Smoothed Rank Regression for the Accelerated Failure Time Competing Risks Model with Missing Cause of Failure

Zhiping Qiu[1,2], **Alan T. K. Wan**[3], **Yong Zhou**[4,5], and **Peter B. Gilbert**[6]

[1]School of Mathematical Sciences, Huaqiao University, Quanzhou 362021, China

[2]Research Center for Applied Statistics and Big Data, Huaqiao University, Xiamen 361021, China

[3]City University of Hong Kong, Kowloon, Hong Kong

[4]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

[5]Institute of Applied Mathematics, Chinese Academy of Science, Beijing 100190, China

[6]Department of Biostatistics, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

Abstract

This paper examines the accelerated failure time competing risks model with missing cause of failure using the monotone class rank-based estimating equations approach. We handle the non-smoothness of the rank-based estimating equations using a kernel smoothed estimation method, and estimate the unknown selection probability and the conditional expectation by non-parametric techniques. Under this setup, we propose three methods for estimating the unknown regression parameters based on 1) inverse probability weighting, 2) estimating equations imputation and 3) augmented inverse probability weighting. We also obtain the associated asymptotic theories of the proposed estimators and investigate the estimators' small sample behaviour in a simulation study. A direct plug-in method is suggested for estimating the asymptotic variances of the proposed estimators. A real data application based on a HIV vaccine efficacy trial study is considered.

Keywords

Accelerated failure time model; competing risks; imputation; inverse probability weighting; missing at random; monotone estimating equation; rank-based estimator; U-statistic

1 Introduction

In many areas of research, investigators are interested in studying the effects of different factors on the hazards or failures from a specific cause but failures may often result from multiple causes. This leads to the problem of competing risks. This problem arises most frequently in clinical trials where patients may fail from causes other than the disease under investigation. Studies on competing risks that focus on the covariate effects on the cause-specific hazard function for the failure type of interest include Cheng, Fine and Wei (1998), Shen and Cheng (1999) and Scheike and Zhang (2003). Some authors have also considered

direct modeling of the sub-distribution of a competing risk (Fine and Gray, 1999; Sun et al., 2006).

The majority of studies on competing risks to-date assume that the cause of failure is always known and observed. Oftentimes, the cause of failure may be unknown. For example, in the "Mashi trial" study concerning HIV-related death of infants born to HIV-infected mothers in Botswana considered by Sun, Wang and Gilbert (2012), the causes of death of live-born infants are known only for 50 and missing for 61 of the 111 observations in the study sample. In another study concerning survival times of HIV patients, Bakoyannis, Siannis and Touloumi (2010) also reported missing causes of death for some sample observations. In our example in Section 6 based on the HVTN 502 'Step' Phase IIb HIV vaccine efficacy trial study, HIV sequences are missing for 23 out of 88 infected participants (Buchbinder et al., 2008). Methods that account for the missing failure causes to-date all assume that the causes of failure are missing at random (MAR), meaning that the probability of missingness is only related to the fully observed variables and not to the partially unobserved cause of failure. Studies that focus on the covariate effects on the cause-specific hazard function assuming multiplicative effects when failure causes may be missing include Goetghebeur and Ryan (1995), Lu and Tsiatis (2001) and Gao and Tsiastis (2005). The former two of these studies use data imputation methods to compute fitted values for the missing failure causes, while the latter study addresses the missing data issue by an augmented inverse probability weighting approach within the framework of a linear transformation model. Lu and Liang (2008) considered an additive hazards model and developed inverse probability weighting and doubly robust methods for estimating the regression coefficients. Other studies on competing risks with missing failure causes that focus on aspects other than the cause-specific hazard function include Bakoyannis, Siannis and Touloumi (2010), who concentrated on the modelling of the cumulative incidence function, and Sun, Wang and Gilbert (2012), who considered quantile regression modeling of the survival time.

Recently, Zheng, Lin and Yu (2016) analysed the competing risks data with missing causes of failure under the accelerated failure time (AFT) model. The AFT model permits a direct measurement of the effects of the covariates on the survival time instead of the hazard function. This facilitates interpretation of results and is considered to be a major advantage of the AFT model over hazards models. One common approach for fitting AFT models is rank-based estimation developed from the weighted log-rank test (Prentice, 1978). This is also the approach taken by Zheng, Lin and Yu (2016) in their study. When the data are right-censored, the rank-based approach leads to estimators that are consistent and asymptotically normal (Tsiatis, 1990; Ying, 1993). In a recent paper, Lee and Lewbel (2013) provided general identification conditions and developed a sieve maximum likelihood estimation procedure for the AFT model with competing risks data. One shortcoming of the rank-based approach is that rank estimating functions are discontinuous. This feature poses formidable challenges to the computation of the regression coefficient estimates and subsequent inference. Jin et al. (2003) proposed a method that goes some way towards resolving this difficulty. They suggested a monotone approximation to the rank estimating function and a relatively straightforward linear programming-based procedure for estimating the regression coefficients. However, as the inference procedure of Jin et al.'s (2003) method involves re-

sampling, their method can be demanding on computation time, especially with large datasets and for models with many covariates.

Another approach is to apply smoothing methods to the non-smooth estimating functions. The objective of this approach is to construct a smooth surrogate estimating function that is asymptotically equivalent to the original non-smooth function. The continuous differentiability of the surrogate equation ensures that solutions can be obtained by standard numerical algorithms. Particularly notable in this respect are the contributions of Brown and Wang (2005) and Heller (2007). Brown and Wang (2005) proposed an induced smoothing method whereby the smoothed estimating functions are obtained by taking expectations with respect to an artificial Gaussian continuous noise variable added to the regression coefficients. Heller (2007) employed a direct approximation of the non-smooth function by a local distribution function. As noted by Johnson and S-trawderman (2009), when applying Heller's method, if one uses the standard Gaussian cumulative distribution function as the local distribution function, this method will yield the same smoothed estimating functions as Brown and Wang's method that replaces the covariate-dependent bandwidth by a fixed bandwidth. Brown and Wang's (2005) induced smoothing method has been generalised to estimating functions with general weight (Chiou, Kang and Yan, 2014), and extended to AFT models with censored data (Brown and Wang, 2007; Zhao, Brown and Wang, 2014), clusterd data (Johnson and Strawderman, 2009), censored and clustered data (Wang and Fu, 2011), and quantile regression (Pang, Lu and Wang, 2012). However, to the best of our knowledge, neither Brown and Wang's nor Heller's methods have been applied to AFT models with missing failure causes, and the purpose of this paper is to take steps in this direction.

In this paper, we consider the AFT competing risk model with MAR causes of failure using the monotone rank estimating equations approach. We overcome the difficulty with regard to the discontinuity of the rank estimating equations using a local distribution function smoothing approach in the spirits of Heller (2007). Under this setup, we consider three procedures for estimating the unknown regression coefficients. The first is based on a non-parametric inverted probability weighting (IPW) approach, similar to that developed by Qi, Wang and Prentice (2005) for the proportional hazards model. This approach uses non-parametric smoothers in estimating the selection probabilities, thus overcoming the difficulty with the mis-specification of propensity score frequently encountered with parametric methods. The second method is based on the estimating equation imputation (EEI) approach proposed under a general setup by Zhou, Wan and Wang (2008). The EEI approach is closely related to the missing information principle; in the context of interest here, studies that apply the missing information principle for the handling of censored data trace back to the work of Buckley and James (1979). The third is an augmented IPW (AIPW) approach in the spirits of Robins, Rotnitzky and Zhao (1994) who considered a general setup. An important appeal of this approach is that it leads to estimators that are doubly robust. The AIPW approach was considered by Wang and Chen (2001) for the proportional hazards model.

It should be emphasized at the outset that although Zheng, Lin and Yu (2016) also considered the modeling of competing risk data with missing failure causes by a rank-based

estimating equations procedure, there are significant differences between ours and Zheng, Lin and Yu's (2016) approaches. First, in order to overcome the difficulty with respect to solving the discontinuous rank estimating equations, Zheng, Lin and Yu (2016) transformed the problem into an optimisation problem, and the subsequent inference procedure involves re-sampling, which can be demanding on computation time. However, for the proposed method, as the estimating equations are differentiable with respect to the unknown parameters, estimates of the parameters can be computed by the Newton-Raphson algorithm, and the associated asymptotic variances can be estimated by a plug-in method. This affords a substantial computational advantage over the method of Zheng, Lin and Yu (2015). Second, we consider the IPW, EEI and AIPW methods for handling missing data but Zheng, Lin and Yu (2016) only discussed the IPW and doubly robust methods based on a Martingale with zero mean. In particular, the EEI method we introduce does not require the estimation of the missing probability. We consider the latter a significant advantage. As will be shown ahead, all three missing data handling methods being considered have identical asymptotic theoretical properties and very comparable finite sample properties.

The remainder of the paper is organised as follows. Section 2 describes the model setup and the smoothed rank estimating equations approach. The three proposed methods for handling missing failure causes and their theoretical properties are discussed and examined in Section 3. Section 4 explores the selection of kernel functions and bandwidth parameters, along with a discussion on dimension reduction. Section 5 focuses on the finite sample properties of estimators, while Section 6 considers applications of the proposed methods based on a real data set. Some concluding remarks are placed in Section 7. Proofs of theorems are contained in an appendix.

2 Notations, Model Descriptions and A Smoothed Rank Estimating Equations Approach

Let the population contain n independent subjects, and for simplicity and without loss of generality, we assume that there are only two mutually exclusive causes of failure, denoted by $J_i = 1, 2$. For the ith ($i = 1, 2, \ldots, n$) subject, let T_{i1} and T_{i2} be the latent failure times associated with $J_i = 1$ and $J_i = 2$ respectively, $T_i = \min(T_{i1}, T_{i2})$ be the uncensored failure time, C_i the right-censoring time, $\delta_i = I(T_i \quad C_i)$ the censoring indicator such that $\delta_i = 1$ if T_i is observed and $\delta_i = 0$ otherwise, and Z_i be the $p \times 1$ vector of covariates. The observable failure time is thus $\tilde{T}_i = \min(T_i, C_i)$. Furthermore, we assume that C_i, T_{i1} and T_{i2} are mutually independent given Z_i.

Suppose that we are only interested in assessing the covariate effect on the failure time of second type. The AFT model postulates the following linear relationship between the natural log of the failure time and the covariates (Kalbfleisch and Prentice, 2002):

$$\log(T_2) = Z^T \beta + \varepsilon, \quad (1)$$

where $\boldsymbol{\beta}$ is an unknown $p \times 1$ regression coefficient vector, and the error term ε has a mean of zero with an unspecified continuous distribution independent of $\mathbf{Z}$. When there is no missing cause of failure, the right-censored competing risks data set comprises i.i.d. observations of $(\tilde{T}_i, \delta_i, \delta_i J_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$. Let the counting process be $N_i(t) = I\{ \log \tilde{T}_i - \mathbf{Z}_i^T \boldsymbol{\beta} \le t, \delta_i J_i = 2\}$, $Y_i(t) = I\{ \log \tilde{T}_i - \mathbf{Z}_i^T \boldsymbol{\beta} \ge t\}$, and $\lambda(t)$ be the unknown hazard function of ε in (1). It can be shown using counting process theory (Fleming and Harrington, 1991) that

$$M_i(t) = N_i(t) - \int_{-\infty}^t Y_i(u)\lambda(u)du, i = 1, 2, \ldots, n,$$

are mean zero martingale processes. By applying similar arguments as Tsiatis (1990), we obtain the following estimating equations for the joint estimation of $\boldsymbol{\beta}$ and $\lambda(t)$:

$$\sum_{i=1}^n dM_i(t) = \sum_{i=1}^n [dN_i(t) - Y_i(t)\lambda(t)dt] = 0, \quad (2)$$

$$\sum_{i=1}^n \int_{-\infty}^{\tau} \mathbf{Z}_i dM_i(t) = \sum_{i=1}^n \int_{-\infty}^{\tau} \mathbf{Z}_i[dN_i(t) - Y_i(t)\lambda(t)dt] = 0, \quad (3)$$

where τ is a constant representing the end time of the study. Equation (3) yields

$$\lambda(t)dt = \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t)}. \quad (4)$$

Substituting (4) into (3) leads to the following estimating equation for $\boldsymbol{\beta}$ in model (1):

$$\sum_{i=1}^n \delta_i I(J_i = 2)\left[\mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j I(\log \tilde{T}_j - \mathbf{Z}_j^T \boldsymbol{\beta} \ge \log \tilde{T}_i - \mathbf{Z}_i^T \boldsymbol{\beta})}{\sum_{j=1}^n I(\log \tilde{T}_j - \mathbf{Z}_j^T \boldsymbol{\beta} \ge \log \tilde{T}_i - \mathbf{Z}_i^T \boldsymbol{\beta})}\right] = 0. \quad (5)$$

Note that the l.h.s. of (5) is not monotone in $\boldsymbol{\beta}$, and this may produce multiple solutions of $\boldsymbol{\beta}$. To reconcile this difficulty, we consider the following monotone rank estimating equation analogous to that proposed by Fygenson and Ritov (1994) for censored data:

$$\tilde{U}_n(\boldsymbol{\beta}) \equiv n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j) I\{ \log(\tilde{T}_j) - \mathbf{Z}_j^T \boldsymbol{\beta} \geq \log(\tilde{T}_i) - \mathbf{Z}_i^T \boldsymbol{\beta} \} = 0. \quad (6)$$

Although the l.h.s. of (6) is monotone in $\boldsymbol{\beta}$, it is still discontinuous with respect to $\boldsymbol{\beta}$ due to the presence of an indicator (jump) function in it. A range of well-developed algorithms including the brutal search method, Nelder-Mead method and linear programming method developed by Jin et al. (2003) can be used for computing $\hat{\boldsymbol{\beta}}$. However, as the asymptotic covariance matrix of the estimators involves the hazards function of an unspecified error distribution, direct estimation of the covariance matrix requires an estimate of this hazards function. Recognising that this estimate can be highly unstable, Jin et al. (2003) proposed a resampling method to estimate the covariance matrix that eliminates the estimation of the hazards function but the computation efforts involved for the resampling method can be immense, especially with large data-sets. This motivates us to develop a differentiable estimating equation to approximate (6). Specifically, define $r_i^{\beta} = \log(\tilde{T}_i) - \mathbf{Z}_i^T \boldsymbol{\beta}$, $i = 1, 2, \ldots,$ n, along the lines of Heller (2007), we consider an approximation to the indicator function $I(r_j^{\beta} \geq r_i^{\beta})$ by a local distribution function $S((r_j^{\beta} - r_i^{\beta})/\sigma_n)$, where $S(u)$ is non-decreasing and satisfies the conditions $\lim_{u \to \infty} S(u) = 1$ and $\lim_{u \to -\infty} S(u) = 0$, where σ_n is a sequence of strictly positive and decreasing numbers satisfying $\lim_{n \to \infty} \sigma_n = 0$. Clearly, when $r_j^{\beta} > r_i^{\beta}$, $S((r_j^{\beta} - r_i^{\beta})/\sigma_n) \to 1$ as $n \to \infty$; on the other hand, when $r_j^{\beta} < r_i^{\beta}$, $S((r_j^{\beta} - r_i^{\beta})/\sigma_n) \to 0$ as $n \to \infty$. A smoothed version of (6) is thus given by

$$n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j) S\left(\frac{r_j^{\beta} - r_i^{\beta}}{\sigma_n} \right) = 0. \quad (7)$$

3 Methods for Handling Missing Causes of Failure

When the causes of failure are only partially available, the estimating equation (7) cannot be applied because J_i's are not observed for all i's. Now, let R_i be the complete-case indicator that is equal to 1 when either $\delta_i = 0$, or $\delta_i = 1$ and J_i is observed, and equal to 0 otherwise. Thus, when the causes of failure are not completely observed, the right-censored competing risks data set comprises i.i.d. observations of $\{(\tilde{T}_i, \delta_i, \mathbf{Z}_i, A_i, R_i, R_i\delta_iJ_i), i = 1, \ldots, n\}$, where A_i's are some auxiliary covariates that may be useful for predicting the missing failure type.

We assume that the cause of failure is MAR (Rubin, 1976). That is, given $\delta_i = 1$ and $W_i = (\tilde{T}_i, \mathbf{Z}_i^T, A_i)^T$, the probability that the failure cause of the ith subject is missing depends only on the observed W_i, but not on the unobserved J_i. Specifically, we assume that the failure cause missing probability is given by

$$r(\mathbf{W}_i) = P(R_i = 1 | J_i, \delta_i = 1, \mathbf{W}_i) = P(R_i = 1 | \delta_i = 1, \mathbf{W}_i). \quad (8)$$

Although the MAR assumption is more restrictive than nonignorable missingness, MAR is justified in many practical situations, and there is a large collection of literature that uses the MAR assumption as the baseline for analysis. Recent examples include Aerts et al. (2002), Wang and Rao (2002), Chen, Ibrahim and Shao (2004), Qi, Wang and Prentice (2005), Lu and Copas (2005), Zhou, Wan and Wang (2008), among others. In the remainder of this section, we develop three methods for dealing with missing data in the context of competing risks data.

3.1 Inverse probability weighting

Write $\mathbf{Q}_i = (\mathbf{W}_i^T, \delta_i)^T$. From Horvitz and Thompson (1952), note that

$$M_i^{(1)}(t) \equiv \frac{R_i}{\pi(\mathbf{Q}_i)} N_i(t) - \int_{-\infty}^{t} Y_i(u)\lambda(u)du, i = 1, 2, \ldots, n,$$

are mean zero processes, where $\pi(\mathbf{Q}_i) = P(R_i = 1 | \delta_i, \mathbf{W}_i) = \delta_i r(\mathbf{W}_i) + (1 - \delta_i)$. By derivations similar to those in Section 2, this leads to the following inverse probability weighted (IPW) estimating equation for $\boldsymbol{\beta}$:

$$n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\pi(\mathbf{Q}_i)} \delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j) S\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n}\right) = 0. \quad (9)$$

In practice, $r(\mathbf{W}_i)$ is often unknown. We may estimate $r(\mathbf{W}_i)$ parametrically as in Gao and Tsiatis (2005), Lu and Liang (2008) and Sun, Wang and Gilbert (2012), or non-parametrically as in Qi, Wang and Prentice (2005), Zhou, Wan and Wang (2008), and Song et al. (2010). Here, we adopt the non-parametric approach which has the advantage over its parametric counterpart of being less prone to biases arising from model mis-specification. We use a kernel method and assume that d is the size of the continuous elements in $\mathbf{W}_i$ and $k(u)$ is a rth-order ($r > d$) kernel function with compact support that satisfies the following conditions: $\int k(u)du = 1$, $\int u^m k(u)du = 0$ for $m = 1, 2, \ldots, r - 1$, $\int u^r k(u)du \neq 0$ and $\int k^2(u)du < \infty$. As well, for any $\mathbf{u} = (u_1, u_2, \ldots, u_d) \in R^d$, define $K_h(\mathbf{u}) = \frac{1}{h^d}\prod_{i=1}^{d} k(u_i/h)$, where h is a bandwidth sequence that satisfies $nh^{2r} \to 0$ and $nh^{2d} \to \infty$ as $n \to \infty$. The Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964) of $r(\mathbf{w})$ is then given by

$$\hat{r}(\mathbf{w}) = \hat{G}_n^{-1}(\mathbf{w})\frac{1}{n}\sum_{i=1}^{n} R_i\delta_i K_h(\mathbf{w}_1 - \mathbf{W}_{1i})I(\mathbf{W}_{2i} = \mathbf{w}_2), \quad (10)$$

where $\widehat{G}_n(w) = \frac{1}{n}\sum_{i=1}^n \delta_i K_h(w_1 - W_{1i})I(W_{2i} = w_2)$, $w = (w_1, w_2)$, and W_{1i} and W_{2i} are matrices that contain the continuous and discrete elements of W_i respectively. Substituting the estimator $\hat{\pi}(W_i)$ into (9) leads to the following IPW estimating equation for β:

$$U_1(\beta) \equiv n^{-3/2} \sum_{i=1}^n \sum_{j=1}^n \frac{R_i}{\hat{\pi}(Q_i)} \delta_i I(J_i = 2)(Z_i - Z_j) S\left(\frac{r_j^\beta - r_i^\beta}{\sigma_n} \right) = 0, \quad (11)$$

where $\hat{\pi}(Q_i) = \delta_i \hat{\pi}(W_i) + (1 - \delta_i)$.

Denote the solution of (11) as $\hat{\beta}_{IPW}$. The development of an asymptotic theory for $\hat{\beta}_{IPW}$ (as well as that of the other estimators in the subsequent sections) requires the following conditions:

(C1) The covariate vector, Z_1, is bounded, and there exists a constant M such that, $\| E(Z_1 - Z_2)(Z_1 - Z_2)^T \| < M < \infty$, and the parameter β lies in a compact set $\mathcal{B}$.

(C2) The sequence σ_n satisfies the conditions: $n\sigma_n \to \infty$ and $n\sigma_n^4 \to 0$ as $n \to \infty$.

(C3) The local distribution function $S(u)$ is continuous with respect to u, and its first derivative $s(u)$ satisfies the condition $\int u^2 s(u)du < \infty$ and is symmetric about zero.

(C4) The bandwidth h satisfies the conditions: $nh^{2r} \to 0$ and $nh^{2d} \to \infty$ as $n \to \infty$.

(C5) The matrix $A = \nabla \tilde{U}_0(\beta_0)$ exists and is nonsingular, where $\tilde{U}_0(\beta) = \lim_{n \to \infty} \frac{1}{\sqrt{n}}\tilde{U}_n(\beta)$,

(C6) Denote $f_{01}(\cdot)$ and $f_{02}(\cdot)$ as the density functions of $\log(T_{11}) - Z_1^T\beta_0$ and $\log(T_{12}) - Z_1^T\beta_0$ respectively. Then $f_{01}(\cdot)$, $f'_{01}(\cdot)$, $f_{02}(\cdot)$ and $f'_{02}(\cdot)$ are bounded functions on $\mathcal{R}$ with

$$\int_{-\infty}^{\infty} \left\{ \frac{f'_{01}(t)}{f_{01}(t)} \right\}^2 f_{01}(t)dt < \infty$$

and

$$\int_{-\infty}^{\infty} \left\{ \frac{f'_{02}(t)}{f_{02}(t)} \right\}^2 f_{02}(t)dt < \infty .$$

(C7) The distribution of $\log(C_1) - Z_1^T\beta_0$ is absolutely continuous and has a bounded density function $h(\cdot)$ on $\mathcal{R}$.

(C8) The function $g(\mathbf{w}) = P(\mathbf{W}_1 = \mathbf{w}, \delta_1 = 1)$ is bounded away from zero. Also, $g(\mathbf{w})$ has r continuous and bounded partial derivatives with respect to the continuous components of $\mathbf{W}_1$ almost surely.

(C9) The conditional probabilities $r(\mathbf{w}) = P(R_1 = 1|\delta_1 = 1, \mathbf{W}_1 = \mathbf{w})$ and $\rho(\mathbf{w}) = P(J_1 = 2|\delta_1 = 1, \mathbf{W}_1 = \mathbf{w})$ are bounded away from zero, and have r continuous and bounded partial derivatives with respect to the continuous components of $\mathbf{W}_1$ almost surely.

Let $\mathbf{S}_i = (\mathbf{Q}_i^T, R_i, J_i)^T$, $i = 1, 2, \ldots, n$, $\mathbf{h}(\mathbf{S}_i, \mathbf{S}_j) = \delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$, and $\mathbf{H}(\mathbf{S}_i, \mathbf{S}_j) = \mathbf{h}(\mathbf{S}_i, \mathbf{S}_j) + \mathbf{h}(\mathbf{S}_j, \mathbf{S}_i)$. Then we have the following theorem on the asymptotic properties of $\hat{\boldsymbol{\beta}}_{IPW}$:

Theorem 1 Let conditions (C1)-(C9) hold. Then $\hat{\boldsymbol{\beta}}_{IPW} \xrightarrow{p} \boldsymbol{\beta}_0$ and

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{IPW} - \boldsymbol{\beta}_0) \xrightarrow{d} N\{0, A^{-1}(\boldsymbol{\beta}_0)\sum(\boldsymbol{\beta}_0)(A^{-1}(\boldsymbol{\beta}_0))^T\},$$

where "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote, respectively, convergence in probability and convergence in distribution,

$$\sum(\boldsymbol{\beta}_0) = \boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0) + \boldsymbol{\Gamma}_2(\boldsymbol{\beta}_0), \boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0) = E(\mathbf{H}_1(\mathbf{S}_1))^{\otimes 2}, \mathbf{H}_1(\mathbf{S}_1) = E(\mathbf{H}(\mathbf{S}_1, \mathbf{S}_2)|\mathbf{S}_1), \boldsymbol{\Gamma}_2(\boldsymbol{\beta}_0) = E\left[(1 - r(\mathbf{W}_i))r^{-1}\right.$$
$$\left.(\mathbf{W}_i)\rho(\mathbf{W}_i)(1 - \rho(\mathbf{W}_i))\delta_i \varphi^{\otimes 2}(\mathbf{W}_i)\right], and \varphi(\mathbf{w}) = E\left\{(\mathbf{Z}_1 - \mathbf{Z}_2)I\{r_2^{\boldsymbol{\beta}_0} \geq r_1^{\boldsymbol{\beta}_0}\}|\mathbf{W}_1 = \mathbf{w}, \delta_1 = 1\right\}.$$

Proof: See the Appendix.

Now, for $i, j = 1, 2, \ldots, n$, define

$$e_{ij}^{(1)\boldsymbol{\beta}} = \frac{R_i}{\hat{\pi}(\mathbf{Q}_i)}\delta_i I(J_i = 2)I\{r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}}\}, \text{ and } d_{ij}^{(1)\boldsymbol{\beta}} = \left(\frac{R_i}{\hat{r}(\mathbf{W}_i)} - 1\right)\delta_i \hat{\rho}(\mathbf{W}_i)I\{r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}}\},$$

where $\hat{\rho}(\mathbf{W}_i)$ is defined in (13). Then from the proof of Theorem 1 in the Appendix and the theory of U-statistic (van der Vaart, 2000, Ch.12), we can show that the asymptotic variance of $\hat{\boldsymbol{\beta}}_{IPW}$ can be consistently estimated by $n^{-1}\hat{A}_{1n}^{-1}(\hat{\boldsymbol{\beta}}_{IPW})\hat{\sum}_{1n}(\hat{\boldsymbol{\beta}}_{IPW})(\hat{A}_{1n}^{-1}(\hat{\boldsymbol{\beta}}_{IPW}))^T$, where

$$\hat{A}_{1n}(\boldsymbol{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{\hat{\pi}(\mathbf{Q}_i)}\sigma_n^{-1}\delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j)^{\otimes 2}s\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n}\right), \text{and } \widehat{\sum}_{1n}(\boldsymbol{\beta}) = \frac{1}{n^3}\sum_i\sum_j\sum_{k \neq j}(\mathbf{Z}_i - \mathbf{Z}_j$$
$$)(\mathbf{Z}_i - \mathbf{Z}_k)^T(e_{ij}^{(1)\boldsymbol{\beta}} - e_{ji}^{(1)\boldsymbol{\beta}} - d_{ij}^{(1)\boldsymbol{\beta}})(e_{ik}^{(1)\boldsymbol{\beta}} - e_{ki}^{(1)\boldsymbol{\beta}} - d_{ik}^{(1)\boldsymbol{\beta}}).$$

3.2 Estimating equations imputation

A second method to consider is based on the estimating equations imputation (EEI) approach proposed by Zhou, Wan and Wang (2008). The EEI method is closely allied to the missing information principle. Let $\rho(W_i) = P(J_i = 2 | \delta_i = 1, W_i) = P(J_i = 2 | R_i = 1, \delta_i = 1, W_i)$. Noting that $E[R_i N_i(t) + (1 - R_i) E\{N_i(t) | Q_i\}] = E[N_i(t)]$, it can be readily shown that

$$M_i^{(2)}(t) \equiv R_i N_i(t) + (1 - R_i) E\{N_i(t) | Q_i\} - \int_{-\infty}^{t} Y_i(u) \lambda(u) du = R_i N_i(t) + (1 - R_i) \delta_i \rho(W_i) N_i^*(t) - \int_{-\infty}^{t} Y_i(u) \lambda(u$$

$$)du, i = 1, 2, \ldots, n,$$

are mean zero processes, where $N_i^*(t) = I\{\log(T_{i2}) - Z_i^T \beta \le t\}$. We can then obtain the following estimating equation by similar argument to those in Section 2:

$$n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i [R_i I(J_i = 2) + (1 - R_i) \rho(W_i)](Z_i - Z_j) S\left(\frac{r_j^\beta - r_i^\beta}{\sigma_n}\right) = 0. \quad (12)$$

In practice, $\rho(W_i)$ may be unknown. Analogous to the kernel estimator of $r(w)$ in Section 3.1, the estimator of $\rho(w)$ is

$$\hat{\rho}(w) = \widehat{M}_n^{-1}(W_i) \frac{1}{n} \sum_{i=1}^{n} I(J_i = 2) R_i \delta_i K_h(w_1 - W_{1i}) I(W_{2i} = w_2), \quad (13)$$

where $\widehat{M}_n(w) = \frac{1}{n} \sum_{i=1}^{n} R_i \delta_i K_h(w_1 - W_{1i}) I(W_{2i} = w_2)$. Thus, the EEI estimator $\hat{\beta}_{EEI}$ is the solution of the estimating equation

$$U_2(\beta) \equiv n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i [R_i I(J_i = 2) + (1 - R_i) \hat{\rho}(W_i)](Z_i - Z_j) S\left(\frac{r_j^\beta - r_i^\beta}{\sigma_n}\right) = 0. \quad (14)$$

The following theorem provides the asymptotic properties of $\hat{\beta}_{EEI}$.

Theorem 2 Let conditions (C1)-(C9) be satisfied. Then $\hat{\beta}_{EEI} \xrightarrow{p} \beta_0$ and

$$\sqrt{n}(\hat{\beta}_{EEI} - \beta_0) \xrightarrow{d} N\{0, A^{-1}(\beta_0) \sum (\beta_0)(A^{-1}(\beta_0))^T\},$$

where $\Sigma(\beta_0)$ is defined in Theorem 1.

Proof: See the Appendix.

It is straightforward to show, from the proof of Theorem 2, that the asymptotic variance of $\hat{\boldsymbol{\beta}}_{EEI}$ can be consistently estimated by $n^{-1}\hat{A}_{2n}^{-1}(\hat{\boldsymbol{\beta}}_{EEI})\hat{\boldsymbol{\Sigma}}_{2n}(\hat{\boldsymbol{\beta}}_{EEI})(\hat{A}_{2n}^{-1}(\hat{\boldsymbol{\beta}}_{EEI}))^{T}$, where

$$\hat{A}_{2n}(\boldsymbol{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i[R_iI(J_i=2)+(1-R_i)\hat{\rho}(\boldsymbol{W}_i)]\sigma_n^{-1}(\boldsymbol{Z}_i-\boldsymbol{Z}_j)^{\otimes 2}s\left(\frac{r_j^{\boldsymbol{\beta}}-r_i^{\boldsymbol{\beta}}}{\sigma_n}\right), \quad \widehat{\boldsymbol{\Sigma}}_{2n}(\boldsymbol{\beta}) = \frac{1}{n^3}\sum_i\sum_j\sum_{k\neq j}$$

$$(\boldsymbol{Z}_i-\boldsymbol{Z}_j)(\boldsymbol{Z}_i-\boldsymbol{Z}_k)^T(e_{ij}^{(2)\boldsymbol{\beta}}-e_{ji}^{(2)\boldsymbol{\beta}}+d_{ij}^{(2)\boldsymbol{\beta}})(e_{ik}^{(2)\boldsymbol{\beta}}-e_{ki}^{(2)\boldsymbol{\beta}}+d_{ik}^{(2)\boldsymbol{\beta}}), e_{ij}^{(2)\boldsymbol{\beta}}=\delta_i[R_iI(J_i=2)+(1-R_i)\hat{\rho}(\boldsymbol{W}_i)]I\{r_j^{\boldsymbol{\beta}}$$

$$\geq r_i^{\boldsymbol{\beta}}\}, i,j=1,2,...,n, \text{ and } d_{ij}^{(2)\boldsymbol{\beta}}=[I(J_i=2)-\rho(\boldsymbol{W}_i)]R_i\delta_i\frac{1-\hat{r}(\boldsymbol{W}_i)}{\hat{r}(\boldsymbol{W}_i)}I\{r_j^{\boldsymbol{\beta}}\geq r_i^{\boldsymbol{\beta}}\}, i,j=1,2,...,n.$$

3.3 Augmented inverse probability weighted estimator

Another common approach for handling data with missing values is the augmented inverse probability weighted (AIPW) method. The AIPW estimator has the so-called double robustness property, that is, the estimator is consistent provided that either $\rho(\boldsymbol{W}_i)$ or $r(\boldsymbol{W}_i)$ is specified correctly (Robins, Rotnitzky and Zhao, 1994; Wang and Chen, 2001).

Using theories developed by Robins, Rotnitzky and Zhao (1994), and noting that $E\left[\frac{R_i}{\pi(\boldsymbol{Q}_i)}N_i(t)+(1-\frac{R_i}{\pi(\boldsymbol{Q}_i)})E\{N_i(t)|\boldsymbol{Q}_i\}\right]=E[N_i(t)]$, it follows that

$$M_i^{(3)}(t) \equiv \frac{R_i}{\pi(\boldsymbol{Q}_i)}N_i(t)+(1-\frac{R_i}{\pi(\boldsymbol{Q}_i)})E\{N_i(t)|\boldsymbol{Q}_i\}-\int_{-\infty}^{t}Y_i(u)\lambda(u)du, i=1,2,...,n,$$

are mean zero processes. We have, analogous to the analysis in Section 2, the following AIPW estimating equations for $\boldsymbol{\beta}$:

$$\boldsymbol{U}_3(\boldsymbol{\beta}) \equiv n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i[\frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)}I(J_i=2)+(1-\frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)})\hat{\rho}(\boldsymbol{W}_i)](\boldsymbol{Z}_i-\boldsymbol{Z}_j)S\left(\frac{r_j^{\boldsymbol{\beta}}-r_i^{\boldsymbol{\beta}}}{\sigma_n}\right)=0,$$

(15)

where $\hat{\pi}(\boldsymbol{Q}_i)=\delta_i\hat{r}(\boldsymbol{W}_i)+(1-\delta_i)$ and $\hat{\rho}(\boldsymbol{W}_i)$ are defined in (10) and (13) respectively.

Let the solution of (15) be $\hat{\boldsymbol{\beta}}_{AIPW}$. The asymptotic properties of $\hat{\boldsymbol{\beta}}_{AIPW}$ are presented in the following theorem:

Theorem 3 Let conditions (C1)-(C9) be satisfied. Then $\hat{\beta}_{AIPW} \xrightarrow{p} \beta_0$ and

$$\sqrt{n}(\hat{\beta}_{AIPW} - \beta_0) \xrightarrow{d} N\{0, A^{-1}(\beta_0) \sum (\beta_0)(A^{-1}(\beta_0))^T\},$$

where $\Sigma(\beta_0)$ is specified in Theorem 1.

Proof: See the Appendix.

From the proof of Theorem 3, a consistent estimator of the asymptotic variance of $\hat{\beta}_{AIPW}$ is given by $n^{-1}\widehat{A}_{3n}^{-1}(\hat{\beta}_{AIPW}) \overset{\wedge}{\sum}_{3n}(\hat{\beta}_{AIPW})(\widehat{A}_{3n}^{-1}(\hat{\beta}_{AIPW}))^T$, where

$$\widehat{A}_{3n}(\beta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_i [\frac{R_i}{\hat{\pi}(Q_i)} I(J_i = 2) + (1 - \frac{R_i}{\hat{\pi}(Q_i)})\hat{\rho}(W_i)]\sigma_n^{-1}(Z_i - Z_j)^{\otimes 2} s\left(\frac{r_j^{\beta} - r_i^{\beta}}{\sigma_n}\right), \widehat{\sum}_{3n}(\beta)$$

$$= \frac{1}{n^3} \sum_i \sum_j \sum_{k \neq j} (Z_i - Z_j)(Z_i - Z_k)^T (e_{ij}^{(3)\beta} - e_{ji}^{(3)\beta})(e_{ik}^{(3)\beta} - e_{ki}^{(3)\beta}), \text{and} e_{ij}^{(3)\beta} = \delta_i[\frac{R_i}{\hat{\pi}(Q_i)} I(J_i = 2) + (1 - \frac{R_i}{\hat{\pi}(Q_i)}$$

$$)\hat{\rho}(W_i)]I\{r_j^{\beta} \geq r_i^{\beta}\}, i, j = 1, 2, ..., n.$$

Remark 1 The three estimators $\hat{\beta}_{IPW}$, $\hat{\beta}_{EEI}$ and $\hat{\beta}_{AIPW}$ are asymptotically equivalent and thus have the same asymptotic efficiency. This is a surprising result because we would expect the *AIPW* method, which combines the *IPW* and *EEI* approaches, to have an improved level of efficiency over the other two methods.

Remark 2 The implementation of the *IPW* and *AIPW* methods requires the estimation of the missing probability $\pi(Q_i)$. On the other hand, unless one wants to estimate the asymptotic variance directly, the *EEI* procedure does not involve the estimation of $\pi(Q_i)$. If we resort to a re-sampling method for estimating the asymptotic variance, then the estimation of $\pi(Q_i)$ will not be required for the *EEI* method. Thus, from a computational point of the view, the *EEI* method has an advantage over the *IPW* and *AIPW* methods.

4 Selection of Kernel Functions and Smoothing Parameters, and Dimension Reduction

4.1 Selection of kernel functions and smoothing parameters

In this section, we discuss the selection of the kernel functions $S(\cdot)$ and $k(\cdot)$ and the smoothing parameters σ_n and h. In all our numerical studies and the real data example, we use the standard Gaussian cumulative distribution function as $S(u)$, as the local distribution function. A recent study on the AFT model under length-biased sampling by Qiu, Qin and Zhou (2016) shows that the finite sample properties of estimators are generally insensitive to the choice of the local distribution function. As for the choice of σ_n, there exist many studies, including Song et al. (2007), Ma and Huang (2007), Lin and Peng (2013), and Qiu,

Qin and Zhou (2016), who showed that smoothing approximation techniques similar to ours are applicable under a wide range of choices of σ_n. Here, we use the "rule of thumb" approach along the lines of Heller (2007) to select this smoothing parameter. Specifically, we set $\sigma_n = \hat{c}n^{-0.26}$, where $\hat{c} = \left[\sum_{i=1}^n R_i I\{J_i \delta_i = 2\}(r_i^{\hat{\beta}_I} - \bar{r}^{\hat{\beta}_I})^2 / (\sum_{i=1}^n R_i I\{J_i \delta_i = 2\} - 1) \right]^{1/2}$,

$r_i^{\hat{\beta}_I} = \log \tilde{T}_i - Z_i^T \hat{\beta}_I$, $\hat{\beta}_I$ is an intial estimator obtained by solving the following estimating equation

$$n^{-3/2} \sum_{i=1}^n \sum_{j=1}^n R_i \delta_i I(J_i = 2)(Z_i - Z_j) I\{r_j^{\beta} \geq r_i^{\beta}\} = 0,$$

and $\bar{r}^{\hat{\beta}_I} = \sum_{i=1}^n R_i I\{J_i \delta_i = 2\} r_i^{\hat{\beta}_I} / \sum_{i=1}^n R_i I\{J_i \delta_i = 2\}$; the purpose of imposing the power constant -0.26 for n in σ_n is to satisfy condition (C2).

The generalised cross-validation method can be used for choosing the bandwidth h when estimating $r(W_i)$ and $\rho(W_i)$. Here, following Wang and Wang (2001) and Qi, Wang and Prentice (2005), we set $h = O(n^{-1/q})$ with $q > 2d$ and the smallest even integer for r such that $r \quad q - d$. More specifically, when the number of continuous elements in W_i is equal to 1, i.e., $d = 1$, we use the univariate second order Epanechnikov kernel $k(u) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}u^2)I(u^2 < 5)$ and set the bandwidth to $h = 4\sigma_{\tilde{T}}n^{-1/3}$, where $\sigma_{\tilde{T}}$ is the sample standard deviation of the observed survival times. When $d = 2$, we use the fourth order Epanechnikov kernel $k(u) = \frac{3}{4\sqrt{5}}(\frac{15}{8} - \frac{7}{8}u^2)(1 - \frac{1}{5}u^2)I(u^2 < 5)$ and set the bandwidth to $(h_1, h_2)^T = (4\sigma_{\tilde{T}}n^{-1/5}, 4\sigma_Z n^{-1/5})^T$, where σ_Z is the sample standard deviation of Z_i. We use this method to select the kernel function $k(u)$ and the smoothing parameters h in the simulation studies and the real example.

4.2 Dimension reduction

The three methods proposed by us are based on non-parametric regression. It is well-known that non-parametric methods all suffer from the curse of dimensionality, meaning that the performance of the methods will deteriorate rapidly as the dimension of the covariates increases. This limits the usefulness of our methods. An alternative is estimate $\pi(W_i)$ and $\rho(W_i)$ parametrically, along the lines of Gao and Tsiatis (2005), Lu and Liang (2008), Sun, Wang and Gilbert (2012), Zheng, Lin and Yu (2016), and others. However, parametric methods can result in substantially biased estimators when the correctness of the parametric specifications is called into question (Han, 2014).

Dimension reduction is one way to circumvent the problem caused by the curse of dimensionality. The objective is to seek low dimensional variables U_1 and U_2 in the observed data such that $E(R|U_1) = E(R|\delta, W) = E(R|Q)$ and $P(J = 2|U_2) = P(J = 2|\delta, W) = P(J = 2|Q)$. It is easy to see that if we replace $E(R|Q)$ and $P(J = 2|Q)$ in the estimating equations

pertaining to our methods by $E(R|U_1)$ and $P(J=2|U_2)$ respectively, the estimating equations remain unbiased.

Many methods have been developed for the selection of U_1 and U_2. For example, we may assume $E(R|Q) = g_1(Q^T\theta_1)$, and $P(J=2|Q) = g_2(Q^T\theta_2)$, where $g_1(\cdot)$ and $g_2(\cdot)$ are unknown functions, and θ_1 and θ_2 are parameters that can be estimated, for example, by sliced inverse regression (Li, 1991). Then $g_1(\cdot)$ and $g_2(\cdot)$ can be estimated by univariate kernel smoothing techniques. Furthermore, other flexible parametric models, such as the generalised additive and the partially linear models, can be used to model the conditional probabilities $E(R|Q)$ and $P(J=2|Q)$. Clearly, the asymptotic properties of the estimators resulting from dimension reduction procedures will differ from those developed in Section 3. It remains an interesting topic for future research to develop the asymptotic properties of estimators under this alternative approach.

5 A Simulation Study

In this section, we focus on finite sample properties and identify, in the context of a simulation experiment, estimation and inference properties of the methods developed in this paper. We also draw comparisons of our methods with the complete-case analysis, which uses only observations that have the failure cause observed.

Experiment 1

Our first experiment is based on the following model containing only one covariate:

$$\log(T_{i2}) = Z_i + \varepsilon_i, i = 1, \ldots, n,$$

where T_{i2} is the failure time associated with the cause of interest, Z_i is a covariate that follows either a Bernoulli(0.5) distribution or a $U[0, 1]$ distribution, and ε_i is an error term following one of the following three distributions: $N(0, 0.5^2)$, $U[-0.5, 0.5]$ and the Generalised Extreme Value $GEV(0, 0.5, 0)$ distributions. All observations of ε_i are converted into mean deviation form in our simulations. Given Z_i, we let $\Phi(\log t - \gamma Z_i)$ be the conditional distribution function of the failure time of the other cause T_{i1}, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and γ is chosen such that the failure of interest arises approximately 60% of the time. The censoring time C_i is generated from $U[0, c]$, where c is a constant that controls the censoring percentage. In all cases, we choose c such that the censoring percentage is about 30%. Depending on the distributional settings of Z_i and ε_i, the percentages of failures due to the cause of interest and the other cause vary between 40% and 42%, and between 28% and 30% respectively. We set $n = 200$ when Z_i follows a Bernoulli(0.5) distribution and $n = 400$ when Z_i follows a $U[0, 1]$ distribution. In addition, we consider the following two missing data scenarios: Scenario 1: $r(W_i) = \exp(\tilde{T}_i - Z_i)/\{1 + \exp(\tilde{T}_i - Z_i)\}$, and Scenario 2: $r(W_i) = 0.5$. Under Scenario 1, the missing percentage varies between 70% and 72% depending on the setting of ε_i, whereas under the second scenario, the missing percentage is approximately 50%.

Our simulation results based on 1000 replications are reported in Table 1, where FULL, CC, IPW, EEI and AIPW refer to results based on the full data set with no missing failure cause, complete-case study, inverse probability weighting, estimating equations imputation and augmented inverse probability weighting respectively, and BIAS, SE, SD and CP denote the empirical bias, the mean of estimated standard error, the empirical standard deviation and the proximity of empirical coverage probability of confidence interval (C.I.) corresponding to the nominal 95% level.

Our results show that by and large, the CC method results in the largest bias and smallest C.I. coverage probability. Of the three proposed methods, the IPW method frequently exhibits the largest bias, but it also yields C.I. coverage probability that is as accurate as those produced by the EEI and AIPW approaches. The biases resulting from the EEI and AIPW approaches are usually quite small and the two approaches also achieve very accurate C.I. coverages. Our results do not suggest any clear preference between the EEI and AIPW approaches; generally speaking, the two approaches lead to indistinguishable results and there is little to choose between them. In all cases, the SE's and their corresponding SD's are very close, indicating that the various non-parametric procedures we use at different stages perform well. As expected, the benchmark estimator based on the full set of data with no missing cause of failure performs best under all performance dimensions being considered. There are no obvious differences between the three types of error distributions, *ceteris paribus*.

Experiment 2

Our second experiment is based on the following model with two covariates:

$$\log(T_{i2}) = \beta_{01} Z_{i1} + \beta_{02} Z_{i2} + \varepsilon_i,$$

where $Z_{i1} \sim U[0, 1]$, $Z_{i2} \sim \text{Bernoulli}(0.5)$ and ε_i is an error term following one of the same three distributions as in Experiment 1. The distribution function of T_{i1}, given Z_{i1} and Z_{i2}, is $\Phi(\log t - \boldsymbol{\gamma}^T \boldsymbol{Z}_i)$, where $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2})^T$. The censoring time C_i is generated from $U[0, c]$, where c is a constant parameter that controls the censoring percentage. As in Experiment 1, we choose $\boldsymbol{\gamma}$ and c such that, on average, 40% of failures are due to the cause of interest, 30% of failures are due to the other cause and the censoring percentage is about 30%. We consider the following missing data scenarios: Scenario 1: $r(\boldsymbol{W}_i) = \exp(4Z_{i1} + 3Z_{i2} - \tilde{T}_i)/\{1 + \exp(4Z_{i1} + 3Z_{i2} - \tilde{T}_i)\}$; Scenario 2: $r(\boldsymbol{W}_i) = 0.5$; Scenario 3: $r(\boldsymbol{W}_i) = 1/\{1 + \exp(Z_{i1}^2 - 2Z_{i2})\}$. For Scenario 1, the missing probability is approximately 65.9% when $\varepsilon_i \sim N(0, 0.25)$, and 64.3% when $\varepsilon_i \sim GEV(0, 0.5, 0)$ and $\varepsilon_i \sim U[-0.5, 0.5]$. For the other two Scenarios, the missing probabilities are approximately 50% and 59% respectively. In all cases, we set $n = 400$ and the number of replications to 1000.

The results are presented in Table 2. By and large, the general comments made above for Experiment 1, where the model contains a single covariate, also apply to the two-covariate case in broad terms. Specifically, the CC method results in estimates with the largest bias in the majority of cases; the IPW, EEI and AIPW methods generally yield comparable results

although the IPW method tends to result in slightly larger estimator bias than the other two methods. Other things being equal, the form of the error distributions does not appear to impact the results significantly.

Experiment 3

This experiment is conducted in response to a query by a referee about the difference in computational time between the proposed smooth approach and the discontinuous rank approach. This experiment is based on the same setup as in Experiment 1, except that we confine our attention to Z_i following a $U[0, 1]$ distribution, ε_i following a $N(0, 0.5^2)$ distribution, and the missing data mechanism of $r(W_i) = \exp(\tilde{T}_i - Z_i)/\{1 + \exp(\tilde{T}_i - Z_i)\}$. Thus, on average, 42% of failures are due to the cause of interest, 28% of failures are due to the other cause, and the data missing percentage is approximately 70%. We only report results based on the IPW missing data handling method. Results based on other missing data handling methods are similar and we omit them for brevity. For the non-smooth approach, estimates of the unknown parameters are obtained as solutions to the estimating equation:

$$n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\hat{\pi}(Q_i)} I(J_i = 2)\delta_i (Z_i - Z_j) I\{e_j^{\boldsymbol{\beta}} \geq e_i^{\boldsymbol{\beta}}\} = 0,$$

where $e_i^{\boldsymbol{\beta}} = \log(\tilde{T}_i) - Z_i^T \boldsymbol{\beta}$, $i = 1, 2, \cdots, n$. Note that the l.h.s. of above equation is the gradient of the following convex function

$$L(\boldsymbol{\beta}) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\hat{\pi}(Q_i)} I(J_i = 2)\delta_i (e_i^{\boldsymbol{\beta}} - e_j^{\boldsymbol{\beta}})^-,$$

where $a^- = |a|I\{a < 0\}$. We use the package "fminsearch" in Matlab to minimise $L(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and obtain the estimates of $\boldsymbol{\beta}$. As discussed previously, for estimating the asymptotic covariance of the estimator, we have to resort to resampling (Jin, Lin, Wei and Ying, 2003). Specifically, we first construct the perturbed objective function

$$L^*(\boldsymbol{\beta}) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\hat{\pi}(Q_i)} I(J_i = 2)\delta_i (e_i^{\boldsymbol{\beta}} - e_j^{\boldsymbol{\beta}})^- \xi_i,$$

where ξ_i follows the exponential distribution with mean 1. Given the data $(Q_i^T, R_i, J_i)^T$, we repeat the resampling process 50 times, and use the standard deviation (SD) of the 50 re-sampled estimates to compute the standard errors of the estimate (SE). Table 3 below reports the results for $n = 400$ observations based on 1000 replications. It can be seen that in addition to delivering more accurate estimates, the smoothed approach has a significant advantage over the non-smoothed approach in terms of computational time.

Experiment 4

This experiment provides insights on the impact of bandwidth choices on the results. This experiment is conducted in response to a question by a referee. Related studies by Ma and Huang (2007), Song et al. (2007), Lin and Peng (2013) and Qiu, Qin and Zhou (2016) have shown that bandwidth choices do not impact the results significantly. Here, we conduct a simple simulation experiment to examine the sensitivity of results to bandwidth choices. We consider the same setup as in Experiment 2, except that we restrict our attention to $\varepsilon_i \sim N(0, 0.25)$ and Scenario 1 of the missing data mechanism. We set the smoothing parameter σ_n to $0.1 \times n^{-0.26}$, $0.3 \times n^{-0.26}$, $0.5 \times n^{-0.26}$, $0.7 \times n^{-0.26}$ and $0.9 \times n^{-0.26}$. The results presented in Table 4 show that for a given estimation method, the results across the different bandwidths are very similar. The assignment of bandwidth σ_n is thus straightforward and does not involve any search.

6 Application to the HVTN 502 Phase IIb 'Step' HIV Vaccine Efficacy Trial

We apply the newly developed methods to the HVTN 502 'Step' Phase IIb trial, which was a randomised, placebo-controlled, preventive vaccine efficacy trial that enrolled HIV-1 uninfected men who have sex with men who were at high risk for acquiring HIV-1 infection, and hadprimary objective to assess whether the incidence of HIV-1 infection differed between the two treatment groups [active vaccination with the Merck adenovirus type 5 (Ad5) vector vaccine (named MRKAd5) vs. placebo](Buchbinder et al., 2008). The Step trial enrolled 1836 HIV-1 uninfected men, of whom 88 acquired the primary study endpoint of HIV-1 infection (52 in the vaccine group and 36 in the placebo group). The primary analysis assessed the vaccine effect on the time to HIV-1 infection with a Cox model, yielding an estimated hazard ratio (vaccine vs. placebo) of 1.50 (95% C.I.: 0.95–2.41, p-value = 0.06), suggesting that, unfortunately, the vaccine elevated the risk of HIV-1 infection.

HIV-1 is extraordinarily genetically diverse, with many genetic types of HIV-1 exposing participants in the Step trial, and a secondary objective of the Step trial was to assess the vaccine effect on the time to HIV-1 infection with specific genetic types of HIV-1. Based on measurement of the HIV-1 sequences from Step participants who had the HIV-1 infection endpoint, there are many ways to define genetic types. Once a definition is specified– such that there are K mutually exclusive and exhaustive genetic types– then the objective at hand is a standard competing risks failure time problem, where T is the time to the first HIV-1 infection and J is the genetic type of the HIV-1 infection, $J \in \{1, \cdots, K\}$. However, HIV-1 sequences were only successfully obtained from 65 of the 88 HIV-1 infected participants, such that J is missing for 23 participants, and a method that handles missing failure causes is needed. Therefore the data set-up fits the purpose for which the newly proposed methods were designed. In addition to needing a method to handle the missing outcome type J from 23 HIV-1 infected participants, a method is needed to account for the fact that the vaccine effect on the incidence of HIV-1 infection appeared to wane over time (Duerr et al., 2012), which, casts doubt about the suitability of the Cox model and motivates use of the AFT model developed in the current manuscript. Because the previous analysis applied a Cox

model to address the secondary objective (Rolland et al., 2011), the newly proposed methods may be a better fit to the application.

We now describe the genetic HIV-1 type J that was analyzed. It is of particular scientific interest to study the vaccine effect on infection with the HIV-1 genetic type defined by high amino acid dissimilarity to a 'hotspot' span of 30 contiguous amino acids in the Gag HIV-1 protein sequence inside the vaccine construct that was targeted by vaccine-induced T cell responses (Hertz et al., 2013); accordingly we define $J = 2$ as 2 or more mismatches of the HIV-1 infected participants hotspot sequence with the corresponding hotspot sequence in the vaccine (based on a multiple sequence alignment). Thenall HIV-1infections with genetic types with 0 or 1 mismatches have $J=1$. The distribution of J across the 88 endpoints is 7 for $J = 1$, 32 for $J = 2$, and 14 missing for HIV-1 infected vaccine recipients, and 5 for $J = 1$, 21 for $J = 2$, and 9 missing for HIV-1 infected placebo recipients.

We employ the AFT model to evaluate the effect of Treatment (Treatment=1, if the participant was assigned to receive the MRKAd5 vaccine, Treatment=0 placebo), on the failure time T, where T is defined as the number of days from randomisation to diagnosis of HIV-1 infection due to the genotype of interest $J = 2$. We also included in the model the demographic factors Age (in years at study entry) and WhiteRace (indicator of reporting white race).

Table 5 reports the estimation results. By all methods, Treatment is statistically significant whereas WhiteRace is not. The results for Treatment show that vaccine recipients have a shorter mean time to diagnosis with genotype $J = 2$ HIV-1 infection than placebo recipients, suggesting that vaccination increased susceptibility to acquisition of $J = 2$ HIV-1 genotypes. In addition, the EEI and AIPW methods found that Age was non-significant, but the CC and IPW methods suggest that Age was significant at the 5% level. Interestingly, the EEI and AIPW methods tended to produce estimates of similar magnitudes, and the same was observed for the CC and IPW methods. For a given coefficient, there was no sign difference in the estimates produced by any of the methods.

In conclusion, the analysis suggests that recipients of the MRKAd5 vaccine may have elevated risk of acquiring HIV-1 infection with HIV-1 genetic types that have too many mismatches to the genetic type represented inside the vaccine construct, when these mismatches occur in the HIV-1 Gag hotspot location to which the vaccine predominantly directs T cell responses. This highlights the importance of designing new HIV-1 vaccine regimens that direct immune responses to many different genetic types of HIV-1, to maximize overall vaccine efficacy of future HIV-1 vaccines.

7 Concluding Remarks

Competing risks are commonplace in clinical trial study. This paper has examined the AFT competing risk model with missing cause of failure using the monotone rank estimating equations approach combined with local distribution function smoothing, and developed three methods for estimating unknown regression coefficients. Our simulation study shows that the three methods work well, and the methods have been applied to two datasets on

bone marrow transplant and HIV vaccine efficacy. We have also discussed methods of dimension reduction that can be undertaken in conjunction with the methods developed when the number of covariates is large. Our proposed methods can also be extended to other semi-parametric models, such as the generalized transformation models and the mean residual lifetime model. These remain for future work.

Acknowledgments

Appendix: Proof of theorems

In our proof of theorems, for convenience purposes we assume that all elements of W_i are continuous. This assumption does not lead to any loss of generality.

Proof of Theorem 1. We divide the proof into two parts.

Part A1. We can write

$$\frac{1}{\sqrt{n}} \nabla U_1(\beta_0) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\pi(Q_i)} \delta_i I(J_i = 2)(Z_i - Z_j)^{\otimes 2} s\left(\frac{r_j^{\beta_0} - r_i^{\beta_0}}{\sigma_n}\right) \frac{1}{\sigma_n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \quad (7.16)$$

$$(\frac{1}{\hat{\pi}(Q_i)} - \frac{1}{\pi(Q_i)}) R_i \delta_i I(J_i = 2)(Z_i - Z_j)^{\otimes 2} s\left(\frac{r_j^{\beta_0} - r_i^{\beta_0}}{\sigma_n}\right) \frac{1}{\sigma_n},$$

where $s(\cdot)$ is the standard normal density function. By some tedious calculations and recognising the fact that $\sup_q |\hat{\pi}(q) - \pi(q)| = O_p\left(h^r + (nh^d)^{-\frac{1}{2}}\right)$, we obtain

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\frac{1}{\hat{\pi}(Q_i)} - \frac{1}{\pi(Q_i)}) R_i \delta_i I(J_i = 2)(Z_i - Z_j)^{\otimes 2} s\left(\frac{r_j^{\beta_0} - r_i^{\beta_0}}{\sigma_n}\right) \frac{1}{\sigma_n} = o_p(h^r + (nh^d)^{-\frac{1}{2}}).$$

For the first item on the r.h.s. of (7.16), note that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{\pi(\boldsymbol{Q}_i)}\delta_i I(J_i=1)(\boldsymbol{Z}_i-\boldsymbol{Z}_j)^{\otimes 2}s\left(\frac{r_j^{\boldsymbol{\beta}_0}-r_i^{\boldsymbol{\beta}_0}}{\sigma_n}\right)\frac{1}{\sigma_n}=\frac{n(n-1)}{n^2}\frac{1}{C_n^2}\sum_{i<j}\frac{1}{2}\{\frac{R_i}{\pi(\boldsymbol{Q}_i)}\delta_i I(J_i=2}$$

$$)(\boldsymbol{Z}_i-\boldsymbol{Z}_j)^{\otimes 2}s\left(\frac{r_j^{\boldsymbol{\beta}_0}-r_i^{\boldsymbol{\beta}_0}}{\sigma_n}\right)\frac{1}{\sigma_n}+\frac{R_j}{\pi(\boldsymbol{Q}_j)}\delta_j I(J_j=2)(\boldsymbol{Z}_j-\boldsymbol{Z}_i)^{\otimes 2}s\left(\frac{r_i^{\boldsymbol{\beta}_0}-r_j^{\boldsymbol{\beta}_0}}{\sigma_n}\right)\frac{1}{\sigma_n}\}.$$

By using the strong law of large numbers for U-statistics, we can obtain

$$\frac{1}{\sqrt{n}}\nabla\boldsymbol{U}_1(\boldsymbol{\beta}_0)\xrightarrow{a.s.}E\{(\boldsymbol{Z}_1-\boldsymbol{Z}_2)^{\otimes 2}[\int_{-\infty}^{\infty}\bar{H}(u)\bar{F}_{01}(u)f_{02}(u)\xi(u)du]+\lim_{n\to\infty}\left(\frac{m_n}{\pi}\right)^{\frac{1}{2}}\int_{-\infty}^{\infty}\tau(w)e^{-m_n w^2}dw]\},$$

where $\tau(w)=\int_{-\infty}^{\infty}\bar{H}(u)\bar{F}_{01}(u)f_{02}(u)\{\xi(u+w)-\xi(u)\}du$, $\xi(s)=f_{01}(s)\bar{F}_{02}(s)\bar{H}(s)+$

$f_{02}(s)\bar{F}_{01}(s)\bar{H}(s)+\bar{F}_{01}(s)\bar{F}_{02}(s)h(s)$, $m_n=n/(2\sigma_n^2)$, $\bar{F}_{01}(\cdot)$ is the survival function of $\log(T_{11})-$

$\boldsymbol{Z}^T\boldsymbol{\beta}_0$, $\bar{F}_{02}(\cdot)$ is the survival function of $\log(T_{12})-\boldsymbol{Z}^T\boldsymbol{\beta}_0$, and $\bar{H}(\cdot)$ is the survival function of

$\log(C_1)-\boldsymbol{Z}^T\boldsymbol{\beta}_0$. Under conditions (C1)-(C9), the function $\tau(\cdot)$ is integrable, continuous and

bounded on $\mathcal{R}$ with $\tau(0)=0$. Thus, the second term on the r.h.s. of (7.16) vanishes (Kanwal, 1998, p.11). Therefore, we have

$$\frac{1}{\sqrt{n}}\nabla\boldsymbol{U}_1(\boldsymbol{\beta}_0)\xrightarrow{a.s.}A=E\left\{(\boldsymbol{Z}_1-\boldsymbol{Z}_2)^{\otimes 2}\left[\int_{-\infty}^{\infty}\bar{H}(u)\bar{F}_{01}(u)f_{02}(u)\xi(u)du\right]\right\}.$$

Part A2. By the similar proof as Heller (2007), we note that

$$\boldsymbol{U}_1(\boldsymbol{\beta}_0)=\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_i}{\pi(\boldsymbol{Q}_i)}\delta_i I(J_i=2)(\boldsymbol{Z}_i-\boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0}\geq r_i^{\boldsymbol{\beta}_0}\}$$

$$-\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\hat{\pi}(\boldsymbol{Q}_i)-\pi(\boldsymbol{Q}_i)}{\pi^2(\boldsymbol{Q}_i)}R_i\delta_i I(J_i=2)(\boldsymbol{Z}_i-\boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0}\geq r_i^{\boldsymbol{\beta}_0}\}$$

$$+\frac{1}{n^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(\hat{\pi}(\boldsymbol{Q}_i)-\pi(\boldsymbol{Q}_i))^2}{\hat{\pi}(\boldsymbol{Q}_i)\pi^2(\boldsymbol{Q}_i)}R_i\delta_i I(J_i=2)(\boldsymbol{Z}_i-\boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0}\geq r_i^{\boldsymbol{\beta}_0}\}+o_p(1)\equiv\boldsymbol{U}_{11}(\boldsymbol{\beta}_0)-\boldsymbol{U}_{12}$$

$$(\boldsymbol{\beta}_0)+\boldsymbol{U}_{13}(\boldsymbol{\beta}_0)+o_p(1).$$

(7.17)

Now, by (C1) and (C8), and recognising that $\sup\limits_{\boldsymbol{q}} |\hat{\pi}(\boldsymbol{q}) - \pi(\boldsymbol{q})| = O_p\left(h^r + (nh^d)^{-\frac{1}{2}}\right)$, we have

$$\|\boldsymbol{U}_{13}(\boldsymbol{\beta}_0)\| \leq C\sqrt{n} \sup\limits_{\boldsymbol{q}} |\hat{\pi}(\boldsymbol{q}) - \pi(\boldsymbol{q})|^2 = O_p\left(\sqrt{n}h^{2r} + \frac{1}{\sqrt{nh^d}}\right), \quad (7.18)$$

where C is an arbitrary constant. Thus, by condition (C4), $\boldsymbol{U}_{13}(\boldsymbol{\beta}_0) = o_p(1)$.

From the definition of $\hat{\pi}(\boldsymbol{Q}_i)$, it follows that

$$\boldsymbol{U}_{12}(\boldsymbol{\beta}_0) = \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\hat{r}(\boldsymbol{W}_i) - r(\boldsymbol{W}_i))\hat{G}_n(\boldsymbol{W}_i)}{\pi^2(\boldsymbol{Q}_i)g(\boldsymbol{W}_i)} R_i\delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

$$- \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\hat{r}(\boldsymbol{W}_i) - r(\boldsymbol{W}_i))(\hat{G}_n(\boldsymbol{W}_i) - g(\boldsymbol{W}_i))}{\pi^2(\boldsymbol{Q}_i)g(\boldsymbol{W}_i)} R_i\delta_i I(J_i = 2) \times (\boldsymbol{Z}_i - \boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

$$= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (R_l - r(\boldsymbol{W}_i))\pi^{-2}(\boldsymbol{Q}_i)g^{-1}(\boldsymbol{W}_i)\delta_l K_h(\boldsymbol{W}_i - \boldsymbol{W}_l)R_i\delta_i I(J_i = 2) \times (\boldsymbol{Z}_i - \boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0}$$

$$\geq r_i^{\boldsymbol{\beta}_0}\} - \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\hat{r}(\boldsymbol{W}_i) - r(\boldsymbol{W}_i))(\hat{G}_n(\boldsymbol{W}_i) - g(\boldsymbol{W}_i))}{\pi^2(\boldsymbol{Q}_i)g(\boldsymbol{W}_i)} R_i\delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

$$\equiv \boldsymbol{U}_{121}(\boldsymbol{\beta}_0) + \boldsymbol{U}_{122}(\boldsymbol{\beta}_0).$$

$$(7.19)$$

But according to the facts that $\sup\limits_{\boldsymbol{w}} |\hat{r}(\boldsymbol{w}) - r(\boldsymbol{w})| = O_p\left(h^r + (nh^d)^{-\frac{1}{2}}\right)$ and

$\sup\limits_{\boldsymbol{w}} |\hat{G}_n(\boldsymbol{w}) - g(\boldsymbol{w})| = O_p\left(h^r + (nh^d)^{-\frac{1}{2}}\right)$, we obtain

$$\boldsymbol{U}_{122}(\boldsymbol{\beta}_0) = O_p\left(\sqrt{n}h^{2r} + \frac{1}{\sqrt{nh^d}}\right) = o_p(1). \quad (7.20)$$

Our next task is to prove

$$U_{121}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(\boldsymbol{W}_i)} - 1 \right) \delta_i \rho(\boldsymbol{W}_i) \boldsymbol{\varphi}(\boldsymbol{W}_i) + o_p(1), \quad (7.21)$$

where $\varphi(\boldsymbol{w}) = E\left\{ (\boldsymbol{Z}_1 - \boldsymbol{Z}_2) S\left(\frac{r_2^{\boldsymbol{\beta}_0} - r_1^{\boldsymbol{\beta}_0}}{\sigma_n} \right) \middle| \boldsymbol{W}_1 = w \right\}.$

Note that

$$U_{121}(\boldsymbol{\beta}_0) = U_{121}^{[1]}(\boldsymbol{\beta}_0) + U_{121}^{[2]}(\boldsymbol{\beta}_0), \quad (7.22)$$

where

$$U_{121}^{[1]}(\boldsymbol{\beta}_0) = \frac{1}{n^{\frac{5}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (R_l - r(\boldsymbol{W}_l)) \pi^{-2}(\boldsymbol{Q}_i) g^{-1}(\boldsymbol{W}_i) \delta_l K_h(\boldsymbol{W}_i - \boldsymbol{W}_l) \times R_i \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0}$$
$$\geq r_i^{\boldsymbol{\beta}_0}\}$$

and

$$U_{121}^{[2]}(\boldsymbol{\beta}_0) = \frac{1}{n^{\frac{5}{2}}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (r(\boldsymbol{W}_l) - r(\boldsymbol{W}_i)) \pi^{-2}(\boldsymbol{Q}_i) g^{-1}(\boldsymbol{W}_i) \delta_l K_h(\boldsymbol{W}_i - \boldsymbol{W}_l) \times R_i \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0}$$
$$\geq r_i^{\boldsymbol{\beta}_0}\}.$$

To analyse $U_{121}^{[1]}(\boldsymbol{\beta}_0)$, similar to Zhou, Wan and Wang (2008), let us define

$$h(\boldsymbol{S}_i, \boldsymbol{S}_j, \boldsymbol{S}_l) = (R_l - r(\boldsymbol{W}_l)) \pi^{-2}(\boldsymbol{Q}_i) g^{-1}(\boldsymbol{W}_i) \delta_l K_h(\boldsymbol{W}_i - \boldsymbol{W}_l) \times R_i \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

and

$$H(\boldsymbol{S}_i, \boldsymbol{S}_j, \boldsymbol{S}_l) = h(\boldsymbol{S}_i, \boldsymbol{S}_j, \boldsymbol{S}_l) + h(\boldsymbol{S}_i, \boldsymbol{S}_l, \boldsymbol{S}_j) + h(\boldsymbol{S}_j, \boldsymbol{S}_i, \boldsymbol{S}_l) + h(\boldsymbol{S}_l, \boldsymbol{S}_i, \boldsymbol{S}_j) + h(\boldsymbol{S}_j, \boldsymbol{S}_l, \boldsymbol{S}_i) + h(\boldsymbol{S}_l, \boldsymbol{S}_j, \boldsymbol{S}_i),$$

where $S_i = (Q_i^T, R_i, J_i)^T$, $i, j, l = 1, 2, \ldots, n$. Thus,

$$U_{121}^{[1]}(\beta_0) = n^{-\frac{5}{2}} \sum_{i \neq j, l = i} h(S_i, S_j, S_l) + n^{-\frac{5}{2}} \sum_{i \neq j, l = j} h(S_i, S_j, S_l) + n^{-\frac{5}{2}} \sum_{i \neq j, l \neq i, l \neq l} h(S_i, S_j,$$
$$S_l) + n^{-\frac{5}{2}} \sum_{i \neq j, l = i} h(S_i, S_j, S_l) + n^{-\frac{5}{2}} \sum_{i \neq j, l = j} h(S_i, S_j, S_l) + n^{-\frac{5}{2}} \sum_{i < j < l} h(S_i, S_j, S_l).$$

(7.23)

Let us consider each of the three terms on the r.h.s. of (7.23). By the theory of U-statistics (van der Vaart, 2000), it can be shown easily that

$$n^{-\frac{5}{2}} \sum_{i \neq j, l = i} h(S_i, S_j, S_l) = O_p(n^{-1}) \text{ and } n^{-\frac{5}{2}} \sum_{i \neq j, l = j} h(S_i, S_j, S_l) = O_p(n^{-1}). \quad (7.24)$$

The third term on the r.h.s. of (7.23) is a U-statistic with symmetric kernel function $H(\cdot, \cdot, \cdot)$. Note that $E\{H(S_i, S_j, S_l)\} = 0$ and $E\{[R_l - r(W_l)]|W_l, \delta_l = 1\} = 0$. Then, by some manipulations, we can show that $E\{h(S_i, S_j, S_l)|S_i\} = E\{h(S_i, S_l, S_j)|S_i\} = E\{h(S_j, S_i, S_l)|S_i\} = E\{h(S_l, S_i, S_j)|S_i\} = 0$. Also, by standard non-parametric procedure, we can write

$$E\{h(S_j, S_l, S_i)|S_i\} = (R_i - r(W_i))\delta_i E\{K_h(W_j - W_i)\pi^{-2}(Q_j)g^{-1}(W_j) \times R_j \delta_j I(J_i = 2)(Z_i - Z_j)I\{r_l^{\beta_0} \geq r_j^{\beta_0}\}|S_i$$
$$\} = \left(\frac{R_i}{r(W_i)} - 1\right)\delta_i \rho(W_i)\varphi(W_i) + O_p(h^r).$$

Similarly,

$$E\{h(S_l, S_j, S_i)|S_i\} = \left(\frac{R_i}{r(W_i)} - 1\right)\delta_i \rho(W_i)\varphi(W_i) + O_p(h^r).$$

Therefore, the projection of the kernel function $H(S_i, S_j, S_l)$ is given by

$$E\{H(S_i, S_j, S_l)|S_i\} = 2\left(\frac{R_i}{r(W_i)} - 1\right)\delta_i \rho(W_i)\varphi(W_i) + O_p(h^r).$$

Thus, by the theory of U-statistics (van der Vaart, 2000, Chap.12),

$$n^{-\frac{5}{2}} \sum_{i < j < l} H(S_i, S_j, S_l) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(W_i)} - 1 \right) \delta_i \rho(W_i) \varphi(W_i) + O_p(\sqrt{n}h^r). \quad (7.25)$$

Combining (7.23), (7.24) and (7.25), it follows that

$$U_{121}^{[1]}(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(W_i)} - 1 \right) \delta_i \rho(W_i) \varphi(W_i) + O_p(\sqrt{n}h^r). \quad (7.26)$$

On the other hand, by some complex calculations as in Zhou, Wan and Wang (2008), we obtain

$$\| U_{121}^{[2]}(\beta_0) \| \leq C\sqrt{n}h^r + o_p(1). \quad (7.27)$$

Thus, by (7.22), (7.26), (7.27) and condition (C4),

$$U_{121}(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(W_i)} - 1 \right) \delta_i \rho(W_i) \varphi(W_i) + o_p(1), \quad (7.28)$$

and this proves (7.21). Further, by combining (7.19)-(7.21), we have

$$U_{12}(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(W_i)} - 1 \right) \delta_i \rho(W_i) \varphi(W_i) + o_p(1). \quad (7.29)$$

Analogous to the above derivation and by the theory of U-statistics, we can also obtain

$$U_{11}(\beta_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j) I\{r_j^{\beta_0} \geq r_i^{\beta_0}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(W_i)} - 1 \right) \delta_i I(J_i = 2)\varphi$$

$$(W_i).$$

$$(7.30)$$

Therefore, by (7.17), (7.18), (7.29) and (7.30), it follows that

$$
U_1(\boldsymbol{\beta}_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i I(J_i = 2)(\mathbf{Z}_i - \mathbf{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{R_i}{r(\mathbf{W}_i)} - 1 \right) \delta_i (I(J_i = 2)
$$

$$
- \rho(\mathbf{W}_i)) \varphi(\mathbf{W}_i) + o_p(1).
$$

(7.31)

Note that the first and second terms on the r.h.s. of (7.31) are uncorrelated. Hence

$$
U_1(\boldsymbol{\beta}_0) \xrightarrow{d} N(0, \sum\nolimits_1, \sum\nolimits_2),
$$

by the Central Limit Theorem. The proof of Theorem 1 can be completed by the Taylor series expansion. We omit the details here for brevity.

Proof of Theorem 2. We divide the proof into two parts.

Part B1. Note that

$$
\frac{1}{\sqrt{n}} \nabla U_2(\boldsymbol{\beta}_0) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ [R_i I(J_i = 2) + (1 - R_i)\rho(\mathbf{W}_i)] \delta_i (\mathbf{Z}_i - \mathbf{Z}_j)^{\otimes 2} s \left(\frac{r_j^{\boldsymbol{\beta}_0} - r_i^{\boldsymbol{\beta}_0}}{\sigma_n} \right) \frac{1}{\sigma_n} \right\} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (1
$$

$$
- R_i)(\hat{\rho}(\mathbf{W}_i) - \rho(\mathbf{W}_i)) \delta_i (\mathbf{Z}_i - \mathbf{Z}_j)^{\otimes 2} s \left(\frac{r_j^{\boldsymbol{\beta}_0} - r_i^{\boldsymbol{\beta}_0}}{\sigma_n} \right) \frac{1}{\sigma_n}.
$$

By some tedious calculations and the fact that $\sup\limits_{\boldsymbol{w}} |\hat{\rho}(\boldsymbol{w}) - \rho(\boldsymbol{w})| = O_p\left(h^r + (nh^d)^{-\frac{1}{2}} \right)$, it follows that

$$
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - R_i)(\hat{\rho}(\mathbf{W}_i) - \rho(\mathbf{W}_i)) \delta_i (\mathbf{Z}_i - \mathbf{Z}_j)^{\otimes 2} s \left(\frac{r_j^{\boldsymbol{\beta}_0} - r_i^{\boldsymbol{\beta}_0}}{\sigma_n} \right) \frac{1}{\sigma_n} = o_p(h^r + (nh^d)^{-\frac{1}{2}}).
$$

Thus, recognising that $E[R_i I(J_i = 2) + (1 - R_i)\rho(\mathbf{W}_i) = E[I(J_i = 2)]$ and by derivations similar to those used in the proof of Theorem 1, we obtain

$$\frac{1}{\sqrt{n}} \nabla U_2(\boldsymbol{\beta}_0) \xrightarrow{a.s} A = E\left\{ (\mathbf{Z}_1 - \mathbf{Z}_2)^{\otimes 2} \left[\int_{-\infty}^{\infty} \bar{H}(u) \bar{F}_{01}(u) f_{02}(u) \xi(u) du \right] \right\}.$$

Part B2. First, note that

$$U_2(\boldsymbol{\beta}_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} [R_i I(J_i = 2) + (1 - R_i) \rho(\mathbf{W}_i)] \delta_i(\mathbf{Z}_i - \mathbf{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n}$$

$$(1 - R_i)(\hat{\rho}(\mathbf{W}_i) - \rho(\mathbf{W}_i)) \delta_i(\mathbf{Z}_i - \mathbf{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + o_p(1) \equiv U_{21}(\boldsymbol{\beta}_0) + U_{22}(\boldsymbol{\beta}_0) + o_p(1).$$

$$(7.32)$$

Now, we can write

$$U_{22}(\boldsymbol{\beta}_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - R_i)(\hat{\rho}(\mathbf{W}_i) - \rho(\mathbf{W}_i)) \widehat{M}_n(\mathbf{W}_i) \times m^{-1}(\mathbf{W}_i) \delta_i(\mathbf{Z}_i - \mathbf{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

$$- n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - R_i)(\hat{\rho}(\mathbf{W}_i) - \rho(\mathbf{W}_i))(\widehat{M}_n(\mathbf{W}_i) - m(\mathbf{W}_i)) \times m^{-1}(\mathbf{W}_i) \delta_i(\mathbf{Z}_i - \mathbf{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

$$\} \equiv U_{22}^{[1]}(\boldsymbol{\beta}_0) + U_{22}^{[2]}(\boldsymbol{\beta}_0),$$

$$(7.33)$$

where $m(\boldsymbol{w}) = \pi(\boldsymbol{w}) g(\boldsymbol{w})$. We can easily show by steps similar to those for proving Theorem 1 that $U_{22}^{[2]}(\boldsymbol{\beta}_0) = o_p(1)$. Also, by the definition of $\hat{M}_n(\mathbf{W}_i)$, we have

$$U_{22}^{[1]}(\boldsymbol{\beta}_0) = n^{-\frac{5}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (1 - R_i)(I(J_l = 1) - \rho(W_i))R_l\delta_l K_h(W_i - W_l) \times m^{-1}(W_i)\delta_i(Z_i$$

$$- Z_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} = n^{-\frac{5}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (1 - R_i)(I(J_l = 1) - \rho(W_l))R_l\delta_l K_h(W_i - W_l) \times m^{-1}$$

$$(W_i)\delta_i(Z_i - Z_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + n^{-\frac{5}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (1 - R_i)(\rho(W_l) - \rho(W_i))R_l\delta_l K_h(W_i - W_l)$$

$$\times m^{-1}(W_i)\delta_i(Z_i - Z_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} \equiv I_1 + I_2.$$

(7.34)

By arguments similar to those used for proving Theorem 1, we have

$$I_2 = o_p(1). \quad (7.35)$$

and

$$I_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I(J_i = 2) - \rho(W_i))R_i\delta_i(1 - r(W_i))r^{-1}(W_i)\boldsymbol{\varphi}(W_i) + O_p(\sqrt{n}h^r). \quad (7.36)$$

Using (7.32)-(7.36) together, and by the theory of U-statistics, we obtain

$$U_2(\boldsymbol{\beta}_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} I(J_i = 2)\delta_i(Z_i - Z_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} - n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - R_i)(I(J_i = 2) - \rho(W_i))\delta_i(Z_i$$

$$- Z_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I(J_i = 2) - \rho(W_i))R_i\delta_i(1 - r(W_i))r^{-1}(W_i)\boldsymbol{\varphi}(W_i) + o_p(1) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} I(J_i$$

$$= 2)\delta_i(Z_i - Z_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I(J_i = 2) - \rho(W_i))\left(\frac{R_i}{r(W_i)} - 1\right)\delta_i\boldsymbol{\varphi}(W_i) + o_p(1).$$

The proof of Theorem 2 may be completed by using steps analogous to those used for proving Theorem 1.

Proof of Theorem 3. We divide the proof into two parts.

Part C1. Write

$$\frac{1}{\sqrt{n}} \nabla U_3(\boldsymbol{\beta}_0) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \left[\frac{R_i}{\pi(\boldsymbol{Q}_i)} I(J_i = 2) + (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)}) \rho(\boldsymbol{W}_i) \right] (\boldsymbol{Z}_i - \boldsymbol{Z}_j) s \left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n} \right) \frac{1}{\sigma_n}$$

$$+ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \left[\frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)} - \frac{R_i}{\pi(\boldsymbol{Q}_i)} \right] (\boldsymbol{Z}_i - \boldsymbol{Z}_j) s \left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n} \right) \frac{1}{\sigma_n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)}) (\hat{\rho}(\boldsymbol{W}_i)$$

$$- \rho(\boldsymbol{W}_i)) (\boldsymbol{Z}_i - \boldsymbol{Z}_j) s \left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n} \right) \frac{1}{\sigma_n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\hat{\pi}(\boldsymbol{Q}_i) - \pi(\boldsymbol{Q}_i)}{\hat{\pi}(\boldsymbol{Q}_i) \hat{\pi}(\boldsymbol{Q}_i)} (\hat{\rho}(\boldsymbol{W}_i) - \rho(\boldsymbol{W}_i)) (\boldsymbol{Z}_i - \boldsymbol{Z}_j) \delta_i R_i s$$

$$\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n} \right) \frac{1}{\sigma_n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\frac{R_i}{\pi(\boldsymbol{Q}_i)} - \frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)}) \delta_i \rho(\boldsymbol{W}_i) (\boldsymbol{Z}_i - \boldsymbol{Z}_j) s \left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n} \right) \frac{1}{\sigma_n}.$$

(7.37)

By steps analogous to those used for proving Theorems 1 and 2, we can show that the last four items of the r.h.s. of (7.37) are $o_p(1)$. Furthermore, noting that $E \left[\frac{R_i}{\pi(\boldsymbol{Q}_i)} I(J_i = 2) + (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)}) \rho(\boldsymbol{W}_i) \right] = E[I(J_i = 2)]$, we have

$$\frac{1}{\sqrt{n}} \nabla U_3(\boldsymbol{\beta}_0) \xrightarrow{a.s.} A = E \left\{ (\boldsymbol{Z}_1 - \boldsymbol{Z}_2)^{\otimes 2} \left[\int_{-\infty}^{\infty} \bar{H}(u) \bar{F}_{01}(u) f_{02}(u) \xi(u) du \right] \right\}.$$

Part C2. Note that

$$U_3(\boldsymbol{\beta}_0) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i [\frac{R_i}{\pi(\boldsymbol{Q}_i)} I(J_i = 2) + (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)}) \rho(\boldsymbol{W}_i)] (\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}$$

$$+ n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i [\frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)} - \frac{R_i}{\pi(\boldsymbol{Q}_i)}] I(J_i = 2) (\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i [(1$$

$$- \frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)}) \hat{\rho}(\boldsymbol{W}_i) - (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)}) \rho(\boldsymbol{W}_i)] (\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} \equiv U_{31}(\boldsymbol{\beta}_0) + U_{32}(\boldsymbol{\beta}_0) + U_{33}(\boldsymbol{\beta}_0).$$

(7.38)

Similar to the proof of Theorem 1, we have

$$U_{31}(\boldsymbol{\beta}_0) = - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\frac{R_i}{r(\boldsymbol{W}_i)} - 1) \delta_i \rho(\boldsymbol{W}_i) \boldsymbol{\varphi}(\boldsymbol{W}_i) + o_p(1). \quad (7.39)$$

Also, note that

$$
U_{33}(\boldsymbol{\beta}_0) = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)})(\hat{\rho}(\boldsymbol{W}_i) - \rho(\boldsymbol{W}_i))(\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\}
$$

$$
+ n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\hat{\pi}(\boldsymbol{Q}_i) - \pi(\boldsymbol{Q}_i)}{\hat{\pi}(\boldsymbol{Q}_i)\hat{\pi}(\boldsymbol{Q}_i)}(\hat{\rho}(\boldsymbol{W}_i) - \rho(\boldsymbol{W}_i))(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\delta_i R_i I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n}
$$

$$
(\frac{R_i}{\pi(\boldsymbol{Q}_i)} - \frac{R_i}{\hat{\pi}(\boldsymbol{Q}_i)})\delta_i \rho(\boldsymbol{W}_i)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} \equiv U_{33}^{[1]}(\boldsymbol{\beta}_0) + U_{33}^{[2]}(\boldsymbol{\beta}_0) + U_{33}^{[3]}(\boldsymbol{\beta}_0).
$$

(7.40)

It is clear that $U_{33}^{[2]}(\boldsymbol{\beta}_0) = o_p(1)$, and similar to the proof of Theorem 2, it follows that

$$
U_{33}^{[1]}(\boldsymbol{\beta}_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (1 - \frac{R_i}{\pi(\boldsymbol{Q}_i)})(I(J_i = 2) - \rho(\boldsymbol{W}_i))R_l \delta_l \times K_h(\boldsymbol{W}_i - \boldsymbol{W}_l)m^{-1}(\boldsymbol{W}_i)\delta_i
$$

$$
(\boldsymbol{Z}_i - \boldsymbol{Z}_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + o_p(1) = O_p(\sqrt{n}h^r) + o_p(1) = o_p(1).
$$

(7.41)

Moreover, by arguments similar to those used for the proof of Theorem 1, we can write

$$
U_{33}^{[3]}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\frac{R_i}{r(\boldsymbol{W}_i)} - 1)\delta_i \rho(\boldsymbol{W}_i)\boldsymbol{\varphi}(\boldsymbol{W}_i) + o_p(1). \quad (7.42)
$$

Thus, using (7.40)-(7.42) together,

$$
U_{33}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\frac{R_i}{r(\boldsymbol{W}_i)} - 1)\delta_i \rho(\boldsymbol{W}_i)\boldsymbol{\varphi}(\boldsymbol{W}_i) + o_p(1). \quad (7.43)
$$

Finally, by combining (7.38), (7.39) and (7.43) and the theory of U-statistics, we obtain

$$U_3(\boldsymbol{\beta}_0) = n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} I(J_i = 2)\delta_i(\mathbf{Z}_i - \mathbf{Z}_j)I\{r_j^{\boldsymbol{\beta}_0} \geq r_i^{\boldsymbol{\beta}_0}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I(J_i = 2) - \rho(\mathbf{W}_i))\left(\frac{R_i}{r(\mathbf{W}_i)} - 1\right)\delta_i\boldsymbol{\varphi}(\mathbf{W}_i)$$

$$+ o_p(1).$$

The proof of Theorem 3 may be completed by using arguments analogous to those used for proving Theorem 1.

References

1. Aerts M, Claeskens G, Hens N, Molenberghs G. 2002; Local multiple imputation. Biometrika. 89:375–388.

2. Bakoyannis G, Siannis F, Touloumi G. 2010; Modelling competing risks data with missing cause of failure. Statist Med. 29:3172–3185.

3. Brown BM, Wang YG. 2005; Standard errors and covariance matrices for smoothed rank estimators. Biometrika. 92:149–158.

4. Brown BM, Wang YG. 2007; Induced smoothing for rank regression with censored survival times. Statist Med. 26:828–836.

5. Buchbinder SP, Mehrotra DV, Duerr A, Fitzgerald DW, Mogg R, Li D, Gilbert PB, Lama JR, Marmor M, Del Rio C, McElrath MJ, Casimiro DR, Gottesdiener KM, Chodakewitz JA, Corey L, Robertson MN. 2008; Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. Lancet. 372(9653): 1881–1893. [PubMed: 19012954]

6. Buckley J, James I. 1979; Linear regression with censored data. Biometrika. 66:429–436.

7. Chen MH, Ibrahim JG, Shao QM. 2004; Propriety of the posterior distribution and existence of the MLE for regression models with covariates missing at random. J Amer Statist Assoc. 99:421–438.

8. Cheng SC, Fine JP, Wei LJ. 1998; Prediction of cumulative incidence function under the proportional hazards model. Biometrics. 54:219–228. [PubMed: 9544517]

9. Chiou SH, Kang S, Yan J. 2014; Fast accelerated failure time modeling for case-cohort data. Statist Comput. 24:559–568.

10. Duerr A, Huang Y, Buchbinder S, Coombs RW, Sanchez J, del Rio C, Casapia M, Santiago S, Gilbert PB, Corey L, Robertson MN. 2012; Extended follow-up confirms early vaccine-enhanced risk of HIV acquisition and demonstrates waning effect over time among participants in a randomized trial of recombinant adenovirus HIV vaccine (Step Study). J Infect Dis. 206:258–266. [PubMed: 22561365]

11. Fine JP, Gray RJ. 1999; A proportional hazards model for the subdistri-bution of a competing risk. J Amer Statist Assoc. 94:496–509.

12. Fleming, TR, Harrington, DP. Counting Processes and Survival Analysis. New York: Wiley; 1991.

13. Fygenson M, Ritov Y. 1994; Monotone estimating equations for censored data. Ann Statist. 22:732–746.

14. 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.

15. Goetghebeur E, Ryan L. 1995; Analysis of competing risks survival data when some failure types are missing. Biometrika. 82:821–834.

16. Han P. 2014; Multiply robust estimation in regression analysis with missing data. J Amer Statist Assoc. 109:26–41.

17. Heller G. 2007; Smoothed rank regression with censored data. J Amer Statist Assoc. 102:552–559.

18. Hertz T, Ahmed H, Friedrich DP, Casimiro DR, Self SG, Corey L, McElrath MJ, Buchbinder S, Horton H, Frahm N, Robertson MN, Graham BS, Gilbert PB. 2013; HIV-1 Vaccine-induced T-Cell

reponse cluster in epitope hotspots that differ from those induced in natural infection with HIV-1. PLoS Patho. 9:e1003404.

19. Horvitz DG, Thompson DJ. 1952; A generalization of sampling without replacement from a finite universe. J Amer Statist Assoc. 47:663–685.

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

21. Jin Z, Lin DY, Wei LJ, Ying Z. 2003; Rank-based inference for the accelerated failure time model. Biometrika. 90:341–353.

22. Johnson LM, Strawderman RL. 2009; Induced smoothing for the semi-parametric accelerated failure time model: asymptotics and extensions to clustered data. Biometrika. 96:577–590. [PubMed: 23049117]

23. Kalbfleisch, JD, Prentice, RL. The Statistical Analysis of Failure Time Data. 2nd. New York: Wiley; 2002.

24. Kanwal, RP. Generalized Functions: Theory and Thehnique. Boston: Birkhäuser; 1998.

25. Lee S, Lewbel A. 2013; Nonparametric identification of accelerated failure time competing risks models. Economet Theor. 29:905–919.

26. Li K. 1991; Sliced inverse regression for dimension reduction. J Amer Statist Assoc. 86:316–327.

27. Lin H, Peng H. 2013; Smoothed rank correlation of the linear transformation regression model. Comput Statist Data An. 19:1370–1402.

28. Lu G, Copas JB. 2005; Missing at random, likelihood ignorability and model completeness. Ann Statist. 32:754–765.

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

30. Lu W, Liang Y. 2008; Analysis of competing risks data with missng cause of failure under additive hazards model. Statist Sinica. 18:219–234.

31. Ma S, Huang J. 2007; Combining multiple markers for classification using ROC. Biometrics. 63:751–757. [PubMed: 17825007]

32. Nadaraya EA. 1964; On estimating regression. Theor Probab Appl. 9:141–142.

33. Pang L, Lu W, Wang H. 2012; Variance estimation in censored quantile regression via induced smoothing. Comput Statist Data An. 56:785–796.

34. Prentice RL. 1978; Linear rank tests with right censored data. Biometrika. 65:167–180.

35. Prentice RL, Kalbfleisch JD. 1978; The analysis of failure times in the presence of competing risks. Biometrics. 34:541–554. [PubMed: 373811]

36. Qi L, Wang YC, Prentice RL. 2005; Weighted estimators for proportional hazards regression with missing covariates. J Amer Statist Assoc. 100:1250–1263.

37. Qiu Z, Qin J, Zhou Y. 2016; Composite Estimating Equation Method for the Accelerated Failure Time Model with Length-biased Sampling Data. Scand J Statist. 43:396–415.

38. Robins JM, Rotnitzky A, Zhao LP. 1994; Estimation of regression coefficients when some regressors are not always observed. J Amer Statist Assoc. 89:846–866.

39. Rolland M, Tovanabutra S, deCamp AC, Frahm N, Gilbert PB, Sanders-Buell E, Heath L, Magaret CA, Bose M, Bradfield A, O'Sullivan A, Crossler J, Jones T, Nau M, Wong K, Zhao H, Raugi DN, Sorensen S, Stoddard JN, Maust BS, Deng W, Hural J, Dubey S, Michael NL, Shiver J, Corey L, Li F, Self SG, Kim J, Buchbinder S, Casimiro DR, Robertson MN, Duerr A, McElrath MJ, McCutchan FE, Mullins JI. 2011; Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. Nat Med. 17:366–371. [PubMed: 21358627]

40. Rubin DB. 1976; Inference and missing data. Biometrika. 63:581–592.

41. Scheike TH, Zhang MJ. 2003; Extensions and applications of the Cox-Aalen survival model. Biometrics. 59:1036–1045. [PubMed: 14969483]

42. Shen Y, Cheng SC. 1999; Confidence bands for cumulative incidence curves under the additive risk model. Biometrics. 55:1093–1100. [PubMed: 11315053]

43. Song X, Ma S, Huang J, Zhou XH. 2007; A semiparametric approach for the nonparametric transformation survival model with multiple covariates. Bio-statistics. 8:197–211.

44. Song XY, Sun LQ, Mu XY, Dinse GE. 2010; Additive hazards regression with censoring indicators missing at random. Can J Statist. 38:333–351.

45. Sun LQ, Liu J, Zhang M, Sun J. 2006; Modeling the subdistribution of a competing risk. Statist Sinica. 16:1367–1385.

46. Sun YQ, Wang HJ, Gilbert PB. 2012; Quantile regression for competing risks data with missing cause of failure. Statist Sinica. 22:703–728.

47. Tsiatis AA. 1990; Estimating regression parameters using linear rank tests for censored data. Ann Statist. 18:354–372.

48. van der Vaart, AW. Asymptotic Statistics. Cambridge, UK: Cambridge University Press; 2000.

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

50. Wang Q, Rao JNK. 2002; Empirical likelihood-based inference under imputation for missing response data. Ann Statist. 30:896–924.

51. Wang S, Wang CY. 2001; A note on kernel-assisted estimators in missing covariate regression. Statist Probabil Lett. 55:439–449.

52. Wang YG, Fu L. 2011; Rank regression for the accelerated failure time model with clustered and censored data. Comput Statist Data An. 55:2334–2343.

53. Watson GS. 1964; Smooth regression analysis. Sankhy (series A). 26:359–372.

54. Ying Z. 1993; A large sample study of rank estimation for censored regression data. Ann Statist. 21:76–99.

55. Zhao YD, Brown BM, Wang YG. 2014; Smoothed rank-based procedure for censored data. Electron J Statist. 8:2953–2974.

56. Zheng M, Lin R, Yu W. 2016; Competing risks data analysis under the accelerated failure time model with missing cause of failure. Ann Inst Statist Math. 68:855–876.

57. Zhou Y, Wan ATK, Wang X. 2008; Estimating equations inference with missing data. J Amer Statist Assoc. 103:1187–1199.

Table 1

Simulation results of Experiment 1

		$Z_i \sim$ Bernoulli(0.5)				$Z_i \sim U[0, 1]$			
		BIAS	**SE**	**SD**	**CP**	**BIAS**	**SE**	**SD**	**CP**
Scenario 1									
	FULL	-0.008	0.094	0.091	93.5%	0.001	0.112	0.111	95.1%
	CC	0.043	0.110	0.104	92.0%	0.061	0.130	0.129	92.2%
$e_i \sim N(0,0.25)$	IPW	-0.000	0.104	0.097	93.7%	0.016	0.125	0.119	93.4%
	EEI	-0.002	0.102	0.097	94.4%	0.017	0.118	0.119	94.9%
	AIPW	-0.006	0.104	0.097	93.9%	0.004	0.125	0.119	94.0%
	FULL	-0.001	0.055	0.055	94.9%	-0.001	0.070	0.067	94.2%
	CC	0.017	0.069	0.065	92.9%	0.020	0.086	0.081	91.9%
$e_i \sim U[-0.5,0.5]$	IPW	0.001	0.063	0.062	94.7%	0.003	0.080	0.077	94.1%
	EEI	0.012	0.064	0.062	94.1%	0.013	0.078	0.078	94.2%
	AIPW	0.002	0.065	0.062	93.8%	0.001	0.081	0.076	93.1%
	FULL	-0.005	0.132	0.129	94.0%	0.000	0.159	0.156	95.1%
	CC	0.097	0.143	0.139	88.1%	0.110	0.176	0.171	88.7%
$e_i \sim GEV(0,0.5,0)$	IPW	0.012	0.141	0.133	93.0%	0.021	0.174	0.166	93.8%
	EEI	-0.010	0.138	0.134	93.3%	0.004	0.166	0.166	95.6%
	AIPW	-0.009	0.143	0.135	92.2%	-0.004	0.175	0.166	94.0%
Scenario 2									
	FULL	-0.002	0.090	0.091	95.2%	0.005	0.113	0.111	94.4%
	CC	0.069	0.129	0.127	90.6%	0.083	0.154	0.154	91.0%
$e_i \sim N(0,0.25)$	IPW	-0.004	0.111	0.106	95.0%	0.002	0.137	0.131	94.2%
	EEI	-0.000	0.102	0.107	95.7%	0.022	0.122	0.131	96.1%
	AIPW	-0.004	0.105	0.106	95.4%	0.002	0.134	0.131	94.6%
	FULL	-0.000	0.057	0.055	94.5%	0.003	0.069	0.067	95.1%
$e_i \sim U[-0.5,0.5]$	CC	0.034	0.081	0.076	91.2%	0.042	0.096	0.093	92.3%

		$Z_i \sim$ Bernoulli(0.5)				$Z_i \sim U[0, 1]$			
		BIAS	SE	SD	CP	BIAS	SE	SD	CP
Scenario 1									
	IPW	-0.001	0.072	0.070	93.9%	0.003	0.089	0.089	95.1%
	EEI	0.014	0.073	0.071	93.7%	0.020	0.082	0.088	96.1%
	AIPW	-0.002	0.074	0.070	93.2%	0.002	0.088	0.085	94.8%
	FULL	-0.001	0.132	0.128	94.7%	-0.004	0.158	0.156	94.8%
	CC	0.118	0.190	0.181	88.7%	0.127	0.225	0.218	91.0%
$\varepsilon_i \sim$ GEV(0,0.5,0)	IPW	-0.001	0.160	0.146	92.5%	-0.000	0.194	0.181	94.1%
	EEI	-0.013	0.146	0.145	95.0%	0.011	0.168	0.176	95.8%
	AIPW	-0.004	0.152	0.146	94.4%	-0.002	0.182	0.178	94.6%

Table 2

Simulation results of Experiment 2

		$\beta_{b1}=1$				$\beta_{b2}=1$			
		BIAS	SE	SD	CP	BIAS	SE	SD	CP
Scenario 1									
	FULL	-0.004	0.116	0.111	94.0%	-0.001	0.066	0.064	94.2%
	CC	0.039	0.141	0.133	92.0%	-0.004	0.080	0.075	93.7%
$\varepsilon_i \sim$ N(0,0.25)	IPW	-0.032	0.139	0.129	92.0%	-0.001	0.076	0.073	94.8%
	EEI	-0.008	0.125	0.126	95.7%	-0.004	0.070	0.072	95.7%
	AIPW	-0.009	0.137	0.126	93.2%	-0.005	0.073	0.073	95.1%
	FULL	0.003	0.068	0.068	95.2%	-0.001	0.038	0.039	95.1%
	CC	0.018	0.080	0.078	93.7%	-0.008	0.045	0.044	94.1%
$\varepsilon_i \sim$ U[-0.5,0.5]	IPW	-0.024	0.081	0.084	95.0%	-0.010	0.044	0.046	95.4%
	EEI	0.012	0.078	0.085	96.8%	0.005	0.045	0.047	96.6%
	AIPW	-0.006	0.088	0.082	95.3%	-0.011	0.048	0.047	96.0%
	FULL	0.005	0.161	0.158	94.4%	0.003	0.091	0.090	95.2%
	CC	0.097	0.202	0.202	92.5%	0.002	0.114	0.113	95.0%
$\varepsilon_i \sim$ GEV(0,0.5,0)	IPW	-0.007	0.191	0.191	95.1%	0.022	0.104	0.106	94.8%
	EEI	-0.002	0.171	0.177	96.0%	-0.001	0.096	0.100	96.1%
	AIPW	0.008	0.189	0.177	94.2%	0.010	0.101	0.102	95.5%
Scenario 2									
	FULL	0.002	0.111	0.112	95.2%	-0.000	0.066	0.065	95.2%
	CC	0.061	0.158	0.155	92.9%	0.072	0.092	0.090	86.2%
$\varepsilon_i \sim$ N(0,0.25)	IPW	-0.002	0.141	0.134	94.0%	0.002	0.081	0.076	93.3%
	EEI	-0.003	0.124	0.133	96.9%	0.005	0.075	0.076	95.7%
	AIPW	-0.002	0.138	0.133	94.7%	-0.000	0.080	0.076	94.0%
	FULL	0.002	0.069	0.068	93.3%	-0.002	0.037	0.039	96.4%
$\varepsilon_i \sim$ U[-0.5,0.5]	CC	0.041	0.098	0.094	90.8%	0.030	0.054	0.054	90.7%

		$\beta_{01}=1$				$\beta_{02}=1$			
		BIAS	SE	SD	CP	BIAS	SE	SD	CP
Scenario 1									
	IPW	0.003	0.088	0.088	94.1%	-0.004	0.048	0.049	95.6%
	EEI	0.006	0.083	0.089	96.1%	0.011	0.048	0.050	96.1%
	AIPW	0.002	0.094	0.087	93.4%	-0.005	0.051	0.050	94.7%
	FULL	0.001	0.164	0.158	94.1%	0.006	0.093	0.090	94.1%
	CC	0.104	0.236	0.225	90.8%	0.128	0.137	0.128	83.0%
$\varepsilon_i \sim$ GEV(0,0.5,0)	IPW	0.002	0.211	0.186	91.8%	0.009	0.118	0.104	91.5%
	EEI	-0.002	0.177	0.183	95.6%	-0.003	0.107	0.103	94.2%
	AIPW	0.005	0.183	0.117	94.0%	0.006	0.113	0.104	92.8%
Scenario 3									
	FULL	0.002	0.109	0.112	95.8%	-0.001	0.065	0.064	95.7%
	CC	0.060	0.139	0.142	93.0%	-0.018	0.085	0.083	94.2%
$\varepsilon_i \sim$ N(0,0.25)	IPW	0.035	0.137	0.134	93.7%	-0.007	0.080	0.075	93.4%
	EEI	0.003	0.120	0.132	97.4%	-0.001	0.076	0.075	94.7%
	AIPW	0.003	0.130	0.131	95.1%	-0.004	0.078	0.075	94.2%
	FULL	0.002	0.068	0.068	95.4%	-0.001	0.040	0.039	94.9%
	CC	0.033	0.089	0.087	92.1%	-0.009	0.053	0.051	93.8%
$\varepsilon_i \sim$ U[-0.5,0.5]	IPW	0.024	0.090	0.090	94.7%	-0.005	0.049	0.050	96.2%
	EEI	-0.000	0.085	0.090	96.9%	0.021	0.049	0.050	93.1%
	AIPW	-0.003	0.097	0.086	93.4%	-0.002	0.052	0.049	94.7%
	FULL	0.007	0.159	0.159	94.6%	0.001	0.090	0.090	95.0%
	CC	0.103	0.210	0.200	90.9%	-0.024	0.124	0.117	93.5%
$\varepsilon_i \sim$ GEV(0,0.5,0)	IPW	0.053	0.206	0.186	91.0%	-0.008	0.113	0.105	93.4%
	EEI	0.012	0.174	0.181	95.6%	-0.022	0.103	0.104	95.0%
	AIPW	0.012	0.196	0.183	93.6%	-0.006	0.107	0.105	94.7%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Simulation results of Experiment 3

METHOD	BIAS	SE	SD	CP	computational time
Smoothed	0.001	0.058	0.058	94.7%	314.051s
Non-smoothed	0.020	0.067	0.068	93.9%	21166.368s

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Simulation results of Experiment 4

σ_n		$\beta_{01} = 1$				$\beta_{02} = 1$			
		BIAS	SE	SD	CP	BIAS	SE	SD	CP
	FULL	-0.001	0.111	0.112	94.7%	-0.002	0.064	0.064	94.2%
	CC	0.041	0.136	0.133	94.0%	-0.003	0.076	0.075	93.8%
$0.1 * n^{-0.26}$	IPW	-0.026	0.133	0.134	94.5%	0.001	0.074	0.075	95.1%
	EEI	-0.003	0.117	0.128	96.2%	-0.002	0.070	0.072	95.4%
	AIPW	-0.006	0.132	0.127	94.5%	-0.004	0.075	0.074	95.4%
	FULL	-0.001	0.112	0.112	94.6%	-0.003	0.066	0.064	94.0%
	CC	0.040	0.137	0.133	93.3%	-0.004	0.079	0.075	94.1%
$0.3 * n^{-0.26}$	IPW	-0.025	0.133	0.133	94.0%	-0.001	0.077	0.075	94.4%
	EEI	-0.002	0.121	0.131	96.0%	-0.003	0.072	0.074	95.3%
	AIPW	-0.006	0.138	0.129	93.4%	-0.007	0.080	0.073	93.8%
	FULL	-0.004	0.116	0.111	94.0%	-0.001	0.066	0.064	94.2%
	CC	0.039	0.141	0.133	92.0%	-0.004	0.080	0.075	93.7%
$0.5 * n^{-0.26}$	IPW	-0.032	0.139	0.129	92.0%	-0.001	0.076	0.073	94.8%
	EEI	-0.008	0.125	0.126	95.7%	-0.004	0.070	0.072	95.7%
	AIPW	-0.009	0.137	0.126	93.2%	-0.005	0.073	0.073	95.1%
	FULL	0.002	0.111	0.113	95.6%	0.002	0.064	0.064	95.4%
	CC	0.050	0.130	0.134	94.2%	0.002	0.076	0.075	94.1%
$0.7 * n^{-0.26}$	IPW	-0.019	0.132	0.137	95.4%	0.006	0.075	0.076	94.8%
	EEI	0.004	0.121	0.134	97.1%	0.002	0.071	0.077	96.3%
	AIPW	0.002	0.131	0.129	94.0%	0.001	0.074	0.076	95.4%
	FULL	0.006	0.120	0.114	94.4%	0.001	0.063	0.065	95.9%
	CC	0.054	0.141	0.136	93.0%	0.000	0.077	0.077	95.4%
$0.9 * n^{-0.26}$	IPW	-0.018	0.144	0.135	93.6%	0.003	0.074	0.075	95.8%
	EEI	0.005	0.130	0.131	96.0%	-0.001	0.067	0.075	96.9%
	AIPW	0.003	0.146	0.129	93.3%	-0.003	0.072	0.073	95.6%

Table 5

Estimation of the effects of WhiteRace, Age and Treatment for the HVTN 502 Step HIV vaccine efficacy trial data

Method	WhiteRace			Age			Treatment		
	EST	$\widehat{SE}$	P-value	EST	$\widehat{SE}$	P-value	EST	$\widehat{SE}$	P-value
CC	-0.460	0.403	0.254	0.040	0.013	0.002	-0.743	0.152	0.000
IPW	-0.501	0.402	0.213	0.039	0.013	0.002	-0.681	0.167	0.000
EEI	-0.321	0.428	0.453	0.022	0.013	0.084	-0.786	0.132	0.000
AIPW	-0.296	0.426	0.488	0.024	0.013	0.064	-0.772	0.132	0.000