QUANTILE REGRESSION FOR COMPETING RISKS DATA WITH MISSING CAUSE OF FAILURE

Yanqing Sun[1], Huixia Judy Wang[2] and Peter B. Gilbert[3,4]

[1]*University of North Carolina, Charlotte,* [2]*North Carolina State University,*
[3]*University of Washington and* [4]*Fred Hutchinson Cancer Research Center*

Abstract: This paper considers generalized linear quantile regression for competing risks data when the failure type may be missing. Two estimation procedures for the regression coefficients, including an inverse probability weighted complete-case estimator and an augmented inverse probability weighted estimator, are discussed under the assumption that the failure type is missing at random. The proposed estimation procedures utilize supplemental auxiliary variables for predicting the missing failure type and for informing its distribution. The asymptotic properties of the two estimators are derived and their asymptotic efficiencies are compared. We show that the augmented estimator is more efficient and possesses a double robustness property against misspecification of either the model for missingness or for the failure type. The asymptotic covariances are estimated using the local functional linearity of the estimating functions. The finite sample performance of the proposed estimation procedures are evaluated through a simulation study. The methods are applied to analyze the 'Mashi' trial data for investigating the effect of formula- versus breast-feeding plus extended infant zidovudine prophylaxis on HIV-related death of infants born to HIV-infected mothers in Botswana.

Key words and phrases: Augmented inverse probability weighted, auxiliary variables, competing risks, double robustness, efficient estimator, estimating equation, inverse probability weighted, local functional linearity, logistic regression, Mashi trial, missing at random, quantile regression.

1. Introduction

This paper was motivated by the need to analyze the 'Mashi' trial data (mashi means milk in Setswana) for examining the effect of formula- versus breast-feeding plus extended infant zidovudine prophylaxis on HIV-related death of infants born to HIV-infected mothers in Botswana (Thior et al. (2006)). Whereas studies including the Mashi trial have shown that formula-feeding increases the overall risk of death while breast-feeding increases the risk of transmitting HIV (Dunn et al. (1992); Beaudry, Dufour, and Marcoux (1995)), the effect of feeding strategy on death due to HIV infection is unknown. Accordingly, it is of interest to assess the treatment effect on HIV-related death, with

HIV-unrelated death considered as a competing risk. This analysis provides additional insight over the analysis of all-cause death by addressing whether the known beneficial effect of formula-feeding to prevent HIV infection leads to a beneficial effect to reduce mortality of HIV infected infants.

Of the 111 live-born infants who died in the Mashi trial, the cause of death is known for 50 and missing for 61. It is well-known that the analysis of only cases with complete information may lead to inefficient and/or biased estimates. To account for the missingness, a number of methods have been developed to estimate the covariate effects under different survival models for the cause-specific hazard functions, for instance the proportional hazards model (Goetghebeur and Ryan (1995), Lu and Tsiatis (2001)), the linear transformation model (Gao and Tsiatis (2005)) and the additive hazard model (Lu and Liang (2008)), among others.

In this paper, we consider a quantile regression model (Koenker and Bassett (1978)) that is a valuable complement to the Cox proportional hazards model (Cox (1972)) and the accelerated failure time model (Buckley and James (1979); Koul, Susarla, and Van Ryzin (1981)). Quantile regression allows the covariate effects to vary at different tails of the event time distribution. Such important heterogeneity in the population may be overlooked by using the Cox model or the accelerated failure time model. General quantile regression methods in survival analysis are developed under the assumption that censoring is independent of failure time; see Ying, Jung, and Wei (1995), Bang and Tsiatis (2002), Portnoy (2003), Neocleous, Vanden Branden, and Portnoy (2006), Peng and Huang (2008), Wang and Wang (2009), among others.

Most relevant to this work, Peng and Fine (2009) studied the quantile for the cumulative incidence function, which is the distribution of the time to failure due to a particular cause of interest. However, their method does not account for missingness of failure cause. Its application based on the complete-cases may be invalid and misleading because of the high percentage of missing causes of death in the Mashi data. We consider generalized linear quantile regression for competing risks data when causes of failure may be missing. Two estimation procedures are discussed under the assumption that the failure cause is missing at random (Rubin (1976)). The first, following the idea of Horvitz and Thompson (1952), uses inverse probability weighting (IPW) of complete-cases, which leverages auxiliary predictors of whether cause of failure is observed. The second approach, adapting the theory of Robins, Rotnitzky, and Zhao (1994), augments the IPW complete-case estimator with auxiliary predictors of the cause of failure of interest.

This work fits in the general area of competing risks failure time analysis, wherein subjects are followed over time and may fail from one of many causes.

The competing risks failure time can be represented by the minimum of the latent failure times, each of which is defined as the time to failure from a particular cause in the absence of all other competing risks. The existing quantile regression methods could be applied within this framework by considering quantile regression of the latent failure time for a particular cause while treating other latent failure times as censoring and by assuming mutual independence of the latent failure times (Tsiatis (1975)). This independence mutual assumption is untestable and is often dubious (because we expect positive correlation of the latent failure times), however, and Peng and Fine (2009), took a different approach that avoids this assumption. In particular, they studied the cumulative incidence function, which is the distribution of the time to failure due to a particular cause in the presence of the other competing risks. This approach evaluates "crude" effects on the cause-specific cumulative incidence, and hence caution is needed in the interpretation of the results (Prentice et al. (1978)). This is the dominant approach for assessing competing risks data given the fundamental non-identifiability of the latent failure times, and the methods developed here take this approach.

The rest of the paper is organized as follows. Two procedures for estimating the regression coefficients are proposed in Section 2. The asymptotic properties of these estimators are derived and their asymptotic efficiencies are compared in Section 3. Procedures for estimating the asymptotic covariances are given in Section 4. The finite-sample performance of the proposed estimation procedures are evaluated in Section 5 through a simulation study. The methods are applied to the Mashi data in Section 6 for investigating the effect of formula- versus breast-feeding plus extended infant zidovudine prophylaxis on HIV-related death of infants. All proofs are in Appendix.

2. Estimation Procedures

2.1. Model descriptions and assumptions

Let T be the survival time of interest. Due to censoring, we only observe (X, δ), where $X = \min(T, C)$, $\delta = I(T \leq C)$ and C is the censoring variable. Let J denote the failure type associated with the uncensored failure time T. The J is meaningless and undefined if T is censored. For convenience we let Z be the $(p+1)$-dimensional concomitant variable including 1 as its first component corresponding to an intercept. A typical right-censored competing risks data set consists of independent and identically distributed (*i.i.d.*) copies $(X_i, \delta_i, \delta_i J_i, Z_i)$, $i = 1, \ldots, n$, of $(X, \delta, \delta J, Z)$.

We consider $J = 1$ as the failure type of interest and set $J = 2$ for all other failure types. The type-1 cumulative incidence function is $F_1(t|Z) = P(T \leq t, J = 1|Z)$, which represents the conditional probability of observing a type-1 failure by time t given the covariate Z. The τth type-1 conditional quantile given

$Z = z$ is defined as $F_1^{-1}(\tau|z) = \inf\{t : F_1(t|z) \geq \tau\}$. Let ν be the end of follow-up time and satisfying the condition C1 given in the Appendix. For identifiability, we require that $\tau \leq \tau_0$ where $\tau_0 = \inf_z P(T \leq \nu, J = 1 \mid z)$. For $\tau \in [\tau_L, \tau_U]$ with $0 < \tau_L \leq \tau_U < \tau_0$, the τth generalized linear quantile regression is

$$F_1^{-1}(\tau|z) = g\{z^T \beta(\tau)\}, \tag{2.1}$$

where $g(\cdot)$ is a known monotone link function and $\beta(\tau)$ is a $(p+1)$-dimensional coefficient vector depending on τ.

To help with understanding model (2.1) and the interpretation of $\beta(\tau)$, we consider the following scenario. Suppose $Z_i = (1, Z_{i,1})$, where $Z_{i,1}$ is the indicator of gender for subject i, say $Z_{i,1} = 1$ for male and $Z_{i,1} = 0$ for female, and T_i is the time (age in years) to death. Suppose $\beta(\tau) = (\beta_0(\tau), \beta_1(\tau))$, with $\beta_0(\tau) = 70$ and $\beta_1(\tau) = -5$ at $\tau = 0.3$. Thus, conditional on gender, the age by which 30% of the population dies of type-1 failure is $70 - 5Z_{i,1}$ under the identity link function. That is, 30% of females die from type-1 failure before age 70 and 30% of males die from type-1 failure before age 65. The gender effect is 5 years – the age at which 30% of individuals die from type-1 failure is 5 years sooner for males than for females.

Let $G(t|Z_i) = P(C_i \geq t|Z_i)$ and let $\hat{G}(t|Z_i)$ be a semiparametric or nonparametric consistent estimator of $G(t|Z_i)$. Peng and Huang (2008) proposed the following estimating equation for $\beta(\tau)$ based on fully observed competing risks data $\{X_i, Z_i, \delta_i, \delta_i J_i, \ i = 1 \ldots, n\}$:

$$S_n(b, \tau) = \sum_{i=1}^n Z_i \left[\frac{I\{T_i \leq g(Z_i^T b), J_i = 1\}\delta_i}{\hat{G}(T_i|Z_i)} - \tau \right] = 0. \tag{2.2}$$

In this paper we consider the quantile regression (2.1) based on the competing risks data with possibly missing failure type. Let R_i be the complete-case indicator: $R_i = 1$ either if $\delta_i = 0$ or if $\delta_i = 1$ and J_i is observed; and $R_i = 0$ otherwise. Auxiliary variables A_i may be helpful for predicting the missing failure type. Since the failure type is defined only for those who are observed to fail, only supplemental information for the observed failures are potentially useful for predicting missingness and for informing about the distribution of the failure type. As such, we denote available auxiliaries by $\delta_i A_i$.

We assume the censoring time C_i is conditionally independent of (T_i, J_i) given Z_i. We also assume the failure type J_i is missing at random (Rubin (1976)); that is, given $\delta_i = 1$ and $W_i = (T_i, Z_i, A_i)$, the probability that the failure type J_i is missing depends only on the observed W_i, not on the value of J_i; this assumption is expressed as

MAR : $\qquad r(W_i) \equiv P(R_i = 1|J_i, \delta_i = 1, W_i) = P(R_i = 1|\delta_i = 1, W_i).$
$$\tag{2.3}$$

Let $\pi(Q_i) = P(R_i = 1|Q_i)$, where $Q_i = (W_i, \delta_i)$. Then

$$\pi(Q_i) = \delta_i r(W_i) + (1 - \delta_i). \tag{2.4}$$

The observed data can be summarized as $O_i = \{X_i, Z_i, \delta_i, R_i, R_i\delta_i J_i, \delta_i A_i\}, i = 1 \ldots, n$. We assume that O_i's are independent and identically distributed.

2.2. Inverse probability weighted estimator

First, following the idea of Horvitz and Thompson (1952), we propose a procedure for estimating $\beta(\tau)$ that uses inverse probability weighting (IPW) of complete-cases. We consider a parametric model $r(W_i, \psi)$ for $r(W_i) = P(R_i = 1|\delta_i = 1, W_i)$, where ψ is an unknown vector of finite-dimensional parameters. For example, $r(W_i, \psi)$ may be a logistic regression model with $\log[r(W_i, \psi)/\{1 - r(W_i, \psi)\}] = W_i^T\psi$. The parameter ψ can be estimated by $\hat{\psi}$, the maximizer of the observed-data likelihood

$$\prod_{i=1}^{n} \{r(W_i, \psi)\}^{R_i I(\delta_i=1)} \{1 - r(W_i, \psi)\}^{1-R_i}. \tag{2.5}$$

Therefore, we can estimate $\pi(Q_i, \psi)$ by $\hat{\pi}(Q_i) = \pi(Q_i, \hat{\psi}) = \delta_i \hat{r}(W_i) + (1 - \delta_i)$ where $\hat{r}(W_i) = r(W_i, \hat{\psi})$.

Modifying (2.2) to accommodate missing failure types leads to the IPW estimating equation for $\beta(\tau)$:

$$S_{1,n}(b, \tau) = n^{-1/2} \sum_{i=1}^{n} Z_i \left[\frac{R_i}{\pi(Q_i, \hat{\psi})} \frac{I\{g^{-1}(T_i) \leq Z_i^T b, J_i = 1\}\delta_i}{\hat{G}(T_i|Z_i)} - \tau \right] = 0. \tag{2.6}$$

We can write $S_{1,n}(b, \tau) = n^{-1/2} \sum_{i=1}^{n} Z_i [I\{g^{-1}(T_i) \leq Z_i^T b\}\hat{\vartheta}_{1,i} - \tau]$, where $\hat{\vartheta}_{1,i} = R_i\delta_i I(J_i = 1)/\{\hat{\pi}(Q_i)\hat{G}(T_i|Z_i)\}$. We refer to the solution of (2.6) as the IPW estimator, denoted by $\hat{\beta}_I(\tau)$.

2.3. Augmented inverse probability weighted estimator

Because the IPW estimator obtained by solving (2.6) uses data from complete cases only, it is inefficient, and it is asymptotically consistent only if the missingness probability $\pi(W_i, \psi)$ is correctly modeled. Adapting the theory of Robins, Rotnitzky, and Zhao (1994) to gain more efficiency and robustness against the misspecification of $\pi(W_i, \psi)$, we propose an improved estimation procedure that augments the IPW complete-case estimator with auxiliary predictors of the failure type of interest.

Let $\rho(W_i) = P(J_i = 1|\delta_i = 1, W_i)$. The missing at random assumption (2.3) implies that J_i is independent of R_i given Q_i:

$$\rho(W_i) = P(J_i = 1|R_i = 1, \delta_i = 1, W_i). \tag{2.7}$$

Let $\rho(W_i, \phi)$ be a parametric model for $\rho(W_i)$, where ϕ is a vector of unknown parameters. From (2.7), it follows that $\rho(W_i)$ can be estimated from the complete cases with $R_i = 1$ and $\delta_i = 1$. The maximum likelihood estimator of ϕ, $\hat{\phi}$, can be obtained by maximizing the likelihood

$$\prod_{i=1}^{n} \{\rho(W_i, \phi)\}^{R_i \delta_i I(J_i=1)} \{1 - \rho(W_i, \phi)\}^{R_i \delta_i I(J_i \neq 1)}. \tag{2.8}$$

Denote $\rho(W_i, \hat{\phi})$ by $\hat{\rho}(W_i)$. We consider the augmented IPW estimating equation for $\beta(\tau)$:

$$S_{2,n}(b, \tau) = 0, \tag{2.9}$$

where

$$S_{2,n}(b, \tau) = n^{-1/2} \sum_{i=1}^{n} Z_i \left(\left[\frac{R_i}{\hat{\pi}(Q_i)} \frac{I\{g^{-1}(T_i) \leq Z_i^T b, J_i = 1\}\delta_i}{\hat{G}(T_i | Z_i)} \right. \right.$$
$$\left. \left. + \left\{ 1 - \frac{R_i}{\hat{\pi}(Q_i)} \right\} \frac{I\{g^{-1}(T_i) \leq Z_i^T b\}\delta_i}{\hat{G}(T_i | Z_i)} \hat{\rho}(W_i) \right] - \tau \right).$$

Let $\hat{\vartheta}_{2,i} = R_i \delta_i I(J_i = 1)/\{\hat{\pi}(Q_i)\hat{G}(T_i|Z_i)\} + \delta_i [1 - \{\hat{\pi}(Q_i)\}^{-1} R_i] \hat{\rho}(W_i)/\hat{G}(T_i|Z_i)$. Then $S_{2,n}(b, \tau) = n^{-1/2} \sum_{i=1}^{n} Z_i [I\{g^{-1}(T_i) \leq Z_i^T b\}\hat{\vartheta}_{2,i} - \tau]$. The solution to the augmented IPW estimating equation (2.9) is referred to as the AIPW estimator and denoted by $\hat{\beta}_A(\tau)$.

Replacing the estimates $\hat{G}(\cdot)$, $\hat{\pi}(\cdot)$, and $\hat{\rho}(\cdot)$ in the estimating function $S_{2,n}(b, \tau)$ by their estimands $G(\cdot)$, $\pi(\cdot)$, and $\rho(\cdot)$, we have $E[S_{2,n}\{\beta(\tau), \tau\}] = 0$ if MAR holds and if one of the parametric models, $r(W_i, \psi)$ and $\rho(W_i, \psi)$, is correctly specified. In fact, under MAR (2.3) and consequently (2.7), $E[S_{2,n}\{\beta(\tau), \tau\}] = E(E[S_{2,n}\{\beta(\tau), \tau\} | Q_i, J_i]) = 0$ if $r(W_i, \psi)$ is correctly specified, and $E[S_{2,n}\{\beta(\tau), \tau\}] = E(E[S_{2,n}\{\beta(\tau), \tau\} | Q_i, \delta_i = 1]) = 0$ if $\rho(W_i, \psi)$ is correctly specified. This leads to the double robustness property of the AIPW estimator that $\hat{\beta}_A(\tau)$ is consistent for $\beta(\tau)$ provided that at least one of $\hat{r}(\cdot)$ and $\hat{\rho}(\cdot)$ is a consistent estimator for $r(\cdot)$ and $\rho(\cdot)$. The missing at random assumption MAR is essential for $r(W_i)$ and $\rho(W_i)$ to be identifiable. Violation of MAR may result in inconsistent estimation of both $r(\cdot)$ and $\rho(\cdot)$, and thus render both the IPW and AIPW estimators inconsistent. This property is further demonstrated in our simulation study in Section 5.

The augmented estimating equation (2.9) follows the ideas of Robins, Rotnitzky, and Zhao (1994) for efficient augmentation, whereas $(X_i, \delta_i, \delta_i J_i, Z_i)$ is considered as the full data and the full data estimating equation is (2.2), as given by Peng and Fine (2009). It is interesting to note that Peng and Fine's estimating

equation (2.2) is, in turn, based on the inverse probability weighting for censoring of the estimating equation for the full data (T_i, J_i, Z_i), while the observed data in their case is $(X_i, \delta_i, \delta_i J_i, Z_i)$. It would be desirable to improve the efficiency of the Peng and Fine (2009) estimator with augmentation. By Robins, Rotnitzky, and Zhao (1994), the efficient augmentation of (2.2) requires the estimation of the conditional expectation $E[I\{g^{-1}(T_i) \leq Z_i^T b, J_i = 1\}|X_i, \delta_i, \delta_i J_i, Z_i]$, which is unobtainable since the conditional distribution of $(T_i, J_i = 1)$ given $(C_i, \delta_i = 0, Z_i)$ is not identifiable based on the observed competing risks data. Its implementation would require some untestable and perhaps unreasonable/conflicting assumptions, such as independence of $(T_i, J_i = 1)$ and $(C_i, \delta_i = 0)$ given Z_i.

The numerical procedure for solving equation (2.9) is equivalent to locating the minimizer of the function:

$$U_{2,n}(b, \tau) = \sum_{i=1}^{n} \hat{\vartheta}_{2,i} \left| g^{-1}(X_i) - Z_i^T b \right| + \left| M + \sum_{i=1}^{n} \hat{\vartheta}_{2,i} Z_i^T b \right| + \left| M - \sum_{i=1}^{n} 2\tau Z_i^T b \right|, \quad (2.10)$$

where M is a large positive number. Equivalency is due to the fact that $U_{2,n}(b, \tau)$ is a convex function in b and its derivative is $2n^{1/2} S_{2,n}(b, \tau)$ when M exceeds $|\sum_{i=1}^{n} \hat{\vartheta}_{2,i} Z_i^T b|$ and $|\sum_{i=1}^{n} 2\tau Z_i^T b|$ for all b within the compact parameter set of $\beta(\tau)$. Under (2.1), $Z_i^T \beta(\tau) = g^{-1}\{F_1^{-1}(\tau|Z_i)\}$. It is necessary that $Z_i^T b \leq \max_{1 \leq i \leq n} |g^{-1}(X_i)|$ holds for all i for some parameter vector b for $\beta(\tau)$ to be identifiable. Hence $|\sum_{i=1}^{n} \hat{\vartheta}_{2,i} Z_i^T b| \leq \sum_{i=1}^{n} |\hat{\vartheta}_{2,i}| \max_{1 \leq i \leq n} |g^{-1}(X_i)|$ and $|\sum_{i=1}^{n} 2\tau Z_i^T b| \leq 2n\tau \max_{1 \leq i \leq n} |g^{-1}(X_i)|$. We further notice that $|\hat{\vartheta}_{2,i}| \leq 3/\{\hat{\pi}(Q_i)\hat{G}(T_i|Z_i)\}$. For most practical applications, where the missingness probabilities are less than 0.9 and fewer than 90% of subjects are censored, it is reasonable to assume that $\sum_{i=1}^{n} 3/\{\hat{\pi}(Q_i) \hat{G}(T_i|Z_i)\} \leq 300n$. Then it suffices to take $M \geq 300n \max_{1 \leq i \leq n} |g^{-1}(X_i)|$. One can use a number greater than 300 in the lower bound for M in more extreme situations. Similarly, the estimating equation (2.6) can be solved by minimizing (2.10) with $\hat{\vartheta}_{2,i}$ replaced by $\hat{\vartheta}_{1,i}$, and the same choice of M can be used in the minimization.

3. Asymptotic Properties

Throughout the rest of the paper, we assume the censoring distribution does not depend on the covariates, i.e., $G(t|Z_i) = G(t)$, and use the Kaplan-Meier estimator $\hat{G}(t)$ to estimate $G(t)$. The independence assumption for C_i and Z_i can be relaxed, in which case the conditional Kaplan-Meier estimator (Beran (1981)) can be used to estimate $G(t|Z_i)$, and the asymptotic distributions for $\hat{\beta}_I(\tau)$ and $\hat{\beta}_A(\tau)$ need to be modified to accommodate the additional variations. This section derives the uniform consistency and weak convergence of the proposed estimators $\hat{\beta}_I(\tau)$ and $\hat{\beta}_A(\tau)$, for τ over the interval $[\tau_L, \tau_U]$, under the conditions C1–C5

given in the Appendix. It also compares the asymptotic efficiency of the two estimators.

Under C5, $n^{1/2}(\hat{\psi}-\psi)$ and $n^{1/2}(\hat{\phi}-\phi)$ are asymptotically linear with influence functions η_i and ζ_i, respectively, such that

$$n^{1/2}(\hat{\psi} - \psi) = n^{-1/2} \sum_{i=1}^{n} \eta_i + o_p(1), \tag{3.1}$$

$$n^{1/2}(\hat{\phi} - \phi) = n^{-1/2} \sum_{i=1}^{n} \zeta_i + o_p(1), \tag{3.2}$$

where $\{(\eta_i, \zeta_i), i = 1, \ldots, n\}$ are *i.i.d.* random variables with $E\eta_i = 0$ and $E\zeta_i = 0$. Under the logistic regression model for $r(W_i, \psi)$, we can write $\eta_i = I_\psi^{-1} \pi'_\psi(Q_i, \psi) \delta_i \{R_i - \pi_i(Q_i)\} / [\pi_i(Q_i)\{1 - \pi_i(Q_i)\}]$, where I_ψ is the asymptotic information matrix of the likelihood function (2.5).

Let $N_i^G(t) = I(X_i \leq t, \delta_i = 0)$, $Y_i(t) = I(X_i \geq t)$, $y(t) = P(X_i \geq t)$ and $M_i^G(t) = N_i^G(t) - \int_0^t Y_i(s)\lambda^G(s)\,ds$, where $\lambda^G(t)$ is the hazard function for the censoring variable C. Let $w_1(b, t) = E[Z_i Y_i(t) I\{X_i \leq g(Z_i^T b)\} \vartheta_{1,i}] y^{-1}(t)$, $w_2(b) = -E(Z_i \{\pi'_\psi(Q_i, \psi)\}^T \{\pi(Q_i)\}^{-1} I[X_i \leq g\{Z_i^T b\}] \vartheta_{1,i})$. Under MAR and the independent censoring assumption, it is easy to see that $w_2(b) = -E(Z_i \{\pi'_\psi(Q_i, \psi)\}^T \{\pi(Q_i)\}^{-1} \rho(W_i) I[T_i \leq g\{Z_i^T b\}])$. Let $a_{1,i}(\tau) = Z_i(I[X_i \leq g\{Z_i^T \beta(\tau)\}] \vartheta_{1,i} - \tau)$, $a_{2,i}(\tau) = Z_i(I[X_i \leq g\{Z_i^T \beta(\tau)\}] \vartheta_{2,i} - \tau)$, $b_i(\tau) = \int_0^\infty w_1(\beta(\tau), s)\,dM_i^G(s)$, $c_i(\tau) = w_2(\beta(\tau))\eta_i$, and $e_i(\tau) = Z_i I[X_i \leq g\{Z_i^T \beta(\tau)\}] \delta_i \{G(T_i)\}^{-1} [1 - R_i\{\pi(Q_i)\}^{-1}] \rho(W_i)$. Define $\xi_{1,i}(\tau) = a_{1,i}(\tau) + b_i(\tau) + c_i(\tau)$ and $\xi_{2,i}(\tau) = a_{2,i}(\tau) + b_i(\tau)$. Let $\beta(\tau)$ be the true regression coefficient at τ.

Theorem 1. *Under* C1−C5, *given in the Appendix, we have* $\lim_{n\to\infty} \sup_{\tau \in [\tau_L, \tau_U]} \|\hat{\beta}_I(\tau) - \beta(\tau)\| = 0$ *and* $\lim_{n\to\infty} \sup_{\tau \in [\tau_L, \tau_U]} \|\hat{\beta}_A(\tau) - \beta(\tau)\| = 0$ *in probability, where* $\|\cdot\|$ *is the Euclidean norm.*

We show in the Appendix that the asymptotic approximations hold for the IPW estimator and the AIPW estimator uniformly in $\tau \in [\tau_L, \tau_U]$ in probability:

$$n^{1/2}\{\hat{\beta}_I(\tau) - \beta(\tau)\} = n^{-1/2} \sum_{i=1}^{n} \left[A\{\beta(\tau)\}\right]^{-1} \xi_{1,i}(\tau) + o_p(1), \tag{3.3}$$

$$n^{1/2}\{\hat{\beta}_A(\tau) - \beta(\tau)\} = n^{-1/2} \sum_{i=1}^{n} \left[A\{\beta(\tau)\}\right]^{-1} \xi_{2,i}(\tau) + o_p(1), \tag{3.4}$$

where $A\{\beta(\tau)\} = E(Z_i^{\otimes 2} f_1[g\{Z_i^T \beta(\tau)\}|Z_i])$ and $f_1(t|z) = \partial F_1(t|z)/\partial t$. The approximations (3.3) and (3.4) lead to the following asymptotic results.

Theorem 2. *Under* C1−C5, *given in the Appendix, we have*

(i) *both* $n^{1/2}\{\hat{\beta}_I(\tau) - \beta(\tau)\}$ *and* $n^{1/2}\{\hat{\beta}_A(\tau) - \beta(\tau)\}$ *converge weakly to mean zero Guassian processes with covariance matrices* $\Phi_1(\tau', \tau) = [A\{\beta(\tau')\}]^{-1}\Sigma_1(\tau', \tau)[A\{\beta(\tau)\}]^{-1}$ *and* $\Phi_2(\tau', \tau) = [A\{\beta(\tau')\}]^{-1}\Sigma_2(\tau', \tau)[A\{\beta(\tau)\}]^{-1}$ *for* $\tau, \tau' \in [\tau_L, \tau_U]$, *respectively, where* $\Sigma_1(\tau', \tau) = E\{\xi_{1,i}(\tau')\xi_{1,i}^T(\tau)\}$ *and* $\Sigma_2(\tau', \tau) = E\{\xi_{2,i}(\tau') \xi_{2,i}^T(\tau)\}$;

(ii) *the AIPW estimator* $\hat{\beta}_A(\tau)$ *is more efficient than the IPW estimator* $\hat{\beta}_I(\tau)$ *with* $\Sigma_1(\tau', \tau) \geq \Sigma_2(\tau', \tau)$.

4. Estimation of the Covariance Matrices

In quantile regression, the estimating functions are not smooth and the asymptotic covariances for the estimators of the regression coefficients involve a subdensity function, which poses difficulties for the estimation of the covariances. Huang (2002) proposed a novel variance estimation procedure for a calibration regression model using the local functional linearity of the estimating functions. Peng and Fine (2009) generalized this technique to the competing risks setting. Our estimators of the asymptotic covariances are constructed following the exposition of Peng and Fine (2009).

First we derive an estimator for $\Sigma_1(\tau, \tau)$. It is shown in the Appendix that

$$
\begin{aligned}
E[\xi_{1,i}(\tau')&\{\xi_{1,i}(\tau)\}^T] \\
&= E\{a_{1,i}(\tau')a_{1,i}^T(\tau)\} - E\{b_i(\tau')b_i^T(\tau)\} - w_2\{\beta(\tau')\}I_\psi^{-1}[w_2\{\beta(\tau)\}]^T. \quad (4.1)
\end{aligned}
$$

Let $(\hat{I}_\psi)^{-1}$ be the estimator of the variance of $\hat{\psi}$ and let $\hat{w}_2\{\hat{\beta}_I(\tau)\} = -n^{-1}\sum_{i=1}^{n} Z_i\{\pi'_\psi(Q_i, \hat{\psi})\}^T \{\hat{\pi}(Q_i)\}^{-1} I[X_i \leq g\{Z_i^T\hat{\beta}_I(\tau)\}]\hat{\vartheta}_{1,i}$. Based on (4.1), $\Sigma_1(\tau, \tau)$ can be consistently estimated by

$$
\begin{aligned}
\hat{\Sigma}_1(\tau, \tau) = &n^{-1}\sum_{i=1}^{n} Z_i^{\otimes 2}\left(\hat{\vartheta}_{1,i}I[X_i \leq g\{Z_i^T\hat{\beta}_I(\tau)\}] - \tau\right)^2 \\
&-n^{-1}\sum_{i=1}^{n}(1 - \delta_i)\left(\frac{\sum_{j=1}^{n} Z_j I[X_i \leq X_j \leq g\{Z_j^T\hat{\beta}_I(\tau)\}]\hat{\vartheta}_{1,j}}{\sum_{j=1}^{n} I(X_j \geq X_i)}\right)^{\otimes 2} \\
&-\hat{w}_2\{\hat{\beta}_I(\tau)\}(\hat{I}_\psi)^{-1}[\hat{w}_2\{\hat{\beta}_I(\tau)\}]^T.
\end{aligned} \quad (4.2)
$$

Next, since $\xi_{2,i}(\tau) = a_{2,i}(\tau) + b_i(\tau)$, with similar arguments to the proof of (4.1) we obtain

$$
E[\xi_{2,i}(\tau')\{\xi_{2,i}(\tau)\}^T] = E\{a_{2,i}(\tau')a_{2,i}^T(\tau)\} - E\{b_i(\tau')b_i^T(\tau)\}. \quad (4.3)
$$

Thus $\Sigma_2(\tau, \tau)$ can be consistently estimated by

$$\hat{\Sigma}_2(\tau, \tau) = n^{-1} \sum_{i=1}^{n} Z_i^{\otimes 2} \left(\hat{\vartheta}_{2,i} I[X_i \leq g\{Z_i^T \hat{\beta}_I(\tau)\}] - \tau \right)^2$$

$$-n^{-1} \sum_{i=1}^{n} (1 - \delta_i) \left(\frac{\sum_{j=1}^{n} Z_j I[X_i \leq X_j \leq g\{Z_j^T \hat{\beta}_I(\tau)\}] \hat{\vartheta}_{2,j}}{\sum_{j=1}^{n} I(X_j \geq X_i)} \right)^{\otimes 2}. \quad (4.4)$$

The estimation of the covariance $\Phi_1(\tau', \tau)$ of $n^{1/2}\{\hat{\beta}_I(\tau') - \beta(\tau')\}$ and $n^{1/2}\{\hat{\beta}_I(\tau) - \beta(\tau)\}$ is outlined as follows.

1. Find a symmetric and nonsingular $(p+1) \times (p+1)$ matrix $E_n(\tau) \equiv \{e_{n,1}(\tau), \ldots, e_{n,p+1}(\tau)\}$ such that $\hat{\Sigma}_1(\tau, \tau) = \{E_n(\tau)\}^2$.

2. Calculate $D_n(\tau) = ([S_{1,n}\{e_{n,1}(\tau), \tau\}]^{-1} - \hat{\beta}_I(\tau), \ldots, [S_{1,n}\{e_{n,p+1}(\tau), \tau\}]^{-1} - \hat{\beta}_I(\tau))$, where $\{S_{1,n}(e, \tau)\}^{-1}$ is the solution to $S_{1,n}(b, \tau) - e = 0$. Similar to (2.6) in Section 2.2, $S_{1,n}(b, \tau) - e = 0$ can be solved by minimizing

$$U_{3,n}(b, \tau) = \sum_{i=1}^{n} \hat{\vartheta}_{1,i} \left| g^{-1}(X_i) - Z_i^T b \right| + \left| M + \sum_{i=1}^{n} \hat{\vartheta}_{2,i} Z_i^T b \right| + \left| M - \sum_{i=1}^{n} 2\tau Z_i^T b \right|$$

$$+ \left| M - 2n^{1/2} e^T b \right|.$$

3. Estimate $\Phi_1(\tau', \tau)$ by $\hat{\Phi}_1(\tau', \tau) = n D_n(\tau')\{E_n(\tau')\}^{-1} \hat{\Sigma}_1(\tau', \tau)\{E_n(\tau)\}^{-1} D_n(\tau)$. In the special case of $\tau' = \tau$, $\hat{\Phi}_1(\tau, \tau) = n\{D_n(\tau)\}^{\otimes 2}$.

The estimation of the covariance $\Phi_2(\tau', \tau)$ of $n^{1/2}\{\hat{\beta}_A(\tau') - \beta(\tau')\}$ and $n^{1/2}\{\hat{\beta}_A(\tau) - \beta(\tau)\}$ follows the same procedure as above by replacing $\hat{\Sigma}_1(\tau', \tau)$ with $\hat{\Sigma}_2(\tau', \tau)$ and $S_{1,n}(e, \tau)$ with $S_{2,n}(e, \tau)$. The proof of the consistency of the variance estimators is similar to that in Peng and Fine (2009), and thus is omitted.

5. Simulation Study

5.1. Assessment of estimation under correctly specified models

The simulation study examines finite-sample performance of the IPW estimator and the AIPW estimator, along with the omniscient estimator (Omni) that assumes complete knowledge of J_i for uncensored failure times, and the complete-case estimator (CC) that deletes observations with missing causes. The Omni and CC estimators are computed via Peng and Fine's (2009).

Let $Z_i = (1, Z_{i,1}, Z_{i,2})$, where $Z_{i,1}$ is a uniform random variable on $(0, 1)$ and $Z_{i,2}$ is Bernoulli with probability of success equal to 0.5. The failure type J_i takes values of 1 and 2 with $P(J_i = 1|Z_i) = p_0 I(Z_{i,2} = 0) + p_1 I(Z_{i,2} = 1)$. The failure time T_i follows the conditional distributions $P(T_i < t|J_i = 1, Z_i) =$

$\Phi(\log t - \gamma^T Z_i)$ and $P(T_i < t | J = 2, Z_i) = \Phi(\log t - \alpha^T Z_i)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0,1)$, $\gamma = (\gamma_0, \gamma_1, \gamma_2)$, and $\alpha = (\alpha_0, \alpha_1, \alpha_2)$. With this set-up, the underlying τth conditional quantile of T_i is

$$\log F_1^{-1}(\tau | Z_i) = \inf \{t : P(\log T_i < t, J_i = 1 | Z_i) \geq \tau\}$$
$$= \beta_0(\tau) + \beta_1(\tau) Z_{i,1} + \beta_2(\tau) Z_{i,2}, \tag{5.1}$$

where $\beta_0(\tau) = \gamma_0 + \Phi^{-1}(\tau/p_0)$, $\beta_1(\tau) = \gamma_1$, and $\beta_2(\tau) = \gamma_2 + \Phi^{-1}(\tau/p_1) - \Phi^{-1}(\tau/p_0)$. The covariate $Z_{i,2}$ has a varying effect on the cumulative incidence quantiles across different quantile levels, whereas $Z_{i,1}$ has a constant effect.

Let the censoring time C_i follow a uniform distribution on $(0, 8)$. We generated the missing failure type indicator R_i from the logistic model: $r(W_i) = P(R_i = 1 | \delta_i = 1, W_i) = \exp(W_i^T \psi) / \{1 + \exp(W_i^T \psi)\}$, where $W_i = (1, Z_{i,1}, Z_{i,2}, X_i, A_i)^T$, $X_i = \min(T_i, C_i)$, and A_i is a univariate auxiliary variable. The values $\psi = (1, -0.9, -1, 2, 0)^T$ and $\psi = (1, -1.4, -1.5, 1, 0)^T$ correspond to 20% and 40% missing failure types, respectively. Here we chose not to include A_i in the missingness model so that we could compare the IPW and AIPW estimators under different degrees of association between A_i and J_i, holding the degree of missingness fixed at the same rate. This set-up suggests that a stronger association between A_i and J_i yields a more efficient AIPW estimator under the same level of missingness.

We consider three different levels of association between A_i and J_i, which correspond to three different choices of $\rho(W_i)$ for the AIPW estimator. In Case 1, the auxiliary variable A_i is independent of failure type J_i given Z_i. For Cases 2 and 3, we let

$$P(A_i = 1 | J_i = 1, Z_i) = \theta, \qquad P(A_i = 2 | J_i = 1, Z_i) = 1 - \theta,$$
$$P(A_i = 2 | J_i = 2, Z_i) = \theta, \qquad P(A_i = 1 | J_i = 2, Z_i) = 1 - \theta,$$

where $0 \leq \theta \leq 1$. Case 2 corresponds to $\theta = 0.8$ and Case 3 corresponds to $\theta = 0.95$. A larger value of θ indicates stronger positive association between A_i and J_i given Z_i. This model set-up results in a logistic regression model for $\rho(W_i)$ with $\text{logit}\{\rho(W_i)\} = \phi_0 + \phi_1 Z_{i,1} + \phi_2 Z_{i2} + \phi_3 A_i$. For Case 1, $\phi_1 = \phi_3 = 0$, $\phi_0 = \log\{p_0/(1 - p_0)\}$ and $\phi_2 = \log[p_1(1 - p_0)/\{p_0(1 - p_1)\}]$. For Cases 2 and 3, $\phi_0 = 3\log\{\theta/(1 - \theta)\}$, $\phi_1 = 0$, $\phi_2 = \log[p_1(1 - p_0)/\{p_0(1 - p_1)\}]$ and $\phi_3 = 2\log\{(1 - \theta)/\theta\}$.

We set $p_0 = 0.8$, $p_1 = 0.6$, $\gamma = (0, 0.5, -0.5)^T$, and $\alpha = (0, 0, -0.5)^T$. Under this setting, on average 55% of the subjects fail from type-1 failure, 25% fail from type-2 failure, and the remaining 20% are right-censored. The performances of the four estimators, Omni, CC, IPW and AIPW, for $\beta(\tau)$ at $\tau = 0.2$ and 0.4 with sample sizes $n = 200$ and $n = 500$ and two missing-causes percentages

Table 1. The average bias (Bias), empirical standard deviation (EmpSD), mean estimated standard deviation (EstSD), and empirical coverage probability (CovP) of 95% confidence intervals at $\tau = 0.2$, based on 500 simulated data sets with 20% missing causes.

Method	Bias×10^3			EmpSD×10^3			EstSD×10^3			CovP×10^2		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
						$n = 200$						
Omni	5	-24	10	274	438	245	304	486	277	92.4	92.0	92.6
CC	68	71	184	288	477	265	301	490	296	91.0	90.8	87.8
IPW	-9	-5	21	337	575	307	341	574	334	92.0	90.0	93.6
AIPW(Case 1)	-8	0	24	316	531	281	334	547	314	91.0	90.4	92.6
AIPW(Case 2)	-10	-2	22	297	497	269	308	521	297	92.2	93.0	94.6
AIPW(Case 3)	-1	-21	17	284	461	246	303	499	280	91.6	91.4	94.2
						$n = 500$						
Omni	7	-7	2	173	278	156	170	293	166	91.8	93.4	93.6
CC	83	64	163	184	297	165	178	293	174	87.0	91.2	79.8
IPW	-2	8	-5	196	338	192	207	347	203	91.6	91.2	92.0
AIPW(Case 1)	4	0	1	188	314	182	198	333	189	92.6	93.2	91.2
AIPW(Case 2)	2	3	2	181	304	171	194	314	182	93.8	93.2	93.4
AIPW(Case 3)	4	1	5	178	285	162	178	290	171	91.0	92.0	93.0

are summarized in Tables 1−4. The tables report the bias, empirical standard deviation, mean estimated standard deviation, and empirical coverage probability of 95% Wald-type confidence intervals based on 500 simulated data sets.

Note that the choices of $\rho(W_i)$ do not change the IPW estimator. The CC estimator had substantial bias for all scenarios. The IPW and AIPW estimators performed comparably to the Omni estimator with very small biases. In addition, the estimated standard deviations matched very well with the empirical ones, and the 95% confidence intervals had reasonable coverage probabilities, except for the CC estimator.

For the analysis of Mashi data presented in the next section, small values of $\tau = 0.005, 0.01$, and 0.02 were considered due to small percentages of HIV-related and HIV-unrelated deaths. Furthermore, the Mashi analysis had a larger sample size. To mimic Mashi, additional simulations at $\tau = 0.01$ with $n = 1200$ were conducted. The results, reported in Table S.1 of the Supplementary Material, show that the biases of the AIPW estimator remain small under 20% and 40% of missing causes. The biases of the IPW estimator are also small under 20% of missing causes. At 40% of missing causes, the biases of the IPW estimator are large compared to those for the AIPW estimator, but these biases for the slope coefficients are still smaller than those of the CC estimator.

Table 5 shows the Pitman relative efficiencies (ratios of variances) for the IPW and AIPW estimators with respect to the Omni estimator. By incorporating

Table 2. The average bias (Bias), empirical standard deviation (EmpSD), mean estimated standard deviation (EstSD), and empirical coverage probability (CovP) of 95% confidence intervals at $\tau = 0.4$, based on 500 simulated data sets with 20% missing causes.

Method	Bias×10³			EmpSD×10³			EstSD×10³			CovP×10²		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
						$n = 200$						
Omni	6	15	8	307	514	308	343	566	338	92.4	93.2	93.6
CC	77	28	148	313	538	334	345	565	351	92.0	93.4	91.4
IPW	-7	32	25	358	630	356	374	639	398	90.6	92.8	94.0
AIPW(Case 1)	-4	28	34	341	599	354	372	640	397	92.2	94.0	93.4
AIPW(Case 2)	-3	22	30	332	572	349	351	587	370	92.6	92.6	92.8
AIPW(Case 3)	5	13	13	315	532	321	334	557	342	92.0	94.8	94.6
						$n = 500$						
Omni	11	-20	18	199	323	193	196	326	199	91.8	93.0	94.2
CC	88	-13	141	198	322	193	196	337	210	89.0	93.2	89.0
IPW	4	-4	15	218	363	217	218	363	230	91.0	91.6	94.8
AIPW(Case 1)	8	-15	20	214	353	220	221	363	224	91.0	94.6	93.8
AIPW(Case 2)	6	-9	18	208	344	207	213	357	213	92.2	95.0	94.2
AIPW(Case 3)	10	-17	19	205	335	196	204	334	206	92.4	94.2	95.0

Table 3. The average bias (Bias), empirical standard deviation (EmpSD), mean estimated standard deviation (EstSD), and empirical coverage probability (CovP) of 95% confidence intervals at $\tau = 0.2$, based on 500 simulated data sets with 40% missing causes.

Method	Bias×10³			EmpSD×10³			EstSD×10³			CovP×10²		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
						$n = 200$						
Omni	5	-24	10	274	438	245	304	486	277	92.4	92.0	92.6
CC	42	226	346	316	551	360	362	606	370	92.2	91.0	79.0
IPW	-51	45	66	430	768	436	450	779	425	89.6	91.0	89.2
AIPW(Case 1)	2	-18	35	363	628	371	381	632	363	91.2	92.4	90.8
AIPW(Case 2)	2	-16	32	338	575	330	343	548	328	91.0	90.8	93.0
AIPW(Case 3)	3	-16	12	286	477	265	309	502	293	91.8	92.8	93.4
						$n = 500$						
Omni	7	-7	2	173	278	156	170	293	166	91.8	93.4	93.6
CC	46	239	325	204	348	212	209	357	221	90.8	86.6	69.4
IPW	1	7	9	256	467	257	266	473	263	93.4	92.4	91.0
AIPW(Case 1)	-1	0	19	221	399	224	222	388	222	92.8	91.2	92.2
AIPW(Case 2)	6	-13	10	207	365	202	213	354	204	92.2	93.4	91.0
AIPW(Case 3)	5	-5	10	176	294	169	187	303	173	92.6	94.0	92.6

information from the missing failure types, AIPW improved efficiency over IPW, with greater improvement when there was a stronger association between the

Table 4. The average bias (Bias), empirical standard deviation (EmpSD), mean estimated standard deviation (EstSD), and empirical coverage probability (CovP) of 95% confidence intervals at $\tau = 0.4$, based on 500 simulated data sets with 40% missing causes.

Method	Bias×10³			EmpSD×10³			EstSD×10³			CovP×10²		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
					$n = 200$							
Omni	6	15	8	307	514	308	343	566	338	92.4	93.2	93.6
CC	43	211	395	347	612	432	388	671	492	91.6	90.6	84.4
IPW	-13	57	49	467	840	543	489	866	593	93.8	94.4	93.4
AIPW(Case 1)	22	-4	52	420	766	505	438	786	508	94.6	95.0	93.0
AIPW(Case 2)	15	-4	24	394	667	416	402	676	435	93.0	93.4	92.8
AIPW(Case 3)	-1	11	19	338	539	364	341	594	354	92.0	95.4	91.8
					$n = 500$							
Omni	11	-20	18	199	323	193	196	326	199	91.8	93.0	94.2
CC	53	174	389	212	367	264	227	382	274	91.4	90.8	72.8
IPW	0	14	26	267	494	310	276	494	322	92.2	92.0	93.8
AIPW(Case 1)	3	-4	42	245	443	284	254	450	293	94.4	95.6	93.8
AIPW(Case 2)	11	-21	33	234	410	257	237	413	269	92.6	93.8	95.4
AIPW(Case 3)	8	-19	27	207	352	212	212	355	218	92.2	94.2	95.0

Table 5. Pitman relative efficiencies of the IPW and AIPW estimators with respect to the *Omni* estimator based on 500 simulated data sets. The maximum standard error of the relative efficiencies is 0.07. MP stands for the missingness proportion of failure causes.

n	IPW			AIPW(Case 1)			AIPW(Case 2)			AIPW(Case 3)		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
				MP=20%, $\tau = 0.2$								
200	0.66	0.58	0.64	0.75	0.68	0.76	0.85	0.78	0.83	0.93	0.90	0.99
500	0.78	0.68	0.66	0.85	0.79	0.74	0.91	0.84	0.84	0.95	0.95	0.93
				MP=20%, $\tau = 0.4$								
200	0.74	0.67	0.75	0.81	0.74	0.76	0.86	0.81	0.78	0.95	0.93	0.92
500	0.83	0.79	0.79	0.87	0.84	0.76	0.91	0.88	0.86	0.94	0.93	0.96
				MP=40%, $\tau = 0.2$								
200	0.41	0.33	0.32	0.57	0.49	0.44	0.65	0.58	0.55	0.92	0.84	0.85
500	0.46	0.36	0.37	0.62	0.49	0.49	0.70	0.58	0.60	0.97	0.90	0.85
				MP=40%, $\tau = 0.4$								
200	0.43	0.37	0.32	0.53	0.45	0.37	0.61	0.59	0.55	0.83	0.91	0.72
500	0.56	0.43	0.39	0.66	0.53	0.46	0.72	0.62	0.56	0.92	0.84	0.82

auxiliary variable A_i and J_i. For Case 3 with $n = 500$, the efficiencies of AIPW were comparable to those of the Omni estimator.

5.2. On robustness of estimation

To assess how sensitive the proposed methods are to model misspecifications for $r(W_i)$ and/or $\rho(W_i)$, and to violations of the missing at random assumption, we consider four additional cases, namely, Cases 4-7. In Case 4, instead of a logistic model we generated the missing failure type indicator R_i from the probit model: $r(W_i) = \Phi(W_i^T \psi)$, where $\psi = (1, -0.9, -1.4, 2, 0)^T$ and $W_i = (1, Z_{i,1}, Z_{i,2}, X_i, A_i)^T$; whereas both the IPW and AIPW estimators still use logistic regression to estimate $r(W_i)$, and excluding X_i from W_i. Case 5 has the same design as Case 2, and Case 6 has the same design as Case 4. In both Cases 5 and 6, $\rho(W_i)$ is estimated by excluding the important variable A_i in the logistic regression. Therefore, $r(W_i)$ is misspecified in Case 4, $\rho(W_i)$ is misspecified in Case 5, and both models are misspecified in Case 6. In Case 7, we generated the missing failure type indicator R_i from the logistic model: $P(R_i = 1 | \delta_i = 1, J_i, W_i^T) = \exp(W_i^T \psi - J_i)/\{1 + \exp(W_i^T \psi - J_i)\}$, where $W_i = (1, Z_{i,1}, Z_{i,2}, X_i, A_i)^T$ and $\psi = (2.5, -0.9, -1, 2, 0)^T$. Since the probability of missingness depends on the unobserved failure type J_i, the missing at random assumption is violated in Case 7. In all four cases, the missing-cause proportion is 20%.

Table 6 reports the bias, empirical standard deviation, mean estimated standard deviation, and empirical coverage probability of 95% Wald-type confidence intervals for the Omni, CC, IPW, and AIPW estimators based on 500 simulated data sets for Cases 4-7 at $\tau = 0.2$ and with $n = 500$. The presentations of the summaries for different τ values are given in Figures S.1-6 of the Supplementary Material. When $r(W_i)$ was misspecified, the IPW estimator performed similar to the CC estimator, both having large biases for estimating $\beta_2(\tau)$. As expected from its double robustness property, the AIPW estimator performed well in Cases 4-5 when one of the two models for $r(W_i)$ and $\rho(W_i)$ was misspecified. Since the IPW estimator does not utilize $\rho(W_i)$, there is no misspecification for the IPW estimator under Case 5. When both models were misspecified in Case 6, the AIPW estimator had slightly larger biases than in Case 4, but still outperformed the CC estimator, in particular, for $\beta_1(\tau)$ and $\beta_2(\tau)$. Since both the IPW and AIPW estimators are developed based on the MAR assumption, it is no surprise that the IPW and AIPW estimators showed no improvement over the CC estimator in Case 7.

6. Analysis of the Mashi data

The Mashi trial investigated the effect of formula- versus breast-feeding plus extended infant zidovudine prophylaxis among HIV-infected expecting mothers in Botswana (Thior et al. (2006)). Five-hundred and ninety-one women were randomized to formula feeding from birth plus 1 month of infant zidovudine (FF),

Table 6. Method robustness. The average bias (Bias), empirical standard deviation (EmpSD), mean estimated standard deviation (EstSD), and empirical coverage probability (CovP) of 95% confidence intervals at $\tau = 0.2$ with $n = 500$, based on 500 simulated data sets with 20% missing causes.

Method	Bias$\times 10^3$			EmpSD$\times 10^3$			EstSD$\times 10^3$			CovP$\times 10^2$		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
Case 4: $r(w)$ is misspecified												
Omni	4	1	3	169	275	161	177	291	166	92.8	92.0	93.8
CC	10	92	333	166	274	157	169	281	162	92.8	91.2	50.0
IPW	18	73	294	171	278	159	170	286	160	92.2	91.2	54.2
AIPW	26	-51	-34	173	284	160	176	281	166	92.4	90.6	93.6
Case 5: $\rho(w)$ is misspecified												
Omni	7	-7	2	173	278	156	170	293	166	91.8	93.4	93.6
CC	83	64	163	184	297	165	178	293	174	87.0	91.2	79.8
IPW	-2	8	-5	196	338	192	207	347	203	91.6	91.2	92.0
AIPW	6	-5	-3	188	315	179	193	334	193	91.6	93.0	93.4
Case 6: both $r(w)$ and $\rho(w)$ are misspecified												
Omni	4	1	3	169	275	161	177	291	166	92.8	92.0	93.8
CC	10	92	333	166	274	157	169	281	162	92.8	91.2	50.0
IPW	18	73	294	171	278	159	170	286	160	92.2	91.2	54.2
AIPW	35	-72	-46	173	287	162	175	285	167	92.8	91.6	92.6
Case 7: missing-at-random assumption is violated												
Omni	7	-7	2	173	278	156	170	293	166	91.8	93.4	93.6
CC	45	33	58	179	293	162	179	297	171	91.8	93.6	91.4
IPW	-5	-41	-95	191	328	179	196	331	188	91.0	90.4	89.6
AIPW	-7	-42	-85	180	300	167	184	295	170	92.0	91.2	89.4

and 588 women were randomized to breast-feeding from birth plus 6 months of infant zidovudine (BF+AZT). Live first-born infants were followed for 18-months for occurrence of the two primary endpoints, HIV infection and death. HIV-PCR tests were administered at visits at birth and at monthly ages of 1, 2, 3, 4, 5, 6, 7, 9, 12, 18 (with little missing data). The primary objectives assessed the treatment effect on these endpoints separately, as well as on the composite endpoint defined as the first event of HIV infection or death. A secondary objective was to assess the treatment effect on death due to HIV infection, which we refer to as HIV-related death. We apply the methods above to assess $F_1^{-1}(\tau|z)$ with $J = 1$ HIV-related death and $J = 2$ HIV-unrelated death.

We take a death to be HIV-related if either (1) the study clinicians deemed the death HIV-1 related ($n = 4$ deaths), or (2) the infant had at least one positive test result from the PCR assay used to test for HIV infection prior to death ($n = 24$ deaths). In addition, we take a death to be HIV-unrelated if the study clinician deemed the death unrelated to HIV/AIDS ($n = 22$ deaths). Of the 111 live-born infants who died, the cause of death is known in 50 cases and

missing for 61.

Considering 20 covariates of the babies or their mothers, we used logistic regression and all-subsets model selection (with criterion Mallow's Cp) to select a model for predicting among cases whether J was observed. The model included the following variables (estimated regression coefficient): the infant had birthweight < 2.5 kilograms (1.21); the randomization assignment of mom/baby to receive Placebo/Placebo was switched to Placebo/Nevirapine part-way during the trial due to a DSMB recommendation (-1.27); the infant had AZT toxicity (1.43); $\log 10$ plasma viral load level of the mom at delivery (0.98); and the baby was hospitalized with a serious adverse event (-1.20). Using the same model selection strategy for analyzing cases with known death-cause, the following variables were included in the model for predicting $J = 1$: the infant received HAART (2.42), and $\log 10$ plasma viral load level of the mom at delivery (1.70).

For assessing the treatment effect of BF+AZT versus FF we used the identity link function. The covariate of interest is $Z = c(1, Z_1)$, where Z_1 is 1 for mother-infant pairs assigned BF+AZT and 0 for FF. The estimation of the quantile is invariant to the link function in this particular case, but the estimated values of the coefficients $\beta_0(\tau)$ and $\beta_1(\tau)$ can be different for different link functions. With the identity link, $\beta_1(\tau)$ represents the treatment effect on the τth type-1 quantile. Let X be the survival time in days. According to the above logistic regressions, we let W include the variables that proved informative for $P(R = 1|\delta = 1, W)$ and/or for the probability of HIV-related death $P(J = 1|\delta = 1, W)$. We considered the subset of data with complete covariate information, that includes 1123 live-born infants (of the 1193 total), among whom 107 died and 49 died with known cause of death (28 are HIV-related). Based on the data, about 2.5% of infants died while known to be HIV infected ($J = 1$), and 54.2% of the infants who died had missing death cause.

We performed the quantile regression at $\tau = 0.005, 0.01, 0.02$ and 0.03. The analysis at $\tau = 0.005$ is interesting because it concerns early death and there were many early deaths in the data set. Table 7 summarizes the analysis results using the IPW and AIPW methods. From Table 7, by the AIPW method, the p-values for testing the treatment effect at $\tau = (0.005, 0.01, 0.02, 0.03)$ were $(0.138, 0.042, 0.062, 0.52)$, respectively. The results indicate that BF+AZT had some positive effect in postponing/reducing HIV-related deaths compared to FF at the quantiles corresponding to $\tau = 0.01$ and 0.02. Using the AIPW method, the HIV-related death rate reached 1% by 184 days for those assigned to BF+AZT, while it reached 1% by 64 days for those assigned FF. In addition, it reached 2% by 276 days for BF+AZT and 113 days for FF. This analysis suggests that it takes longer for the BF+AZT group to reach the same percentage of HIV-related deaths than the FF group, by 120–163 days. The estimated treatment

Table 7. Analysis of the Mashi data with the IPW and AIPW methods.

	IPW			AIPW			Rel. Efficiency
	Coef Est.	S.E.	p-value	Coef Est.	S.E.	p-value	IPW vs AIPW
$\tau = 0.005$							
Intercept	52.0	19.2	0.007	37.0	22.1	0.09	1.32
Treatment	38.0	61.7	0.54	102.0	68.8	0.14	1.24
$\tau = 0.01$							
Intercept	64.0	31.4	0.04	64.0	33.6	0.06	1.15
Treatment	66.0	108.9	0.54	120.0	59.1	0.04	0.29
$\tau = 0.02$							
Intercept	94.0	113.0	0.41	113.0	95.9	0.24	0.72
Treatment	182.0	107.3	0.09	163.0	87.5	0.06	0.66
$\tau = 0.03$							
Intercept	207.0	164.5	0.21	214.0	147.0	0.15	0.80
Treatment	91.0	186.7	0.63	84.0	132.1	0.52	0.50

effect using the AIPW estimator decreased at $\tau = 0.03$, and the standard error increased because of the small number of deaths after the 0.03 quantile. The estimated treatment effect was also small at $\tau = 0.005$. The IPW method did not identify a significant treatment effect at any of the quantile levels evaluated. This is attributed to the limited number of deaths and the high percentage of unknown death causes among those who died, and the AIPW method was able to recover some of the lost information by modeling the probability of HIV-related death under the missing at random assumption (2.3). The large differences in the IPW and AIPW estimation of the treatment effect at $\tau = 0.005$ and 0.01 in Table 7 reflect the fact that the IPW estimation is not numerically stable.

The difference in the performances of the AIPW and IPW estimators for the Mashi data analysis is consistent with what we observed in the simulation study. That is, the AIPW estimator shows large efficiency gain over the IPW estimator when A_i and J_i are strongly correlated, and is still more efficient than the IPW estimator even when A_i and J_i are independent. We infer that both the efficiency of the AIPW method and the informativeness of the auxiliary variables for HIV-related death contributed to the efficiency gain.

We stress that the quantile regression based on the cumulative incidence function studies the "crude" effect on the time to HIV-related death in the presence of other competing risks, i.e., HIV-unrelated death. This analysis is directly interpretable and relevant. However, it should not be used to infer the "net effect"; this would require strong untestable assumptions and/or sensitivity analysis.

In conclusion, this analysis provides additional insights over the primary study results that showed that infants assigned to formula-feed (FF) had a higher rate of all-cause mortality by 7 months of age than infants assigned BF+AZT, but a lower rate of HIV infection (Thior et al. (2006)). Prior to the current analysis, a beneficial effect of either BF+AZT or FF on HIV-related death was plausible: for BF+AZT because breast-feeding decreases the general early death rate; for FF because, by decreasing the rate of early HIV infection, it reduces the number of infants that could potentially die from HIV. The analysis here supports that the beneficial effect of formula-feeding to reduce HIV infections is overwhelmed by the stronger deleterious effect of formula-feeding to increase early deaths in HIV-infected infants. These results support breast-feeding plus antiretroviral prophylaxis during the first several months of life for infants born to HIV-infected mothers in Botswana.

Acknowledgements

Appendix

The following regularity conditions are assumed in Sections 3 and 4.

C1. There exists $\nu > 0$ such that $P(C = \nu) > 0$ and $P(C > \nu) = 0$.

C2. Z is uniformly bounded with $\sup_i \|Z_i\| < \infty$.

C3. For $0 < \tau_L \leq \tau_U < \tau_0 = \inf_z P(T \leq \nu, J = 1 \mid z)$, $\beta(\tau)$ is Lipschitz continuous for $\tau \in [\tau_L, \tau_U]$, and $f_1(t \mid z)$ is bounded in t and z, where $f_1(t \mid z) = \partial F_1(t \mid z)/\partial t$.

C4. For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{b \in \mathcal{B}(\rho_0)} \text{eigmin} A(b) \geq c_0$, where $\mathcal{B}(\rho) = \{b \in R^{p+1} : \inf_{\tau \in [\tau_L, \tau_U]} \|b - \beta(\tau)\| \leq \rho\}$, $A(b) = E[Z^{\otimes 2} f_1\{g(Z^T b)|Z\}]$, $\text{eigmin} A(b)$ is the minimum of the eigenvalues of $A(b)$, and $u^{\otimes 2} = uu^T$.

C5. $\pi(Q, \psi)$ and $\rho(Q, \phi)$ are twice differentiable with respect to ψ and ϕ, respectively; $\pi(Q, \psi) \geq \alpha > 0$; $\pi'_\psi(Q) = d\pi(Q, \psi)/d\psi$ is uniformly bounded; both $\rho(W, \phi)$ and $\rho'_\phi(W) = d\rho(Q, \phi)/d\phi$ are uniformly bounded.

The first four conditions are similar to those of Peng and Fine (2009). The condition C5 requires that the probability of non-missingness be bounded away from zero, as well as other boundedness conditions that are needed to establish weak convergence of the empirical processes.

Proof of Theorem 1. Let $S_n(b, \tau) = n^{-1/2} \sum_{i=1}^n Z_i\big[I\{g^{-1}(T_i) \le Z_i^T b\}\hat{\vartheta}_i - \tau\big]$. The following proof of the consistency holds for both the estimators that are the roots of $S_n(b, \tau)$ by taking $\hat{\vartheta}_i = \hat{\vartheta}_{1,i}$ for the IPW estimator and $\hat{\vartheta}_i = \hat{\vartheta}_{2,i}$ for AIPW estimator. Let $\vartheta_{1,i} = R_i\delta_i I(J_i = 1)\{\pi(Q_i)G(T_i)\}^{-1}$ and $\vartheta_{2,i} = R_i\delta_i I(J_i = 1)\{\pi(Q_i)G(T_i)\}^{-1} + \big[1 - \{\pi(Q_i)\}^{-1}R_i\big]\,\delta_i\rho(W_i)\,\{G(T_i)\}^{-1}$. We use $\vartheta_i = \vartheta_{1,i}$ for the IPW estimator and $\vartheta_i = \vartheta_{2,i}$ for the AIPW estimator. For brevity, $\sup_b$ and $\sup_\tau$ denote the supremum taken over $b \in \mathbb{R}^{p+1}$ and $\tau \in [\tau_L, \tau_U]$, respectively.

Let $S_n^G(b, \tau) = n^{-1/2} \sum_{i=1}^n Z_i\big[I\{g^{-1}(T_i) \le Z_i^T b\}\vartheta_i - \tau\big]$ and $\mu(b, \tau) = E(Z_i[F_1\{g(Z_i^T b)|\,Z_i\} - \tau])$. Under the missing at random assumption (2.3) and the conditional independence between (T_i, J_i) and C_i given Z_i, $E\{n^{-1/2}S_n^G(b, \tau)\} = E[E\{n^{-1/2}S_n^G(b, \tau)|Q_i, J_i\}] = \mu(b, \tau)$.

By Condition C1 and C5, for every $r > 0$, we have, $\sup_{t < \nu}|\hat{G}(t) - G(t)| = o_p(n^{-1/2+r})$, $|\hat{\psi} - \psi| = o_p(n^{-1/2+r})$ and $|\hat{\phi} - \phi| = o_p(n^{-1/2+r})$. This, coupled with C2 and C5, implies that $\sup_{\tau,b}\|n^{-1/2}\{S_n(b, \tau) - S_n^G(b, \tau)\}\| = o_p(n^{-1/2+r})$. It follows from arguments similar to those of Peng and Fine (2009) that $\sup_{\tau,b}\|n^{-1/2} S_n^G(b, \tau) - \mu(b, \tau)\| = o_p(1)$, and thus $\sup_{\tau,b}\|n^{-1/2}S_n(b, \tau) - \mu(b, \tau)\| = o_p(1)$. This, together with $\mu\{\beta(\tau), \tau\} = 0$, implies the uniform consistency of both $\hat{\beta}_I(\tau)$ and $\hat{\beta}_A(\tau)$ under C4.

Proof of Theorem 2. Let $\hat{\beta}(\tau)$ be the root of $S_n(b, \tau)$. First we show that $S_n\{\beta(\tau), \tau\}$ converges weakly to a mean zero Gaussian process and derive its asymptotic covariance matrix. Note that

$$S_n\{\beta(\tau), \tau\} = n^{-1/2}\sum_{i=1}^n Z_i\big[I\{g^{-1}(T_i) \le Z_i^T\beta(\tau)\}\vartheta_i - \tau\big]$$

$$+ n^{-1/2}\sum_{i=1}^n Z_i I\{g^{-1}(T_i) \le Z_i^T\beta(\tau)\}(\hat{\vartheta}_i - \vartheta_i). \qquad (A.1)$$

The asymptotic approximation for (A.1) is obtained below for the IPW and AIPW estimators, respectively.

For the IPW estimator, $\hat{\vartheta}_i$ and ϑ_i of (A.1) correspond to $\hat{\vartheta}_{1,i}$ and $\vartheta_{1,i}$, respectively, and $\hat{\beta}(\tau) = \hat{\beta}_I(\tau)$. Let $\delta_i^* = R_i\delta_i I(J_i = 1)$. We have

$$\hat{\vartheta}_{1,i} - \vartheta_{1,i} = -\left\{\frac{\hat{G}(X_i) - G(X_i)}{\hat{G}(X_i)G(X_i)\pi(Q_i)} + \frac{\hat{\pi}(Q_i) - \pi(Q_i)}{\hat{\pi}(Q_i)\pi(Q_i)\hat{G}(X_i)}\right\}\delta_i^*. \qquad (A.2)$$

From Pepe (1991),

$$\sup_{t\in[0,\nu)} \|n^{1/2}\{\hat{G}(t) - G(t)\} + n^{-1/2}G(t)\sum_{j=1}^{n}\int_0^t y^{-1}(s)dM_j^G(s)\| \xrightarrow{P} 0, \quad (A.3)$$

where $M_j^G(\cdot)$ and $y(\cdot)$ are defined in Section 3 just before Theorem 1.

By (3.1),

$$n^{1/2}\{\hat{\pi}(Q_i) - \pi(Q_i)\} = n^{-1/2}\{\pi'_\psi(Q_i, \psi)\}^T \sum_{j=1}^{n}\eta_j + o_p(1). \quad (A.4)$$

By (A.2), (A.3) and (A.4), the second term of (A.1) is

$$n^{-1/2}\sum_{i=1}^{n} Z_i I\{g^{-1}(T_i) \le Z_i^T\beta(\tau)\}\delta_i^*$$

$$\times \left[\{\hat{G}(X_i)\pi(Q_i)\}^{-1}n^{-1}\sum_{j=1}^{n}\int_0^{X_i} y^{-1}(s)\, dM_j^G(s) \right.$$

$$\left. - \{\hat{\pi}(Q_i)\pi(Q_i)\hat{G}(X_i)\}^{-1}n^{-1}\{\pi'_\psi(Q_i, \psi)\}^T \sum_{j=1}^{n}\eta_j \right] + o_p(1).$$

Writing $\int_0^{X_i} y^{-1}(s)dM_j^G(s) = \int_0^\infty Y_i(s)y^{-1}(s)dM_j^G(s)$ and changing the order of the summations, the above is

$$n^{-1/2}\sum_{j=1}^{n}\int_0^\infty \left(n^{-1}\sum_{i=1}^{n} Z_i I\{g^{-1}(T_i) \le Z_i^T\beta(\tau)\}\delta_i^*\{\hat{G}(X_i)\pi(Q_i)\}^{-1}Y_i(s)y^{-1}(s) \right)$$

$$dM_j^G(s) - n^{-1/2}\sum_{j=1}^{n}\left(n^{-1}\sum_{i=1}^{n} Z_i I\{g^{-1}(T_i) \le Z_i^T\beta(\tau)\}\delta_i^*\{\hat{\pi}(Q_i)\pi(Q_i)\hat{G}(X_i)\}^{-1} \right.$$

$$\left. \{\pi'_\psi(Q_i, \psi)\}^T \right)\eta_j. \quad (A.5)$$

Let $\mathcal{F} = \{Z_i I\{X_i \le g(Z_i^Tb)\}\delta_i^*\{\pi(Q_i)G(X_i)\}^{-1}Y_i(t); \text{ for } b \in \mathbb{R}^{p+1}, t \in [0,\nu)\}$. The function class $\mathcal{F}$ is Donsker, and thus Glivenko-Cantelli (van der Vaart and Wellner (1996)) because the class of indicator functions is Donsker, and Z_i, $\{\pi(Q_i)G(X_i)\}^{-1}$ and δ_i^* are uniformly bounded. It follows from the Glivenko-Cantelli Theorem that $n^{-1}\sum_{i=1}^{n} Z_i Y_i(t)\{\pi(Q_i)G(X_i)\}^{-1}I\{X_i \le g(Z_i^Tb)\}\delta_i^*y^{-1}(t)$ $\xrightarrow{P} E[Z_i I\{X_i \le g(Z_i^Tb)\}\delta_i^*\{\pi(Q_i)G(X_i)\}^{-1}Y_i(t)]y^{-1}(t)$, uniformly in both $b \in \mathbb{R}^{p+1}$ and $t \in [0,\nu)$. The limit is $w_1(b,t)$, defined in Section 3 just before Theorem 1, under MAR and the independent censoring assumption. Since $\hat{G}(X_i) =$

$G(X_i) + O_p(n^{-1/2})$ and $\hat{\pi}(Q_i) = \pi(Q_i) + O_p(n^{-1/2})$ uniformly in $i \in \{1, \ldots n\}$, $n^{-1} \sum_{i=1}^{n} Z_i I\{g^{-1}(T_i) \leq Z_i^T \beta(\tau)\} \delta_i^* \{\hat{G}(X_i)\pi(Q_i)\}^{-1} Y_i(t) y^{-1}(t) \xrightarrow{P} w_1\{\beta(\tau), t\}$ uniformly in $\tau \in [\tau_L, \tau_U]$ and $t \in [0, \nu)$.

Similarly, $-n^{-1} \sum_{i=1}^{n} Z_i I\{g^{-1}(T_i) \leq Z_i^T \beta(\tau)\} \delta_i^* \{\hat{\pi}(Q_i)\pi(Q_i)\hat{G}(X_i)\}^{-1} \{\pi_\psi'$ $(Q_i, \psi)\}^T \xrightarrow{P} w_2\{\beta(\tau)\}$ uniformly in $\tau \in [\tau_L, \tau_U]$, where $w_2(b)$ is defined in Section 3 just before Theorem 1.

By (A.1) and (A.5), the next asymptotic equivalence follows by applying Lemma 2 of Gilbert, McKeague, and Sun (2008) to (A.5):

$$S_n\{\beta(\tau), \tau\} = n^{-1/2} \sum_{i=1}^{n} \xi_{1,i}(\tau) + o_p(1). \tag{A.6}$$

uniformly in $\tau \in [\tau_L, \tau_U]$, where $\xi_{1,i}(\tau) = a_{1,i}(\tau) + b_i(\tau) + c_i(\tau)$, and $a_{1,i}(\tau)$, $b_i(\tau)$ and $c_i(\tau)$ are defined in Section 3 just before Theorem 1.

For the AIPW estimator, $\hat{\vartheta}_i$ and ϑ_i of (A.1) correspond to $\hat{\vartheta}_{2,i}$ and $\vartheta_{2,i}$, respectively, and $\hat{\beta}(\tau) = \hat{\beta}_A(\tau)$. Then $\hat{\vartheta}_{2,i} - \vartheta_{2,i}$ is

$$\hat{\vartheta}_{1,i} - \vartheta_{1,i} + \frac{\delta_i}{\hat{G}(X_i)}\left\{1 - \frac{R_i}{\hat{\pi}(Q_i)}\right\}\hat{\rho}(W_i) - \frac{\delta_i}{G(X_i)}\left\{1 - \frac{R_i}{\pi(Q_i)}\right\}\rho(W_i)$$

$$= \hat{\vartheta}_{1,i} - \vartheta_{1,i} + \left\{\frac{\delta_i}{\hat{G}(X_i)} - \frac{\delta_i}{G(X_i)}\right\}\left\{1 - \frac{R_i}{\hat{\pi}(Q_i)}\right\}\hat{\rho}(W_i)$$

$$+ \frac{\delta_i}{G(X_i)}\left\{\frac{R_i}{\pi(Q_i)} - \frac{R_i}{\hat{\pi}(Q_i)}\right\}\hat{\rho}(W_i) + \frac{\delta_i}{G(X_i)}\left\{1 - \frac{R_i}{\pi(Q_i)}\right\}\{\hat{\rho}(W_i) - \rho(W_i)\}.$$

Now, we apply the decompositions (A.3), (A.4), and (3.2), and plug them into (A.1). By the Glivenko-Cantelli Theorem, we can show that

$$n^{-1} \sum_{i=1}^{n} \delta_i Z_i Y_i(t)\{G(X_i)\}^{-1} I\{X_i \leq g(Z_i^T b)\}\{1 - R_i \pi^{-1}(Q_i)\}\rho(W_i) y^{-1}(t) \xrightarrow{P} 0,$$

$$n^{-1} \sum_{i=1}^{n} \delta_i Z_i \rho_\phi'(W_i, \phi)\{G(X_i)\}^{-1} I\{X_i \leq g(Z_i^T b)\}\{1 - R_i \pi^{-1}(Q_i)\} \xrightarrow{P} 0,$$

$$n^{-1} \sum_{i=1}^{n} \delta_i R_i Z_i \pi_\psi'(Q_i, \psi)\{\pi^2(Q_i)G(X_i)\}^{-1}\rho(W_i) I\{X_i \leq g(Z_i^T b)\} \xrightarrow{P} w_3(b),$$

uniformly in both $b \in \mathbb{R}^{p+1}$ and $t \in [0, \nu)$, where $w_3(b) = E[\delta_i Z_i \pi_\psi'(Q_i, \psi)$ $\{\pi(Q_i)G(X_i)\}^{-1}\rho(W_i) I\{X_i \leq g(Z_i^T b)\}]$. It is easy to see that $w_3(b) = -w_2(b)$.

Using similar techniques as for the IPW estimator, we obtain

$$S_n\{\beta(\tau), \tau\} = n^{-1/2} \sum_{i=1}^{n} \xi_{2,i}(\tau) + o_p(1). \tag{A.7}$$

uniformly in $\tau \in [\tau_L, \tau_U]$ in probability, where $\xi_{2,i}(\tau) = a_{2,i}(\tau) + b_i(\tau)$ and $a_{2,i}(\tau) = Z_i(I[X_i \leq g\{Z_i^T \beta(\tau)\}]\vartheta_{2,i} - \tau)$.

We have derived the asymptotic approximations of $S_n\{\beta(\tau), \tau\}$ in (A.6) and (A.7) for the IPW estimator and AIPW estimator, respectively. It is obvious that the function class $\{c_i(\tau), \tau \in [\tau_L, \tau_U]\}$ is Donsker. Applying the similar arguments for $\mathcal{F}$, the function classes $\{a_{1,i}(\tau), \tau \in [\tau_L, \tau_U]\}$ and $\{a_{2,i}(\tau), \tau \in [\tau_L, \tau_U]\}$ are Donsker by the Lipschitz continuity of $\beta(\cdot)$ implied by C3, and by using the fact that the Donsker Property is preserved under the Lipschitz transformation. It is not difficult to show that $\int_0^\infty w_1(b, s) \, dM_i^G(s)$ is Lipschitz in b. Hence the function class $\{b_i(\tau), \tau \in [\tau_L, \tau_U]\}$ is Donsker. The Donsker property is preserved under addition. As a result, $S_n\{\beta(\tau), \tau\}$ converges weakly to a mean zero Guassian process with covariance matrix $\Sigma_1(\tau', \tau) = E\{\xi_{1,i}(\tau') \, \xi_{1,i}^T(\tau)\}$ by (A.6) for the IPW estimator, and it converges weakly to a mean zero Guassian process with covariance matrix $\Sigma_2(\tau', \tau) = E\{\xi_{2,i}(\tau')\xi_{2,i}^T(\tau)\}$ by (A.7) for the AIPW estimator, for $\tau, \tau' \in [\tau_L, \tau_U]$.

Next, simple algebraic manipulations show that $S_n\{\hat{\beta}(\tau), \tau\} - S_n\{\beta(\tau), \tau\} = (I) + (II)$, where

$$(I) = n^{-1/2} \sum_{i=1}^n Z_i \vartheta_i \big(I[X_i \leq g\{Z_i \hat{\beta}(\tau)\}] - I[X_i \leq g\{Z_i \beta(\tau)\}]\big),$$

$$(II) = n^{-1/2} \sum_{i=1}^n Z_i(\hat{\vartheta}_i - \vartheta_i)\big(I[X_i \leq g\{Z_i \hat{\beta}(\tau)\}] - I[X_i \leq g\{Z_i \beta(\tau)\}]\big).$$

From Lemma 1 of Peng and Fine (2009) and the uniform consistency of $\hat{\beta}(\tau)$, it follows that the difference between (I) and $n^{1/2}[\mu\{\hat{\beta}(\tau), \tau\} - \mu\{\beta(\tau), \tau\}]$ converges to zero uniformly in $\tau \in [\tau_L, \tau_U]$ in probability. By the first order Taylor expansion of $\hat{\vartheta}_i$ around ϑ_i, (3.1), (3.2), (A.3) and applying Lemma 1 of Peng and Fine (2009), we can show that $(II) = o(1)$ uniformly in $\tau \in [\tau_L, \tau_U]$ in probability. Taylor expansions of $\mu(b, \tau)$ around $b = \beta(\tau)$, along with the fact that $\hat{\beta}(\tau)$ uniformly converges to $\beta(\tau)$, gives that

$$S_n\{\hat{\beta}(\tau), \tau\} - S_n\{\beta(\tau), \tau\} = [A\{\beta(\tau)\} + \epsilon_n(\tau)] \cdot n^{1/2}\{\hat{\beta}(\tau) - \beta(\tau)\} + o_p(1),$$

where $\sup_\tau \|\epsilon_n(\tau)\| \xrightarrow{P} 0$. Given $S_n\{\hat{\beta}(\tau), \tau\} = o_p(n^{-1/2})$, this further implies that

$$n^{1/2}\{\hat{\beta}(\tau) - \beta(\tau)\} = -[A\{\beta(\tau)\}]^{-1}S_n\{\beta(\tau), \tau\} + \epsilon_n^*(\tau),$$

where $\sup_\tau \|\epsilon_n^*(\tau)\| \xrightarrow{P} 0$. The asymptotic approximations (3.3) and (3.4) for the IPW estimator and the AIPW estimator follow from (A.6) and (A.7), respectively. Hence, $n^{1/2}\{\hat{\beta}(\tau) - \beta(\tau)\}$ converges weakly to a mean zero Guassian

process with covariance matrix $\Phi_1(\tau', \tau) = \left[A\{\beta(\tau')\}\right]^{-1} \Sigma_1(\tau', \tau) \left[A\{\beta(\tau)\}\right]^{-1}$ for the IPW estimator and $\Phi_2(\tau', \tau) = \left[A\{\beta(\tau')\}\right]^{-1} \Sigma_2(\tau', \tau)\left[A\{\beta(\tau)\}\right]^{-1}$ for the AIPW estimator.

Finally, we show that the AIPW estimator is more efficient than the IPW estimator by showing that $\Sigma_2(\tau', \tau) \leq \Sigma_1(\tau', \tau)$. Note that $a_{2,i}(\tau) = a_{1,i}(\tau) + e_i(\tau)$ and $\xi_{1,i}(\tau) = \xi_{2,i}(\tau) + \{c_i(\tau) - e_i(\tau)\}$, where $e_i(\tau)$ is defined in Section 3 just before Theorem 1. By (A.6) and (A.7), it suffices to show $E[\xi_{2,i}(\tau')\{c_i(\tau) - e_i(\tau)\}^T] = 0$.

Under MAR, R_i and J_i are conditionally independent given Q_i, we have

$$E\{a_{2,i}(\tau')c_i^T(\tau)\}$$
$$= E\left(\frac{R_i\delta_i\rho(W_i)}{\pi(Q_i)G(X_i)}Z_iI[X_i \leq g\{Z_i^T\beta(\tau')\}]\eta_i^T\right)w_2^T\{\beta(\tau)\}$$
$$- E\left(\frac{\{R_i - \pi(Q_i)\}\delta_i\rho(W_i)}{\pi(Q_i)G(X_i)}Z_iI[X_i \leq g\{Z_i^T\beta(\tau')\}]\eta_i^T\right)w_2^T\{\beta(\tau)\},$$

which equals zero by $E(\eta_i \mid Q_i) = 0$. By $E(\eta_i|Q_i) = 0$, we also have $E\{b_i(\tau')c_i^T(\tau)\} = 0$. Therefore $E\{\xi_{2,i}(\tau')c_i^T(\tau)\} = 0$. Similarly, $E\{\xi_{2,i}(\tau')e_i^T(\tau)\} = 0$. Hence $E\{\xi_{1,i}(\tau')\xi_{1,i}^T(\tau)\} = E\{\xi_{2,i}(\tau')\xi_{2,i}^T(\tau)\} + E[\{c_i(\tau') - e_i(\tau')\}\{c_i(\tau) - e_i(\tau)\}^T]$.

Proof of (4.1). Let $f_i = Z_i[I\{g^{-1}(T_i) \leq Z_i^T\beta(\tau)\}]\vartheta_{1,i}$. Under the MAR assumption, R_i and J_i are conditionally independent given Q_i, and we have

$$E\{a_{1,i}(\tau')b_i^T(\tau)\} = -E\left(f_i \int_0^\infty [w_1\{\beta(\tau), s\}]^T\lambda^G(s)Y_i(s)\,ds\right)$$
$$= -\int_0^\infty w_1\{\beta(\tau'), s\}[w_1\{\beta(\tau), s\}]^T y(s)\lambda^G(s)\,ds$$
$$= -E\{b_i(\tau')b_i^T(\tau)\},$$

$$E\{a_{1,i}(\tau')c_i^T(\tau)\} = E\left(\vartheta_{1,i}Z_iI[X_i \leq g\{Z_i^T\beta(\tau')\}]\eta_i^T\right)w_2^T(\beta(\tau))$$
$$= E\left(\frac{R_i\delta_i\rho(W_i)}{\pi(Q_i)G(X_i)}Z_iI[X_i \leq g\{Z_i^T\beta(\tau')\}]\eta_i^T\right)w_2^T(\beta(\tau))$$
$$= -w_2\{\beta(\tau')\}I_\psi^{-1}[w_2\{\beta(\tau)\}]^T,$$

where the last equation is obtained by the definition of η_i following (3.1). It is easy to see that $E\{b_i(\tau')c_i^T(\tau)\} = 0$ and $E\{c_i(\tau')c_i^T(\tau)\} = w_2\{\beta(\tau')\}I_\psi^{-1}[w_2\{\beta(\tau)\}]^T$ since $E(\eta_i\eta_i^T) = I_\psi^{-1}$. It follows that

$$E[\xi_{1,i}(\tau')\{\xi_{1,i}(\tau)\}^T]$$
$$= E[\{a_{1,i}(\tau') + b_i(\tau')\}\{a_{1,i}(\tau) + b_i(\tau)\}^T] - w_2\{\beta(\tau')\}I_\psi^{-1}[w_2\{\beta(\tau)\}]^T$$
$$= E\{a_{1,i}(\tau')a_{1,i}^T(\tau)\} - E\{b_i(\tau')b_i^T(\tau)\} - w_2\{\beta(\tau')\}I_\psi^{-1}[w_2\{\beta(\tau)\}]^T.$$

References

Bang, H. and Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics* **58**, 643-649.

Beaudry, M., Dufour, R. and Marcoux, S. (1995). Relation between infant feeding and infections during the first six months of life. *J. Pediatrics* **126**, 191–197.

Beran, R. (1981). Nonparametric Regression With Randomly Censored Survival Data. Technical report, University of California, Berkeley.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Dunn, D. T., Newell, M. L., Ades, A. E. and Peckham, C. S. (1992). Risk of human immunod-eficiency virus type 1 transmission through breastfeeding. *The Lancet* **340**, 585-588.

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

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

Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika* **82**, 821-834.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47**, 663-685.

Huang, Y. (2002). Calibration regression of censored lifetime medical cost. *J. Amer. Statist. Assoc.* **97**, 318-327.

Koenker, R. and Bassett, G. S. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276-1288.

Lu, W. and Liang, Y. (2008). Analysis of competing risks data with missing cause of failure under additive hazards model. *Statist. Sinica* **19**, 219-234.

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

Neocleous, T., Vanden Branden, K., and Portnoy, S. (2006). Correction to censored regression quantiles by S. Portnoy, **98** (2003), 1001-1012. *J. Amer. Statist. Assoc.* **101**, 860-861.

Peng, L. and Fine, J. P. (2009). Competing risks quantile regression. *J. Amer. Statist. Assoc.* **104**, 1440-1453.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103**, 637–649.

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *J. Amer. Statist. Assoc.* **86**, 770-778.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98**, 1001-1012.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Fluornoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure time in the presence of competing risks. *Biometrics* **34**, 541-554.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Thior, I., Lockman, S., Smeaton, L. M., Shapiro, R. L., Wester, C., Heymann, S. J., Gilbert, P. B., Stevens, L., Peter, T., Kim, S., van Widenfelt, E., Moffat, C., Ndase, P., Arimi, P., Kebaabetswe, P., Mazonde, P., Makhema, J., McIntosh, K., Novitsky, V., Lee, T.H., Marlink, R., Lagakos, S., Essex M. and the Mashi Study Team (2006). Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child HIV transmission in Botswana: a randomized trial: the Mashi Study. *J. Amer. Statist. Medical Assoc.* **296**, 794-805.

Tsiatis, A. A.(1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. USA* **72**, 20-22.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.

Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104**, 1117-1128.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178-184.

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

E-mail: yasun@uncc.edu

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA.

E-mail: hwang3@ncsu.edu

Department of Biostatistics, University of Washington, and Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

E-mail: pgilbert@scharp.org