \ldots < t_D$. We then need to restructure the data into the counting process style of input. That is, create the variable `Tstart` to represent the start of the time interval and the variable `Tstop` to mark the end of the interval. A subject i who is censored or who experienced event-1 will have one record in this dataset, with `Tstart` $= 0$ and `Tstop` $= X_i$. For a subject i who experienced event-2, the record will be expanded to D rows, with the first row representing `Tstart` $= 0$ and

Tstop $= t_1$, the second representing Tstart $= t_1$ and Tstop $= t_2$, and so forth. Next, the IPCW weights of censored or event-1 subjects will be set to 1, i.e., w.cens $= 1$. For event-2 subjects, set w.cens $= 1$ if $X_i >$ Tstop and w.cens $= \hat{G}(\texttt{Tstop})/\hat{G}(X_i)$ if $X_i <$ Tstop. Table 1(b) illustrates how the data would look like after this step.

4. Fit a Cox model for the main event using case weights equal to w.ipw $*$ w.cens and use a robust variance estimator for the standard errors. For instance in R, we would fit model (1) for the data in Table 1(b) using the command

```
coxph(Surv(Tstart,Tstop,etype==1) ~ Z + cluster(subject),
   data=wData, weight=w.ipw*w.cens).
```

4. Augmented Inverse Probability Weighted Estimator

Although the IPW estimator is consistent when π is correctly specified, it may perform poorly otherwise. Moreover, it may lose efficiency by only using the complete cases. We can potentially improve these limitations by adopting the augmented inverse probability weighting (AIPW) technique developed by Robins et al. (1994). This method augments the IPW score equations with a term that uses information from both the complete cases and missing cases. We continue to assume MAR and begin by considering the probability of observing an event-1 failure conditional on the observed data:

$$\rho(\boldsymbol{W}_i) = \Pr\{\varepsilon_i = 1 | \boldsymbol{W}_i, \varepsilon_i > 0\}.$$

We can think of this as an imputation model for $\varepsilon = 1$. Now define $N_i^{\varepsilon}(t) = I(T_i \leqslant t)I(\varepsilon_i > 0)$. Notice that we can write $N_i(t) = I(\varepsilon_i = 1)N_i^{\varepsilon}(t)$ and $Y_i(t) = I(\varepsilon_i = 2)N_i^{\varepsilon}(t-) + I(T_i \geqslant t)$. It follows that $dN_i^c(t) = \omega_i(t)I(\varepsilon_i = 1)dN_i^{\varepsilon}(t)$ and $\omega_i(t) = I(\varepsilon_i = 2)N_i^{\varepsilon}(t-)\hat{G}(t)/\hat{G}(X_i) + I(X_i \geqslant t)$. Also define $d\widetilde{N}_i^c(t) = E\{dN_i^c(t)|\boldsymbol{W}_i, \varepsilon_i > 0\}$ and $\widetilde{\omega}_i(t) = E\{\omega_i(t)|\boldsymbol{W}_i, \varepsilon_i > 0\}$. Direct calculation gives

$$d\widetilde{N}_i^c(t) = \rho(\boldsymbol{W}_i)dN_i^\varepsilon(t), \quad \text{and}$$
$$\widetilde{\omega}_i(t) = \{1-\rho(\boldsymbol{W}_i)\}N_i^\varepsilon(t-)\hat{G}(t)/\hat{G}(X_i) + I(X_i \geqslant t).$$

For the complete cases, we are able to observe $N_i^c(t)$ (and $\omega_i(t)$ also), but for subjects with unknown failure type, we only observe $N_i^\varepsilon(t)$. We can however compute $d\widetilde{N}_i^c(t)$ and $\widetilde{\omega}_i(t)$ for everyone. Thus the proposed AIPW estimating equations can be constructed as

$$\begin{aligned}\boldsymbol{U}_{aw}(\boldsymbol{\beta}) = &\sum_{i=1}^n \frac{R_i}{\pi(\boldsymbol{W}_i)}\int_0^\tau \left\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},t)\right\} dN_i^c(t) \\ &-\sum_{i=1}^n \frac{R_i-\pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)}\int_0^\tau \left\{\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},t)\right\} d\widetilde{N}_i^c(t), \end{aligned} \tag{6}$$

where $\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},t) = S_{aw}^{(1)}(\boldsymbol{\beta},t)/S_{aw}^{(0)}(\boldsymbol{\beta},t)$ and

$$\begin{aligned}S_{aw}^{(k)}(\boldsymbol{\beta},t) = &\frac{1}{n}\sum_{j=1}^n \frac{R_j}{\pi(\boldsymbol{W}_j)}\omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k} \\ &-\frac{1}{n}\sum_{j=1}^n \frac{R_j-\pi(\boldsymbol{W}_j)}{\pi(\boldsymbol{W}_j)}\widetilde{\omega}_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}, \qquad k=0,1,2.\end{aligned}$$

Note that unlike $\boldsymbol{U}_w(\boldsymbol{\beta})$, $\boldsymbol{U}_{aw}(\boldsymbol{\beta})$ is augmented with a second term and in addition, uses augmented averages $S_{aw}^{(0)}(\boldsymbol{\beta},t)$ and $\boldsymbol{S}_{aw}^{(1)}(\boldsymbol{\beta},t)$, each taking contributions from both the complete and incomplete cases. For completeness of presentation, first we assume that π and ρ are known in Theorem 3 and provide the asymptotic properties of the resulting estimator, $\hat{\boldsymbol{\beta}}^{aw}$. In Theorem 4, we then examine the estimators obtained by using auxiliary models for π and ρ and show that they share the same asymptotic properties as $\hat{\boldsymbol{\beta}}^{aw}$ and possess the property of double-robustness.

THEOREM 3. *Under the regularity conditions (see Appendix), $\hat{\boldsymbol{\beta}}^{aw}$ is consistent for $\boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}^{aw}-\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1}\Xi_{aw}\boldsymbol{\mathcal{I}}_\beta^{-1})$, where $\Xi_{aw} = E\{\mathrm{M}^{\otimes 2}\} + E\{\frac{1-\pi(\boldsymbol{W})}{\pi(\boldsymbol{W})} Var(\mathrm{M}|\boldsymbol{W}, \varepsilon > 0)\}$.*

In practice, $\rho(\boldsymbol{W})$ is estimated from the observed data. This probability can be determined from the relationship of the cause-specific hazards of the latent failure times of the main event T_1 and competing event T_2, i.e.,

$$\frac{\lambda_1^{csh}(t|\boldsymbol{Z},\boldsymbol{A})}{\lambda_1^{csh}(t|\boldsymbol{Z},\boldsymbol{A})+\lambda_2^{csh}(t|\boldsymbol{Z},\boldsymbol{A})},$$

where $\lambda_j^{csh}(\cdot|\boldsymbol{Z},\boldsymbol{A})$ is the conditional cause-specific hazards of T_j, $j=1,2$. Rather than modeling two separate cause-specific hazards, we posit a parametric model $\rho(\boldsymbol{W};\boldsymbol{\eta})$ where $\boldsymbol{\eta}$ is of finite dimension. Again, we employ a standard logistic formulation in which $\rho(\boldsymbol{W}_i;\boldsymbol{\eta}) = 1/\{1+\exp(-\widetilde{\boldsymbol{W}}_i^T\boldsymbol{\eta})\}$. In the presence of missingness, obtaining an estimate of $\boldsymbol{\eta}$ is problematic. MAR however implies $\Pr(\varepsilon_i=1|\boldsymbol{W}_i,\varepsilon_i>0)=\Pr(\varepsilon_i=1|\boldsymbol{W}_i,\varepsilon_i>0,R_i=0)=\Pr(\varepsilon_i=1|\boldsymbol{W}_i,\varepsilon_i>0,R_i=1)$, signifying that we can estimate $\boldsymbol{\eta}$ from the uncensored complete cases. If $\rho(\boldsymbol{W}_i;\boldsymbol{\eta})$ is correctly specified, a consistent estimator for the true value $\boldsymbol{\eta}_0$ is the MLE $\hat{\boldsymbol{\eta}}$.

Theorem 4. *Under the regularity conditions (see Appendix), $\hat{\boldsymbol{\beta}}^{aw}(\boldsymbol{\gamma},\hat{\boldsymbol{\eta}})$, $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\boldsymbol{\eta})$ and $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ are consistent for $\boldsymbol{\beta}_0$ and have the same asymptotic distribution as $\hat{\boldsymbol{\beta}}^{aw}$ as long as either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{w};\boldsymbol{\eta})$ is specified correctly.*

Consistent estimation of the variances in Theorems 3 and 4 is very similar to that of the IPW estimators. Thus, we only demonstrate this for $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$. First we set an augmented Breslow-type estimator for $d\Lambda_{10}(t)$:

$$d\hat{\Lambda}_{10}(t)=\frac{1}{nS_{aw}^{(0)}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}),\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)}\sum_{j=1}^{n}\left\{\frac{R_j}{\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\eta}})}dN_j^c(t)-\frac{R_j-\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\eta}})}{\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\eta}})}d\widetilde{N}_j^c(t)\right\}.$$

Also define $d\hat{M}_i^c(t) = dN_i^c(t)-\omega_i(t)\exp\{[\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})]^T\boldsymbol{Z}_i\}d\hat{\Lambda}_{10}(t)$, $d\hat{\widetilde{M}}_i^c(t) = d\widetilde{N}_i^c(t;\hat{\boldsymbol{\eta}}) - \widetilde{\omega}_i(t;\hat{\boldsymbol{\eta}})\exp\{[\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})]^T\boldsymbol{Z}_i\}d\hat{\Lambda}_{10}(t)$,

$$\hat{\mathbf{M}}_{aw,i}=\int_0^\tau\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}),\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)\}d\hat{M}_i^c(t),\quad\text{and}$$
$$\hat{\widetilde{\mathbf{M}}}_{aw,i}=\int_0^\tau\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}),\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)\}d\hat{\widetilde{M}}_i^c(t),$$

with $d\widetilde{N}_i^c(t;\hat{\boldsymbol{\eta}}) = \rho(\boldsymbol{W}_i;\hat{\boldsymbol{\eta}})dN_i^\varepsilon(t)$ and $\widetilde{\omega}_i(t;\hat{\boldsymbol{\eta}}) = \{1-\rho(\boldsymbol{W}_i;\hat{\boldsymbol{\eta}})\}N_i^\varepsilon(t-)\hat{G}(t)/\hat{G}(X_i)+I(X_i \geqslant t)$. Then a consistent variance estimator is $\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}}^{-1}\hat{\Xi}_{aw}\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}}^{-1}$ where $\hat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\beta}} = -\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{U}_{aw}(\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}))$ and

$$\hat{\Xi}_{aw} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\eta}})}\hat{\mathrm{M}}_{aw,i}^{\otimes 2} + \frac{1}{n}\sum_{i=1}^{n}\frac{R_i\,(1-\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\eta}}))}{\pi^2(\boldsymbol{W}_i;\hat{\boldsymbol{\eta}})}(\hat{\mathrm{M}}_{aw,i} - \hat{\widetilde{\mathrm{M}}}_{aw,i})^{\otimes 2}.$$

5. Simulation Studies

In this section we conduct simulations to assess the properties of the IPW and AIPW estimators for small to moderate sample sizes. We also compare their performance to the full cohort (no missing cases), the compete-case analysis, extra-category analysis, and multiple imputation approach that uses the `R` package `mitools` (Lumley, 2010).

Sample sizes of $n = 200$ or 500 were chosen. Two failure types were considered, both of which depend on a single covariate Z. Z_i was generated from a balanced Bernoulli distribution, then the failure time for the event of interest $(T_i, \varepsilon_i = 1)$ was drawn from the subdistribution given by $F_1(t|Z_i) = 1-[1-p\{1-\exp(-t)\}]^{\exp(\beta_1 Z_i)}$. We set $p = 0.3$ to give about 30% type-1 failures when $Z = 0$ in the absence of censoring. Taking $\Pr(\varepsilon_i = 2|Z_i) = 1 - \Pr(\varepsilon_i = 2|Z_i)$, type-2 failure times were then generated from the subdistribution obtained from $\Pr(T_i \leqslant t|\varepsilon_i = 2, Z_i) = 1 - \exp\{-\exp(\beta_2 Z_i)t\}$. In all the simulations, we chose $\beta_1 = 0.5$ and $\beta_2 = -0.5$. The censoring times were independently simulated from Uniform(1,2) which produced 28% censoring on the average.

We assumed that the missingness mechanism follows a logistic model with complete-data probability that depends on the observed time X_i, the covariate Z_i, an auxiliary variable A_i drawn from a standard normal distribution, and the two-way interactions with X_i. That is, $\pi(\boldsymbol{W}_i;\boldsymbol{\gamma}) = \{1 + \exp(-\boldsymbol{\gamma}^T\widetilde{\boldsymbol{W}}_i)\}^{-1}$ where $\widetilde{\boldsymbol{W}}_i = (1, X_i, Z_i, A_i, X_iZ_i, X_iA_i)^T$. We set $\boldsymbol{\gamma} = (0, 1, -0.5, 0.25, -0.5, -0.25)^T$, producing about 33% missing among the uncensored cases.

To derive the IPW and AIPW estimators, we fitted two different models for π, one being the true model $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ and the other a severely misspecified constant model $\pi(\boldsymbol{W}_i;\gamma_0) =$

$\{1+\exp(-\gamma_0)\}^{-1}$. For comparison, we also included IPW analysis that uses the true complete-data probability. A model that includes all the two-way interaction terms was fitted for $\Pr(\varepsilon = 1|\boldsymbol{W}_i, \varepsilon_i > 0)$, i.e., $\rho(\widetilde{\boldsymbol{W}}_i;\boldsymbol{\eta}) = \{1+\exp(-\boldsymbol{\eta}^T\widetilde{\boldsymbol{W}}_i)\}^{-1}$ where $\widetilde{\boldsymbol{W}}_i = (1, X_i, Z_i, A_i, X_iZ_i, X_iA_i, Z_iA_i)$. This was also the model used to derive the MI estimates. Since the true ρ is a function of the two conditional CSH, note that $\rho(\widetilde{\boldsymbol{W}};\boldsymbol{\eta})$ is misspecified, a scenario that is more likely encountered in practice.

We calculated the bias, standard deviation of the estimates (SD), average of the standard error estimates (SE), and empirical coverage probability (CP) of the sample 95% confidence intervals from the 1000 simulated datasets. The results are shown in Table 2. As expected, the full cohort analysis produced virtually unbiased estimates and coverage probability close to the nominal level. The naïve methods CC and EC both exhibited substantial bias and severe under coverage. Despite an incorrect imputation model, the MI performed reasonably well in terms of bias. However, the estimated standard errors were too conservative. The simulation results confirm the equivalence of IPW using true probabilities (IPWt) and that using estimated probabilities (IPWc) of complete-data, both of which showed unbiasedness and correct coverage level. Although the IPW that uses estimated weights from an incorrect model (IPWi) showed some bias and slight undercoverage, it still performed better than either CC or EC. The AIPW estimator that uses the correct model for the π (AIPWc) performed equally well as its IPW counterpart for $n = 200$ and showed a reduced bias for $n = 500$. Despite having both the π and ρ misspecified, the performance of the AIPW method (AIPWi) is still impressive; a slight undercoverage though can be seen for $n = 500$. The IPW and AIPW estimators are, as generally perceived, potentially inferior in terms of efficiency compared to a correctly specified MI model, but this may be not the case when the MI model is misspecified. The results for $n = 500$ show that the SD of MI is higher than that of IPW and AIPW.

6. Example

We applied our two estimators to study the effects that several risk factors have on the development of CHD among HIV-infected individuals who have a history of either cocaine or alcohol dependence/abuse. We took data collected from January 2000 to July 2007 from a national cohort study and merged it with data from the 1999 Large Health Study of Veteran Enrollees. Development of CHD or the occurrence of death related to CHD constitute the event of interest, while deaths unrelated to CHD comprise the competing event. In this data, 37 were confirmed to have developed CHD or had a CHD-related death whereas 969 were confirmed have CHD-unrelated deaths. However, there were 28 deaths whose cause could not be classified. Although this number appears to be a small, its magnitude is not too far from the observed number of CHD or CHD-related deaths and thus cannot be ignored.

The main covariate of interest is the presence of coinfection with hepatitis-C virus (HCV). Other covariates considered are age at enrollment, race (black or others), obesity, presence of hypertension, presence of diabetes, current smoker or not, presence of hypercholesterolemia (TC), HIV-1 RNA level>500 at baseline (RNA), and CD4 count level ($<200, 200-499$ or $\geqslant 500$) at baseline. Of the $4,984$ participants, the median age at enrollment was 48 years. All were infected with HIV with 58% being coinfected with HIV and HCV.

We analyzed the data under model (1) and employed our two estimators to account for missing data. For comparison, we also fitted the model using an MI approach and the naïve approaches CC and EC. Both the fitted missingness model and the fitted imputation model include all the main effects of the predictors aforementioned, plus the log-transform of the observed failure time. The results are shown in Table 3.

The conclusions that can be drawn from the CC and EC analyses are identical. Looking at the primary covariate of interest, all the methods agree that presence of HCV increase the risk of CHD or CHD-related death in this population. However, the magnitude of the

IPW and AIPW estimates are higher relative to the MI estimate which agreed more with the naïve approaches. The methods also tend to agree in many of the other covariates but not all. Although not significant, it is striking that the estimated effect of race was increased by more than twice in IPW or AIPW relative to CC or EC. Diabetes appears to be significant and marginally with IPW and AIPW methods, respectively. Interestingly, only AIPW was able to declare smoking as a significant risk factor.

7. Discussion

In this article, we derived an inverse probability weighted estimator that is theoretically valid and computationally simple under a proportional subdistribution hazards model. We also demonstrated how it can be implemented in standard software. We should point out however that its consistency relies on a correctly specified model for the complete-data probability. Moreover, it is calculated from the complete cases resulting to a potential loss in efficiency. This led us to develop a second estimator that augments the IPW score equations with a term that incorporates information from the missing cases and that uses an imputation model for the type-1 failure probability. Recognizing that our proposed AIPW estimator could only be used with specialized software, we developed a user-friendly `R`-implemented `C++` macro that will be available for download.

Our AIPW estimator is doubly-robust being valid when either the complete-data probability model or the imputation model is correct. On one hand, Kang and Schafer (2007) showed through simulation that the usual doubly-robust estimator can be severely biased when both models are misspecified. For this reason, Tan (2006) and Cao et al. (2009) proposed improvements in doubly-robust estimation. On the other hand, our simulations demonstrated that the AIPW estimator could potentially be robust to misspecification of both models under the proportional subdistribution hazards framework, a finding whose confirmation requires additional studies.

It is important to realize that both the IPW and AIPW estimators operate under the assumption of MAR. The MI approach as suggested by Bakoyannis et al. (2010) also works under the same assumption. A natural question to ask is which approach should be taken in analyzing data with missing cause of failure. The advantages and disadvantages of these approaches have been the subject of some debate in the missing data literature (e.g., Carpenter et al. (2006); Vansteelandt et al. (2010)). If we can correctly specify the imputation model, then MI is generally favored for efficiency. However, correct specification is difficult in many settings, including the situation that we considered in this paper. As noted earlier, a correctly specified imputation model depends on the CSH of all event types. In practice, we would tend to rely on simpler models such as logistic regression. Even if we were willing to model all the conditional CSH, the chance of specifying all of them correctly decreases as the number of failure types increases. On the other hand, models for the complete-data probability are generally simpler; we may thus have more confidence and success in correctly specifying them. This is a practical appeal that IPW has. If we are torn choosing between MI or IPW, the AIPW estimator can offer protection from misspecifying the model for either π or ρ, but not necessarily both. It may lose some efficiency relative to a correctly specified MI model, but its robustness remains a strong attraction for use in real-life data analysis.

References

Andersen, J., Goetghebeur, E., and Ryan, L. (1996). Missing cause of death information in the analysis of survival data. *Statistics in Medicine* **15,** 2191–2201.

Bakoyannis, G., Siannis, F., and Touloumi, G. (2010). Modelling competing risks data with missing cause of failure. *Statistics in Medicine* **29,** 3172–3185.

Cao, W., Tsiatis, A., and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96,** 723–734.

Carpenter, J., Kenward, M., and Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A* **169,** 571–584.

Fine, J. and Gray, R. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94,** 496–509.

Foutz, R. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* **72,** 147–148.

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

Geskus, R. (2011). Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring. *Biometrics* **67,** 39–49.

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

Gray, B. (2010). *cmprsk: Subdistribution Analysis of Competing Risks.* R package version 2.2-1.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47,** 663–685.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* Wiley-Interscience.

Kang, J. and Schafer, J. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22,** 523–539.

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

Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statistica Sinica* **18,** 219–234.

Lumley, T. (2010). *mitools: Tools for multiple imputation of missing data.* R package version 2.0.1.

Qi, L., Wang, C., and Prentice, R. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association* **100,** 1250–1263.

Robins, J. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *Aids Epidemiology, Methodological issues*, Jewell, N., Dietz, K., and Farewell, V. (eds), 297-331. Boston: Birkhauser.

Robins, J., Rotnitzky, A., and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89,** 846.

Ruan, P. and Gray, R. (2008). Analyses of cumulative incidence functions via non-parametric multiple imputation. *Statistics in Medicine* **27,** 5709–5724.

Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association* **101,** 1619–1637.

Vansteelandt, S., Carpenter, J., and Kenward, M. (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* **6,** 37–48.

Appendix

The following regularity conditions are needed in the proofs: (1) $\Lambda_{10}(\tau) < \infty$; (2) $\Pr\{\omega(t) \geqslant \epsilon > 0, \forall t \in [0,\tau]\} > 0$; (3) $\boldsymbol{\mathcal{I}}_\beta$ is positive definite; (4) $\pi(\boldsymbol{W}) \geqslant \epsilon > 0$; and (5) $\boldsymbol{Z}$ is time independent and bounded.

The Lemma from Qi et al. (2005) is useful in many portions of the proofs. We refer to this as Lemma 1 henceforth.

Proof of Theorem 1. First we show that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta},t) \right\| \xrightarrow{a.s} 0, \tag{A.1}$$

where $\mathcal{B}$ is a compact neighborhood of $\boldsymbol{\beta}_0$. This result can be deduced using similar arguments as in step A1 of the proof of theorem 1 of Qi et al. (2005) by noting that the functions $\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)$ are bounded and $s^{(0)}(\boldsymbol{\beta},t)$ is bounded away from 0 on $[0,\tau]\times\mathcal{B}$, and that $\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)$ is an equicontinuous family at $\boldsymbol{\beta}$.

Next we show the asymptotic normality of $n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta})$. Using straightforward calculations, one can everywhere replace $N_i^c(t)$ by $M_i^c(t)$ in (4). Thus we can write $\boldsymbol{U}_w(\boldsymbol{\beta}) = A_1 - A_2$ where

$$A_1 = \sum_{i=1}^n \int_0^\infty \{\boldsymbol{Z}_i - \boldsymbol{e}(\boldsymbol{\beta},t)\} \frac{R_i}{\pi(\boldsymbol{W}_i)} dM_i^c(t), \quad \text{and}$$

$$A_2 = \int_0^\infty \left\{\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta},t) - \boldsymbol{e}(\boldsymbol{\beta},t)\right\} \sum_{i=1}^n \frac{R_i}{\pi(\boldsymbol{W}_i)} dM_i^c(t).$$

Following the proof of theorem 1 of Qi et al. (2005) but with their $\boldsymbol{E}_{SW}(\boldsymbol{\beta},\pi,t)$ replaced by $\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta},t)$, V by R, π by $\pi(\boldsymbol{W})$, and $d\bar{M}_n(t)$ by $d\bar{M}_n^c(t) = n^{-1/2}\sum_{i=1}^n R_i\pi^{-1}(\boldsymbol{W}_i)dM_i^c(t)$, it can be shown by applying Lemma 1 and the strong embedding theorem (Shorack and

Wellner, 1986, pp.47-48) that $n^{-1/2}A_2 \xrightarrow{\mathcal{P}} 0$. It follows that $n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta})$ can be approximated by the sum of mean-0 iid random variables $n^{-1/2}\sum_i R_i\pi^{-1}(\boldsymbol{W}_i)\mathbf{M}_i$, with variance equal to $\Xi_w = \mathrm{Var}\{R\pi^{-1}(\boldsymbol{W})\mathbf{M}\} = E\{\pi^{-1}(\boldsymbol{W})\mathbf{M}^{\otimes 2}\}$. The asymptotic normality follows from the central limit theorem.

We now establish the limit of $-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta})$. By applying (A.1) and Lemma 1, we see that $\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left\|-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}_\beta\right\| \xrightarrow{a.s} 0$.

Our next task is to establish the existence and consistency of $\hat{\boldsymbol{\beta}}^w$. Now, $n^{-1}\boldsymbol{U}_w(\boldsymbol{\beta}) \xrightarrow{\mathcal{P}} 0$ follows from the weak convergence of $n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta})$ and $\boldsymbol{\mathcal{I}}_\beta$ is positive definite by condition (3). Similar arguments as in the proof of theorem 2 of Foutz (1977) can be made to show that $\hat{\boldsymbol{\beta}}^w$ exists and is unique in $\mathcal{B}$ with probability converging to 1 as $n\to\infty$, and $\hat{\boldsymbol{\beta}}^w \xrightarrow{\mathcal{P}} \boldsymbol{\beta}_0$.

Finally, we show the asymptotic normality of $n^{1/2}\hat{\boldsymbol{\beta}}^w$. By routine Taylor series expansion,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^w - \boldsymbol{\beta}_0) = -\left[\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}*}\right]^{-1} n^{-1/2}\boldsymbol{U}_w(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^*$ is in the line segment formed by $\hat{\boldsymbol{\beta}}^w$ and $\boldsymbol{\beta}_0$. From the previous results, we have $n^{1/2}(\hat{\boldsymbol{\beta}}^w - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\mathcal{I}}_\beta^{-1}\Xi_w\boldsymbol{\mathcal{I}}_\beta^{-1})$.

Proof of Theorem 2. We begin by showing that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\|\boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\| \xrightarrow{\mathcal{P}} 0. \tag{A.2}$$

It is easily seen that (A.1) can be extended to

$$\sup_{\substack{t\in[0,\tau],\\(\boldsymbol{\beta},\boldsymbol{\gamma})\in\mathcal{B}\times\mathcal{C}}} \left\|\boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\| \xrightarrow{a.s} 0, \tag{A.3}$$

where $\mathcal{C}$ is a compact neighborhood of $\boldsymbol{\gamma}_0$. By a Taylor series expansion about $\boldsymbol{\gamma}$, we can write $\boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t)$ as

$$\boldsymbol{S}_w^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t) - (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\frac{1}{n}\sum_{j=1}^n \frac{R_j}{\pi^2(\boldsymbol{W}_j;\boldsymbol{\gamma})}\frac{\partial\pi(\boldsymbol{W}_j;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k} + o_p(1).$$

Then if $\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})$ is correctly specified, the second term converges to 0 in probability by the property of maximum likelihood estimators. This coupled with (A.3) shows (A.2).

To show the asymptotic normality of $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}})$, write $\boldsymbol{U}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}}) = B_1 - B_2$ where

$$B_1 = \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})} \int_0^\tau \{\boldsymbol{Z}_i - \boldsymbol{e}(\boldsymbol{\beta}.t)\}\, dM_i^c(t) \quad \text{and}$$
$$B_2 = \int_0^\tau \left\{\bar{\boldsymbol{Z}}_w(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t) - \boldsymbol{e}(\boldsymbol{\beta},t)\right\} \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})} dM_i^c(t).$$

If we can show that $n^{-1/2}\sum_{i=1}^n R_i\pi^{-1}(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})M_i^c(t)$ converges to 0 almost surely, then Lemma 1 and the strong embedding theorem implies that B_2 converges to 0. By a Taylor series expansion about $\boldsymbol{\gamma}$, we have that B_2 is approximately

$$n^{-1/2}\sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})} M_i^c(t) - (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) n^{-1/2} \sum_{i=1}^{n} \frac{R_i}{\pi^2(\boldsymbol{W}_i;\boldsymbol{\gamma})} \frac{\partial \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T} M_i^c(t)$$

which clearly converges to 0 by the boundedness of π away from 0. Next, a Taylor series expansion of B_1 about $\boldsymbol{\gamma}$ gives

$$\sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})} \mathbf{M}_i - (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) B_{\beta\gamma} + o_p(1),$$

where $B_{\beta\gamma} = \sum_{i=1}^n R_i\pi^{-2}(\boldsymbol{W}_i;\boldsymbol{\gamma})\{\partial\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^T\}\mathbf{M}_i$. By the martingale property, $n^{-1/2}B_{\beta\gamma}$ (and hence the second term in the equation above) converges to 0. Thus, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}})$ can be approximated by a sum of mean-0 iid random variables with variance $E\{\pi^{-1}(\boldsymbol{W};\boldsymbol{\gamma})\mathbf{M}^{\otimes 2}\}$. Normality immediately follows from the central limit theorem.

Now, (A.2) and Lemma 1 gives $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left\| -\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_w(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}}) - \boldsymbol{\mathcal{I}}_\beta \right\| \xrightarrow{\mathcal{P}} 0$. The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}^w(\hat{\boldsymbol{\gamma}})$ can then be established in the same fashion as in Theorem 1.

Proof of Theorem 3. We first show that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \left\| \boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},t) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta},t) \right\| \xrightarrow{a.s} 0. \tag{A.4}$$

We can write $\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},t) = n^{-1}\sum_j \omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k} + n^{-1}\sum_j [R_j - \pi(\boldsymbol{W}_j)]\pi^{-1}(\boldsymbol{W}_j)\{\omega_j(t) -$

$\widetilde{\omega}_j(t)\} \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}$. Since the first term converges to $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, t)$, it is left to show that second term converges to 0. This however immediately follows from MAR and recalling the definition of $\pi(\boldsymbol{W})$ and $\rho(\boldsymbol{W})$.

We now establish the asymptotic normality of $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta}, t)$. Note that we can replace $dN_i(t)$ and $d\widetilde{N}_i^c(t)$ in (6) by, respectively, $dM_i^c(t)$ and $d\widetilde{M}_i^c(t)$ where $d\widetilde{M}^c(t) = E\{dM^c(t)|\boldsymbol{W}, \varepsilon > 0\} = d\widetilde{N}^c(t) - \widetilde{\omega}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z})d\Lambda_{10}(t)$. By extending the proof of Theorem 1 and applying (A.4), the strong embedding theorem, and Lemma 1, we can write

$$n^{-1/2}\boldsymbol{U}_{aw}(\beta) = n^{-1/2}\sum_{i=1}^{n}\mathbf{M}_i + n^{-1/2}\sum_{i=1}^{n}\frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)}(\mathbf{M}_i - \widetilde{\mathbf{M}}_i) + o_p(1), \tag{A.5}$$

with $\widetilde{\mathbf{M}} = \int_0^\tau \{\boldsymbol{Z} - \boldsymbol{e}(\boldsymbol{\beta}, t)\}d\widetilde{M}^c(t)$. Thus, $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta})$ is approximately a sum of mean-0 iid random variables with variance

$$\begin{aligned}\boldsymbol{\Xi}_{aw} &= E\left\{\mathbf{M} + \frac{R - \pi(\boldsymbol{W})}{\pi(\boldsymbol{W})}(\mathbf{M} - \widetilde{\mathbf{M}})\right\}^{\otimes 2} \\ &= E\left\{\mathbf{M}^{\otimes 2}\right\} + E\left\{\frac{1 - \pi(\boldsymbol{W})}{\pi(\boldsymbol{W})}\mathrm{Var}(\mathbf{M}|\boldsymbol{W}, \varepsilon > 0)\right\},\end{aligned}$$

Applying the central limit theorem proves the asymptotic normality of $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta})$.

We now have to show that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left\| -\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_{aw}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}_\beta \right\| \xrightarrow{\mathcal{P}} 0. \tag{A.6}$$

Taking the derivative of (6) with respect to $\boldsymbol{\beta}$, we have

$$\begin{aligned}-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\beta}^T}\boldsymbol{U}_{aw}(\boldsymbol{\beta}) = &\int_0^\tau \frac{\partial}{\partial\boldsymbol{\beta}^T}\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t)\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{R_i}{\pi(\boldsymbol{W}_i)}dM_i^c(t) - \frac{R_i - \pi(\boldsymbol{W}_i)}{\pi(\boldsymbol{W}_i)}d\widetilde{M}_i^c(t)\right\} \\ &+ \int_0^\tau \frac{\partial}{\partial\boldsymbol{\beta}^T}\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t)S_{aw}^{(0)}(\boldsymbol{\beta}, t)d\Lambda_{10}(t),\end{aligned}$$

where

$$\frac{\partial}{\partial\boldsymbol{\beta}^T}\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{S}_{aw}^{(2)}(\boldsymbol{\beta}, t)}{S_{aw}^{(0)}(\boldsymbol{\beta}, t)} - \left(\frac{\boldsymbol{S}_{aw}^{(1)}(\boldsymbol{\beta}, t)}{S_{aw}^{(0)}(\boldsymbol{\beta}, t)}\right)^{\otimes 2}.$$

Because $n^{-1}\sum_i R_i\pi^{-1}(\boldsymbol{W}_i)dM_i^c(t)$ and $n^{-1}\sum_i [R_i-\pi(\boldsymbol{W}_i)]\pi^{-1}(\boldsymbol{W}_i)d\widetilde{M}_i^c(t)$ converges to 0, applying (A.4) and Lemma 1 proves (A.6). It is then straightforward to show the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}^{aw}$ following the same steps used in Theorem 1.

Proof of Theorem 4. Here we just derive the asymptotic results for $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ (similar steps can be followed for $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\boldsymbol{\eta})$ and $\boldsymbol{\beta}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$). The estimating equation is given by

$$\begin{aligned}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}) &= \sum_{i=1}^n \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \left\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)\right\}dN_i^c(t)\\ &\quad -\sum_{i=1}^n \frac{R_i-\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\int_0^\tau \left\{\boldsymbol{Z}_i-\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)\right\}d\tilde{N}_i^c(t;\hat{\boldsymbol{\eta}}),\end{aligned}$$

where $\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},t)=S_{aw}^{(1)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)/S_{aw}^{(0)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)$, and

$$\begin{aligned}S_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t) &= \frac{1}{n}\sum_{j=1}^n \frac{R_j}{\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})}\omega_j(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}\\ &\quad -\frac{1}{n}\sum_{j=1}^n \frac{R_j-\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_j;\hat{\boldsymbol{\gamma}})}\tilde{\omega}_j(t;\hat{\boldsymbol{\eta}})\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_j)\boldsymbol{Z}_j^{\otimes k}, \qquad k=0,1,2.\end{aligned}$$

We begin by showing that

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}}\left\|\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\|\xrightarrow{\mathcal{P}}0. \tag{A.7}$$

We note that (A.4) can be extended to

$$\sup_{\substack{t\in[0,\tau],\\(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})\in\mathcal{B}\times\mathcal{C}\times\mathcal{H}}}\left\|\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},t)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta},t)\right\|\xrightarrow{a.s}0, \tag{A.8}$$

where $\mathcal{H}$ is a compact neighborhood of $\boldsymbol{\eta}_0$. An expansion on $\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)$ about $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$, results to $\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t)=\boldsymbol{S}_{aw}^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},t)-D_{11}-D_{12}+o_p(1)$ where

$$\begin{aligned}D_{11} &= (\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi^2(\boldsymbol{W}_i;\boldsymbol{\gamma})}\frac{\partial\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\{\omega_i(t)-\tilde{\omega}_i(t;\boldsymbol{\eta})\} \quad \text{and}\\ D_{12} &= (\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})\frac{1}{n}\sum_{i=1}^n \frac{R_i-\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i)\frac{\partial}{\partial\boldsymbol{\eta}^T}\tilde{\omega}_i(t;\boldsymbol{\eta}).\end{aligned}$$

Clearly, D_{11} and D_{12} converge to 0 when either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{W};\boldsymbol{\eta})$ is correct. Applying (A.8) gives the desired result.

To establish the normality of $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$, write $\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}}) = D_{21} - D_{22}$ where

$$D_{21} = \sum_{i=1}^{n} \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\mathbf{M}_i - \sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}\widetilde{\mathbf{M}}(\hat{\boldsymbol{\eta}}), \quad \text{and}$$

$$D_{22} = \int_0^{\tau} \left\{\bar{\boldsymbol{Z}}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}},t) - \boldsymbol{e}(\boldsymbol{\beta},t)\right\} \sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})} dM_i^c(t) - \frac{R_i - \pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})}{\pi(\boldsymbol{W}_i;\hat{\boldsymbol{\gamma}})} d\widetilde{M}_i^c(t;\hat{\boldsymbol{\eta}}) \right\},$$

with $\widetilde{\mathbf{M}}(\boldsymbol{\eta}) = \int_0^{\tau} \{\boldsymbol{Z}_i - \boldsymbol{e}(\boldsymbol{\beta},t)\}\, d\widetilde{M}_i^c(t;\boldsymbol{\eta})$ and $d\widetilde{M}_i^c(t;\boldsymbol{\eta}) = d\widetilde{N}_i(t;\boldsymbol{\eta}) - \widetilde{\omega}_i(t;\boldsymbol{\eta}) \exp(\boldsymbol{\beta}^T\boldsymbol{Z}_i) d\Lambda_{10}(t)$. Again by expansion about $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$, it can be shown that the summation in D_{22} converges to 0 when either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{W};\boldsymbol{\eta})$ is correct. Applying (A.7), the strong embedding theorem, and Lemma 1 leads to $n^{-1/2}D_{22}$ converging to 0 in probability. Following a similar expansion, we can further decompose $n^{-1/2}D_{21}$ into $n^{-1/2}D_{31} - n^{-1/2}D_{32} - n^{-1/2}D_{33} + o_p(1)$ where

$$\begin{aligned}
n^{-1/2}D_{31} &= n^{-1/2}\sum_{i=1}^{n} \left\{ \frac{R_i}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\mathbf{M}_i - \frac{R_i - \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}\widetilde{\mathbf{M}}_i(\boldsymbol{\eta}) \right\}, \\
n^{-1/2}D_{32} &= n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\pi^2(\boldsymbol{W}_i;\boldsymbol{\gamma})} \frac{\partial \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T} \left\{\mathbf{M}_i - \widetilde{\mathbf{M}}_i(\boldsymbol{\eta})\right\} \quad \text{and} \\
n^{-1/2}D_{33} &= n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})\frac{1}{n}\sum_{i=1}^{n} \frac{R_i - \pi(\boldsymbol{W}_i;\boldsymbol{\gamma})}{\pi(\boldsymbol{W}_i;\boldsymbol{\gamma})} \frac{\partial}{\partial \boldsymbol{\eta}^T}\widetilde{\mathbf{M}}_i(\boldsymbol{\eta}).
\end{aligned}$$

If either $\pi(\boldsymbol{W};\boldsymbol{\gamma})$ or $\rho(\boldsymbol{W};\boldsymbol{\eta})$ is correct, it can be immediately deduced that $n^{-1/2}D_{32}$ and $n^{-1/2}D_{33}$ converges to 0 in probability. Thus, $n^{-1/2}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ is approximately a sum of mean-0 iid random variables that has the same form as (A.5). Asymptotic normality follows from the central limit theorem.

Next by (A.7) and Lemma 1, convergence of $-\frac{1}{n}\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{U}_{aw}(\boldsymbol{\beta},\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ to $\boldsymbol{\mathcal{I}}_{\beta}$ can be shown. The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}^{aw}(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\eta}})$ can then be established.

Table 1: Preparing data for inverse probability weighted analysis

(a) Sample dataset

subect	X	etype	w.ipw	Ghat	Z
1	2	0	1.00	0.9	3
2	4	1	1.25	0.8	6
3	6	2	1.50	0.7	9
4	8	1	1.75	0.6	12
5	10	1	2.00	0.5	15

(b) Sample dataset in counting process style of input

subject	X	Tstart	Tstop	etype	w.ipw	w.cens	Z
1	2	0	2	0	1.00	1.000	3
2	4	0	4	1	1.25	1.000	6
3	6	0	4	2	1.50	1.000	9
3	6	4	8	2	1.50	0.750	9
3	6	8	10	2	1.50	0.625	9
4	8	0	8	1	1.75	1.000	12
5	10	0	10	1	2.00	1.000	15

`X` is the followup time, `etype` is the event type, `w.ipw` is the IPW weight, `w.cens` is the IPCW weight, and `Ghat` is the Kaplan-Meier estimate of the censoring survival distribution.

Table 2: Simulation results comparing the bias, average of the standard error estimates (SE), empirical standard deviation (SD), and empirical coverage probability (95%) of the different analysis methods for cohort sizes of $n = 200$ and $n = 500$.

	$n = 200$				$n = 500$			
	Bias	SE	SD	CP	Bias	SE	SD	CP
FC	0.029	0.273	0.273	0.952	0.019	0.171	0.172	0.945
CC	-0.233	0.375	0.374	0.906	-0.229	0.234	0.235	0.831
EC	-0.410	0.375	0.369	0.803	-0.405	0.233	0.234	0.591
MI	-0.035	0.430	0.358	0.982	-0.010	0.280	0.235	0.976
IPWt	0.018	0.369	0.369	0.961	0.020	0.231	0.231	0.952
IPWc	0.023	0.373	0.367	0.962	0.020	0.232	0.230	0.957
IPWi	-0.101	0.363	0.363	0.949	-0.099	0.227	0.230	0.922
AIPWc	0.013	0.383	0.366	0.970	0.007	0.229	0.229	0.952
AIPWi	0.012	0.352	0.360	0.955	0.008	0.220	0.228	0.943

FC for full cohort analysis; CC for complete cases; EC for missing failure type as an extra category; MI for multiple imputation; IPWt, IPWc, and IPWi for inverse weighting method using the true, correct model for, and incorrect model for π; AIPWc and AIPWi for augmented inverse weighting method using the correct and incorrect model for π.

Table 3: Analysis of Coronary Heart Disease Data

Covariate		CC	EC	MI	IPW	AIPW
HCV	Estimate	1.198	1.188	1.220	1.367	1.364
	Std. Err.	0.441	0.438	0.450	0.455	0.413
	p-value	0.007	0.007	0.007	0.003	0.001
AGE	Estimate	0.069	0.067	0.068	0.079	0.079
	Std. Err.	0.022	0.022	0.025	0.025	0.035
	p-value	0.002	0.002	0.007	0.002	0.024
RACE	Estimate	-0.147	-0.117	-0.176	-0.313	-0.319
	Std. Err.	0.357	0.353	0.358	0.377	0.423
	p-value	0.681	0.740	0.622	0.406	0.452
DIABETES	Estimate	0.448	0.406	0.519	0.776	0.797
	Std. Err.	0.369	0.365	0.422	0.368	0.422
	p-value	0.225	0.265	0.219	0.035	0.059
SMOKER	Estimate	0.706	0.683	0.692	0.923	0.925
	Std. Err.	0.547	0.543	0.523	0.564	0.404
	p-value	0.197	0.208	0.186	0.102	0.022
OBESE	Estimate	-1.062	-1.083	-1.048	-1.121	-1.084
	Std. Err.	0.748	0.749	0.720	0.758	0.814
	p-value	0.156	0.148	0.145	0.139	0.183
HYPERTENSIVE	Estimate	0.795	0.774	0.833	0.914	0.910
	Std. Err.	0.333	0.333	0.365	0.331	0.316
	p-value	0.017	0.020	0.023	0.006	0.004
TC	Estimate	0.849	0.836	0.869	1.002	1.022
	Std. Err.	0.354	0.354	0.348	0.357	0.401
	p-value	0.017	0.018	0.012	0.005	0.011
RNA	Estimate	0.064	0.076	0.057	0.071	0.060
	Std. Err.	0.354	0.356	0.361	0.372	0.439
	p-value	0.857	0.832	0.875	0.848	0.892
CD4 ($\geqslant 500$)	Estimate	-0.973	-0.961	-0.972	-0.821	-0.813
	Std. Err.	0.476	0.475	0.455	0.482	0.562
	p-value	0.041	0.043	0.033	0.088	0.148
CD4 ($200 - 499$)	Estimate	-0.464	-0.438	-0.524	-0.545	-0.545
	Std. Err.	0.372	0.369	0.389	0.370	0.398
	p-value	0.213	0.236	0.177	0.141	0.171

CC for complete cases; EC for missing failure type as an extra category; MI for multiple imputation; IPW inverse weighting; and AIPW for augmented inverse weighting.