Supplementary Material for ‘Semiparametric Marginal Regression Analysis for Clustered Competing Risks Data with Missing Cause of Failure’

Wenxian Zhou¹, Giorgos Bakoyannis¹,* , Ying Zhang², Constantin T. Yiannoutsos¹

¹Department of Biostatistics, Indiana University, Indianapolis, Indiana, U.S.A.
²Department of Biostatistics, University of Nebraska Medical Center, Omaha, Nebraska, U.S.A.

*gbakogia@iu.edu

1. R Code

Our methodology can be easily implemented using standard software that allows for weights. In this Appendix we illustrate the use of the coxph function in the R package survival. Let data be a data set with the variables: cluster id clusterid, observed time x, cause of failure c, missingness indicator r, a covariate of interest z, an auxiliary variable a, and cluster size clustersize.

For illustration purposes, we analyze the cause specific hazard for cause 1 here. The first stage of the analysis is to estimate γ₀ through generalized estimating equations (GEE) for logistic regression using the observations with an observed cause of failure. This can be done using the following code.

```
cause <- 1

data$include <- 1*(data$r==1 & data$c>0)/data$clustersize
```
data$outcome <- 1*(data$c==cause)
model <- geeglm(outcome ~ x + z + a, family = "binomial",
                data = data, id = clusterid, weight = include, corstr = "independence")

Then, one needs to calculate the estimated probability of the cause of failure $\pi_1(W_{ij}, \hat{\gamma}_n)$.

data$yhat <- predict(model, data, type = "response")

The second stage of the analysis is to maximize the weighted partial pseudolikelihood. This can be implemented with the `coxph` function using the weight option and some simple data management. The data management steps are used to “remove” the weights from the risk sets for the observations with a missing cause of failure.

data$d <- data$r*(data$c == cause) + (1-data$r)*(data$yhat > 0)
data$weight <- data$r + (1-data$r)*data$yhat
data$weight <- data$weight + (data$weight == 0)
dt0 <- data[data$r==0, ]
dt0$weight <- 1 - dt0$weight
dt0$d <- 0
data1 <- rbind(data, dt0)
data1$weight <- data1$weight/data1$clustersize

Then the point estimates of the regression coefficient can be calculated using the augmented dataset `data1` as follows.

    mod <- coxph(Surv(x, d) ~ z, weight = weight, data = data1)
    beta1 <- coef(mod)

The cause-specific baseline cumulative hazard function can be calculated with the `basehaz` function as follows.
Finally, we can get the estimated baseline cumulative incidence function using equation (1) in the main manuscript, based on the estimated regression coefficients and baseline cumulative hazard functions for all causes of failure. For example, in the case of two causes of failure, the baseline cumulative incidence function can be estimated as follows.

```r
Haz1 <- H1$hazard
Haz2 <- H2$hazard
S <- exp(-Haz1 - Haz2)
S.minus <- c(1, S[1: (length(S) - 1)])
Haz1.minus <- c(0, Haz1[1: (length(Haz1) - 1)])
CIF1 <- cumsum(S.minus * (Haz1 - Haz1.minus))
```

The computation of our closed-form standard error estimators is more involved. However, these standard error estimators are readily available in our R functions that can be found at [https://github.com/wz11/ClusteredMPPLE](https://github.com/wz11/ClusteredMPPLE). We must note that the uncertainty in the estimate $\hat{\gamma}_n$ that appears in the weights is properly incorporated into our standard error estimators. This is due to the fact that our standard error estimators involve appropriately the empirical versions of the influence functions $\omega_{ij}$ for $\hat{\gamma}_n$ (see Section 2.3 of the main manuscript and regularity condition C4 in Section 2 of this Supplementary Material). Alternatively, one can use the nonparametric cluster bootstrap approach for standard error estimation (Bakoyannis, 2021).

2. **Asymptotic Theory Proofs**

The asymptotic properties of the proposed estimators are justified based on empirical process theory (Kosorok, 2008; van der Vaart and Wellner, 1996). We use the following standard empirical
process notations throughout the Appendix B. For any measurable functions \( f : D \rightarrow \mathbb{R} \),

\[
\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(D_i), \quad \text{and} \quad P f = \int_{D} f dP = E f(D_1).
\]

Also, let \( K_0 \) denote a generic constant that may change from place to place. For simplicity, we will omit the subindex \( i \), indicating a specific cluster, from expectations. By the i.i.d. assumption across clusters, the expectations correspond to expectations of (functions of) random variables from an arbitrary cluster. In the proofs we will consider an arbitrary cause of failure \( l \in \{1, \ldots, k\} \).

Before stating the regularity conditions, we define the negative of the Hessian matrix of the expected partial pseudolikelihood as

\[
H_l(\beta) = E \left\{ \frac{1}{M} \sum_{j=1}^{M} \int_0^\tau V(t, \beta) d\tilde{N}_{jl}(t; \gamma_0) \right\},
\]

where

\[
V(t, \beta) = \frac{E\left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta^T Z_j) Z_j^{\otimes 2} \right\}}{E\left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta^T Z_j) \right\}} - \left[ \frac{E\left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta^T Z_j) Z_j \right\}}{E\left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta^T Z_j) \right\}} \right]^{\otimes 2}.
\]

In this work we assume the following regularity conditions.

**C1.** \( \Lambda_{0,l}(t) \) is a non-decreasing continuous function with \( \Lambda_{0,l}(\tau) < \infty \), for \( l = 1, \ldots, k \) and \( E\{Y(\tau) \mid Z, M\} > 0 \) almost surely.

**C2.** The true regression coefficients \( \beta_{0,l} \in \mathcal{B}_l \subset \mathbb{R}^p \), where \( \mathcal{B}_l \) is bounded and convex set for \( l = 1, \ldots, k \) and \( \beta_{0,l} \) is in interior of \( \mathcal{B}_l \).

**C3.** The inverse of the link function \( g \) for the marginal probability model of the cause of failure \( \pi_l(W_{ij}, \gamma_0), \ l = 1, \ldots, k \), has continuous derivative \( \dot{g} \) with respect to \( \gamma_0 \) on compact sets. The parameter space \( \Gamma \) of \( \gamma_0 \) is a bounded subset of \( \mathbb{R}^{p'} \).

**C4.** The estimating function for the model of the cause of failure is Lipschitz continuous in \( \gamma \), and the estimator \( \hat{\gamma}_n \) is strongly consistent and asymptotically linear, i.e. \( \sqrt{n}(\hat{\gamma}_n - \gamma_0) = \)
\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j=1}^{M_i} \omega_{ij} + \alpha_p(1), \text{ where } \omega_{ij} \text{ is the influence function of } j \text{th subject in } i \text{th cluster, satisfying } E(\omega_{ij}) = 0 \text{ and } E\|\omega_{ij}\|^2 < \infty. \]

C5. The covariates of interest \( Z \), the auxiliary covariates \( A \), and the cluster size \( M \) are bounded, in the sense that there exist constants \( K \in \mathbb{R}^+ \) and \( m_0 \in \mathbb{N}^+ \) such that \( P(\|Z\| \vee \|A\| \leq K) = 1 \) and \( P(M \leq m_0) = 1 \).

C6. The Hessian matrix of the expected partial pseudolikelihood \(-\mathbf{H}_l(\beta)\) is negative definite on \( B_l \) for all \( l = 1, \ldots, k \).

C7. The observations \( W_{ij}, \Delta_{ij}, \epsilon_{ij}, \) and \( R_{ij} \) are identically distributed conditionally on cluster size \( M_i \), in the sense that \( E(W_{ij}|M_i) = E(W_{i1}|M_i), E(\Delta_{ij}|M_i) = E(\Delta_{i1}|M_i), E(\epsilon_{ij}|M_i) = E(\epsilon_{i1}|M_i), \) and \( E(R_{ij}|M_i) = E(R_{i1}|M_i) \), for all \( i = 1, \ldots, n, j = 1, \ldots, M_i, \) and \( l = 1, \ldots, k \).

Regularity conditions C3 and C4 are satisfied when the marginal model for \( \pi_l(W_{ij}, \gamma_0) \) is correctly specified with a standard link function, such as the logit link, and parameters estimated through generalized estimating equations under a working independence assumption. The assumptions on the parametric models for \( \pi_l(W_{ij}, \gamma_0) \) can be evaluated using the cumulative residual processes \( E\left[ \frac{1}{M} \sum_{j=1}^{M_i} R_j \{N_{lj}(t) - \pi_l(W_{lj}, \gamma_0)N_j(t)\} \right], l = 1, \ldots, k - 1, t \in [0, \tau], \) which can be estimated by \( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j=1}^{M_i} R_{ij} \{N_{lj}(t) - \pi_l(W_{ij}, \hat{\gamma}_n)N_{ij}(t)\}. \) If the model is correctly specified, the cumulative residual process is equal to 0 for \( t \in [0, \tau] \). A formal goodness of fit test can be conducted using the simulation-based approach by Pan and Lin (2005). A graphical evaluation of goodness of fit can also be performed by plotting the observed residual process and the 95% simultaneous confidence band around the line \( f(t) = 0, t \in [0, \tau] \) (Bakoyannis and others, 2019, 2020).

Before presenting the proofs of the theorems stated in the main manuscript, we provide explicit formulas for the elements of the influence functions. First, for Theorem 2.2, define

\[ \psi_{ijl} = \mathbf{H}_l^{-1}(\beta_{0,l}) \int_0^\tau \{Z_{ij} - E(t, \beta_{0,l})\} d\tilde{M}_{ijl}(t; \beta_{0,l}, \gamma_0), \]
where $H_t(\beta_{0,t})$ is the negative of the Hessian matrix of the expected partial pseudolikelihood and

$$E(t, \beta_{0,t}) = \frac{E\{\frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta_{0,t}^T Z_j) Z_j\}}{E\{\frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta_{0,t}^T Z_j)\}},$$

and $\tilde{M}_{ijl}(t; \beta_{0,t}, \gamma_0) = \tilde{N}_{ijl}(t; \gamma_0) - \int_0^t Y_{ij}(u) \exp(\beta_{0,t}^T Z_{ij}) d\Lambda_{0,t}(u),$ where

$$\Lambda_{0,t}(t) = \int_0^t E\{\frac{1}{M} \sum_{j=1}^{M} d\tilde{N}_{ijl}(u; \gamma_0)\}/E\{\frac{1}{M} \sum_{j=1}^{M} Y_j(u) \exp(\beta_{0,t}^T Z_j)\}. $$

The empirical version of $\psi_{ijl}$ is

$$\hat{\psi}_{ijl} = H_{n,l}^{-1}(\hat{\beta}_{n,l}; \hat{\gamma}_n) \int_0^T \{Z_{ij} - E_n(t, \hat{\beta}_{n,l})\} d\tilde{M}_{ijl}(t; \hat{\beta}_{n,l}, \hat{\gamma}_n),$$

where

$$H_{n,l}(\hat{\beta}_{n,l}; \hat{\gamma}_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_i} \int_0^T V_{n,l}(t, \hat{\beta}_{n,l}) d\tilde{N}_{ijl}(t; \hat{\gamma}_n),$$

$$V_{n,l}(t, \hat{\beta}_{n,l}) = \frac{\sum_{p=1}^{n} \frac{1}{M_p} \sum_{q=1}^{M_p} Y_{pq}(t) \exp(\hat{\beta}_{n,l}^T Z_{pq}) Z_{pq}^T \exp(\hat{\beta}_{n,l}^T Z_{pq}) Z_{pq}}{\sum_{p=1}^{n} \frac{1}{M_p} \sum_{q=1}^{M_p} Y_{pq}(t) \exp(\hat{\beta}_{n,l}^T Z_{pq})},$$

$$E_n(t, \hat{\beta}_{n,l}) = \frac{\sum_{p=1}^{n} \frac{1}{M_p} \sum_{q=1}^{M_p} Y_{pq}(t) \exp(\hat{\beta}_{n,l}^T Z_{pq}) Z_{pq}}{\sum_{p=1}^{n} \frac{1}{M_p} \sum_{q=1}^{M_p} Y_{pq}(t) \exp(\hat{\beta}_{n,l}^T Z_{pq})}.$$ 

and

$$\tilde{M}_{ijl}(t; \hat{\beta}_{n,l}, \hat{\gamma}_n) = \tilde{N}_{ijl}(t; \hat{\gamma}_n) - \int_0^t Y_{ij}(u) \exp(\hat{\beta}_{n,l}^T Z_{ij}) d\hat{\Lambda}_{n,l}(u).$$

Also, define the non-random quantity

$$R_t = H_t^{-1}(\beta_{0,t}) E \left[ \frac{1}{M} \sum_{j=1}^{M} (1 - R_j) \int_0^T \{ Z_j - E(t, \beta_{0,t}) \} dN_j(t) \hat{\pi}_t(W_j, \gamma_0)^T \right],$$

where $\hat{\pi}_t(W_j, \gamma) = \partial\{\pi_t(W_j, \gamma)\}(\partial \gamma)^{-1}|_{\gamma=\gamma_0}$. The empirical version of $R_t$ is

$$\hat{R}_t = H_{n,l}^{-1}(\hat{\beta}_{n,l}; \hat{\gamma}_n) \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} (1 - R_{ij}) \int_0^T \{ Z_{ij} - E_n(t, \hat{\beta}_{n,l}) \} dN_{ij}(t) \hat{\pi}_t(W_{ij}, \hat{\gamma}_n)^T \right].$$

Next, for Theorem 2.3, define

$$\phi_{ijl}(t) = \int_0^t \frac{d\tilde{M}_{ijl}(s; \beta_{0,j}, \gamma_0)}{E\{\frac{1}{M} \sum_{j=1}^{M} Y_j(s) \exp(\beta_{0,t}^T Z_j)\}} - (\psi_{ijl} + R_t(\omega_{ij}))^T \int_0^t E(s, \beta_{0,t}) d\Lambda_{0,t}(s).$$
and the non-random function

\[ R^*_t(t) = E \left[ \frac{1}{M} \sum_{j=1}^{M} (1 - R_j) \pi_j(W_j, \gamma_0) \int_0^t \frac{dN_j(s)}{E(\frac{1}{M} \sum_{j=1}^{M} Y_j(s) \exp(\beta_{n,l}^T Z_j))} \right]^T. \]

The empirical versions of the latter functions are

\[ \dot{\phi}_{ijl}(t) = \int_0^t \frac{dM_{ijl}(t; \beta_{n,l}, \gamma_n)}{\frac{1}{M} \sum_{i=1}^{n} \sum_{j=1}^{M} Y_{ij}(s) \exp(\beta_{n,l}^T Z_{ij})} - \left( \dot{\psi}_{ijl} + \dot{R}_t(\dot{\omega}_{ij}) \right) \int_0^t E_n(s, \beta_{n,l}) d\hat{\Lambda}_{n,l}(s), \]

and

\[ \dot{R}^*_t(t) = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{1}{M} \sum_{j=1}^{M} (1 - R_{ij}) \pi_{ij}(W_{ij}, \gamma_n) \int_0^t \frac{dN_{ij}(s)}{E(\frac{1}{M} \sum_{j=1}^{M} Y_{ij}(s) \exp(\beta_{n,l}^T Z_{ij}))} \right]^T, \]

respectively.

Finally, for Theorem 2.4, define the influence function

\[ \phi_{ijl}^F(t; z_0) = \int_0^t \exp \left\{ - \sum_{l=1}^{k} \Lambda_{0,l}(s; z_0) \right\} d\phi_{ijl}(s; z_0) \]

\[ - \int_0^t \left\{ \sum_{l=1}^{k} \phi_{ijl}^A(s; z_0) \right\} \exp \left\{ - \sum_{l=1}^{k} \Lambda_{0,l}(s; z_0) \right\} d\Lambda_{0,l}(s; z_0), \]

where

\[ \phi_{ijl}^A(t; z_0) = \{ z_0^T (\psi_{ijl} + R_t(\omega_{ij}) \Lambda_{n,l}(t) + \phi_{ijl}(t) + R^*_t(t) \omega_{ij}) \} \exp(\beta_{n,l}^T z_0). \]

The empirical version of \( \phi_{ijl}^F(t; z_0) \) is

\[ \dot{\phi}_{ijl}^F(t; z_0) = \int_0^t \exp \left\{ - \sum_{l=1}^{k} \hat{\Lambda}_{n,l}(s; z_0) \right\} d\hat{\phi}_{ijl}^A(s; z_0) \]

\[ - \int_0^t \left\{ \sum_{l=1}^{k} \hat{\phi}_{ijl}^A(s; z_0) \right\} \exp \left\{ - \sum_{l=1}^{k} \hat{\Lambda}_{n,l}(s; z_0) \right\} d\hat{\Lambda}_{n,l}(s; z_0), \]

where

\[ \hat{\phi}_{ijl}^A(t; z_0) = \{ z_0^T (\hat{\psi}_{ijl} + \hat{R}_t(\hat{\omega}_{ij}) \hat{\Lambda}_{n,l}(t) + \hat{\phi}_{ijl}(t) + \hat{R}^*_t(t) \hat{\omega}_{ij}) \} \exp(\hat{\beta}_{n,l}^T z_0). \]

### 2.1 Proof of Theorem 2.1

To prove the consistency of \( \hat{\beta}_{n,l} \) we use the consistency conditions for general Z-estimators (Kosorok, 2008). In light of condition C5, the population analog of the partial pseudoscore func-
tion is
\[ G_l(\beta) = P \left[ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j) \int_0^T \{ Z_j - E(t, \beta) \} d\hat{N}_{jl}(t; \gamma_0) \right]. \]

By conditions C5 and C7 we have
\[ G_l(\beta) = E \left[ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j) E \left[ \int_0^T \{ Z_j - E(t, \beta) \} d\hat{N}_{jl}(t; \gamma_0) \mid M \right] \right] \]
\[ = E \left( E \left[ \int_0^T \{ Z_j - E(t, \beta) \} d\hat{N}_{jl}(t; \gamma_0) \mid M \right] \right) \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j) \]
\[ = E \left[ \int_0^T \{ Z_j - E(t, \beta) \} d\hat{N}_{jl}(t; \gamma_0) \right]. \tag{2.1} \]

for any \( j = 1, \ldots, M \). Now, by the definition of \( \hat{N}_{ijl}(t; \gamma_0) \), the assumption of correct specification
of the marginal model \( \pi_l(W_{ij}, \gamma_0) \), and the fact that \( \Delta_{ijl} N_{ij}(t) = N_{ijl}(t) \), it follows that
\[ G_l(\beta) = E \left[ \int_0^T \{ Z_j - E(t, \beta) \} dN_{jl}(t) \right], \]
for any \( j = 1, \ldots, M \). Letting \( \Lambda_{0,l}(t) = \int_0^t \lambda_{0,l}(s) ds \), \( G_l(\beta_{0,l}) \) can be expressed as
\[ G_l(\beta_{0,l}) = E \left[ \int_0^T \{ Z_j - E(t, \beta_{0,l}) \} d\{ N_{jl}(t) - Y_j(t) \exp(\beta_{0,l}^T Z_j) d\Lambda_{0,l}(t) \} \right] \]
\[ + E \left[ \int_0^T \{ Z_j - E(t, \beta_{0,l}) \} Y_j(t) \exp(\beta_{0,l}^T Z_j) d\Lambda_{0,l}(t) \right]. \tag{2.2} \]

Note that, by the independent censoring assumption conditionally on \( Z_{ij} \) and under the assumed
marginal proportional cause-specific hazards model, we have
\[ E\{ N_{jl}(t) \mid Z_j \} = P(X_j \leq t, C_j = l \mid Z_j) \]
\[ = P(U_j > T_j, T_j \leq t, C_j = l \mid Z_j) \]
\[ = \int_0^t P(U_j > s \mid Z_j) dF_l(s; Z_j) \]
\[ = \int_0^t P(U_j > s \mid Z_j) P(T_j > s \mid Z_j) \frac{dF_l(s; Z_j)}{P(T_j > s \mid Z_j)} \]
\[ = \int_0^t P(X_j \geq s \mid Z_j) \exp(\beta_{0,l}^T Z_j) d\Lambda_{0,l}(s) \]
\[ = E \left\{ \int_0^t Y_j(s) \exp(\beta_{0,l}^T Z_j) d\Lambda_{0,l}(s) \mid Z_j \right\} \] (2.3)
for any \( j = 1, \ldots, M \). Also, by conditions C5 and C7 and similar calculations to those in (2.1), we have that
\[
E(t, \beta_{0,l}) = \frac{E \left\{ Y_j(t) \exp(\beta^T Z_j) \right\}}{E \left\{ Y_j(t) \exp(\beta^T Z_j) \right\}}.
\]
This fact, equality (2.3), and some algebra lead to the conclusion that
\[
E \left[ \int_0^\tau \{ Z_j - E(t, \beta_{0,l}) \} N_{jl}(t) - Y_j(t) \exp(\beta_{0,l}^T Z_j) d\Lambda_{0,l}(t) \right] = 0.
\]
and
\[
E \left[ \int_0^\tau \{ Z_j - E(t, \beta_{0,l}) \} Y_j(t) \exp(\beta_{0,l}^T Z_j) d\Lambda_{0,l}(t) \right] = 0.
\]
Therefore, by (2.2), it follows that \( G_l(\beta_{0,l}) = 0 \). Condition C6 implies that \( \beta_{0,l} \) is the unique root of \( G_l(\beta) = 0, \beta \in B_l \). To complete the proof of the strong consistency of \( \hat{\beta}_{n,l} \), it remains to show that
\[
\sup_{\beta \in B_l} \| G_{n,l}(\beta, \hat{\gamma}_n) - G_l(\beta) \| \to_{as} 0.
\]
Using empirical process notation, the empirical version of the pseudoscore function defined in Section 2.2 can be written as
\[
G_{n,l}(\beta, \hat{\gamma}_n) = P_n \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \{ Z_j - E_n(t, \beta) \} d\tilde{N}_{jl}(t; \hat{\gamma}_n) \right\}.
\]
The difference between the empirical partial pseudoscore function and expected partial pseudoscore function can be decomposed as
\[
G_{n,l}(\beta, \hat{\gamma}_n) - G_l(\beta) = A_{n,l} + B_{n,l} - C_{n,l}(\beta) - D_{n,l}(\beta) - E_{n,l}(\beta),
\]
where
\[
A_{n,l} = P_n \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) Z_j \{ d\tilde{N}_{jl}(t; \hat{\gamma}_n) - d\tilde{N}_{jl}(t; \gamma_0) \} \right\},
\]
\[
B_{n,l} = (P_n - P) \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) Z_j d\tilde{N}_{jl}(t; \gamma_0) \right\},
\]
\[ C_{n,l}(\beta) = \mathbb{P}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{E}_n(t, \beta) \{ d\tilde{N}_j(t; \gamma_n) - d\tilde{N}_j(t; \gamma_0) \} \right], \]

\[ D_{n,l}(\beta) = \mathbb{P}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \{ \mathcal{E}_n(t, \beta) - \mathcal{E}(t, \beta) \} d\tilde{N}_j(t; \gamma_0) \right], \]

and

\[ E_{n,l}(\beta) = (\mathbb{P}_n - P) \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{E}(t, \beta) d\tilde{N}_j(t; \gamma_0) \right\}. \]

By conditions C3-C5,

\[ \| A_{n,l} \| = \mathbb{E}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{Z}_j \{ d\tilde{N}_j(t; \gamma_n) - d\tilde{N}_j(t; \gamma_0) \} \right] \]

\[ = \mathbb{E}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{Z}_j (1 - R_j) \{ \pi_l(W_j, \gamma_n) - \pi_l(W_j, \gamma_0) \} dN_j(t) \right] \]

\[ \leq K_0 \| \gamma_n - \gamma_0 \| \mathbb{E}_n \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{Z}_j (1 - R_j) dN_j(t) \right\} \]

\[ \to_{as} 0. \]

By the strong law of large numbers and condition C5,

\[ B_{n,l} = (\mathbb{P}_n - P) \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{Z}_j d\tilde{N}_j(t; \gamma_0) \right\} \to_{as} 0. \]

By conditions C2-C5,

\[ \sup_{\beta \in \mathcal{B}_l} \| C_{n,l}(\beta) \| = \sup_{\beta \in \mathcal{B}_l} \mathbb{E}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{E}_n(t, \beta) \{ d\tilde{N}_j(t; \gamma_n) - d\tilde{N}_j(t; \gamma_0) \} \right] \]

\[ \leq \sup_{\beta \in \mathcal{B}_l} \mathbb{E}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{E}_n(t, \beta) \{ \pi_l(W_j, \gamma_n) - \pi_l(W_j, \gamma_0) \} dN_j(t) \right] \]

\[ \leq K_0 \| \gamma_n - \gamma_0 \| \sup_{\beta \in \mathcal{B}_l} \mathbb{E}_n \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \mathcal{E}_n(t, \beta) dN_j(t) \right\} \]

\[ \to_{as} 0. \]

For \( D_{n,l}(\beta) \), the classes of functions \( \{ M^{-1} \sum_{j=1}^{m_0} I(M \geq j) Y_j(t) \exp(\beta^T \mathcal{Z}_j), t \in [0, \tau], \beta \in \mathcal{B}_l \} \)

and \( \{ M^{-1} \sum_{j=1}^{m_0} I(M \geq j) Y_j(t) \exp(\beta^T \mathcal{Z}_j) \mathcal{Z}_j, t \in [0, \tau], \beta \in \mathcal{B}_l \} \) are Donsker by condition C5.
and, thus, also Glivenko-Cantelli. Therefore, using conditions C2 and C5,

$$\sup_{\beta \in B_i} \|D_{n,t}(\beta)\| = \sup_{\beta \in B_i} \left\| \mathbb{P}_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^T I(M \geq j) \{ \mathbf{E}_n(t, \beta) - \mathbf{E}(t, \beta) \} d\tilde{N}_j(t; \gamma_0) \right] \right\|$$

$$\leq \sup_{t \in [0, \tau], \beta \in B_i} \|\mathbf{E}_n(t, \beta) - \mathbf{E}(t, \beta)\| \sup_{\beta \in B_i} \mathbb{P}_n \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^T I(M \geq j) d\tilde{N}_j(t; \gamma_0) \right\}$$

$$\leq K_0 \sup_{t \in [0, \tau], \beta \in B_i} \|\mathbf{E}_n(t, \beta) - \mathbf{E}(t, \beta)\| \rightarrow_{a.s.} 0.$$ 

For \( \mathbf{E}_{n,t}(\beta_i) \), consider the class of functions

$$\mathcal{L}_t^{(p)} = \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^T I(M \geq j) \mathbf{E}^{(p)}(t, \beta) d\tilde{N}_j(t; \gamma_0), \beta \in B_i \right\}$$

$$= \left\{ \frac{1}{M} \sum_{j=1}^{m_0} \{ R_j \Delta_{jt} + (1 - R_j) \pi_l(W_j, \gamma_0) \} \int_0^T I(M \geq j) \mathbf{E}^{(p)}(t, \beta) dN_j(t), \beta \in B_i \right\},$$

where

$$\mathbf{E}^{(p)}(t, \beta) = \frac{E \left\{ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j) Y_j(t) \exp(\beta^T \mathbf{Z}_j) \mathbf{Z}^{(p)}_j \right\}}{E \left\{ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j) Y_j(t) \exp(\beta^T \mathbf{Z}_j) \right\}},$$

with \( \mathbf{Z}^{(p)}_j \) being the \( p \)th component of \( \mathbf{Z}_j \). For an arbitrary probability measure \( Q \), define the norm

$$\|f\|_{Q,2} = (\int f^2 dQ)^{1/2}.$$ Now, for any finitely discrete probability measure \( Q \) and \( \forall \beta_1, \beta_2 \in B_i \) and \( f_{\beta_1}^{(p)}, f_{\beta_2}^{(p)} \in \mathcal{L}_t^{(p)} \)

$$\left\| f_{\beta_1}^{(p)} - f_{\beta_2}^{(p)} \right\|_{Q,2} \leq \left\| \frac{1}{M} \sum_{j=1}^{m_0} \{ R_j \Delta_{jt} + (1 - R_j) \pi_l(W_j, \gamma_0) \} \int_0^T I(M \geq j) \mathbf{E}^{(p)}(t, \beta_1) - \mathbf{E}^{(p)}(t, \beta_2) dN_j(t) \right\|_{Q,2}$$

$$\leq \left\| \frac{1}{M} \sum_{j=1}^{m_0} \int_0^T \mathbf{E}^{(p)}(t, \beta_1) - \mathbf{E}^{(p)}(t, \beta_2) dN_j(t) \right\|_{Q,2}$$

$$\leq K_0 \| \beta_1 - \beta_2 \|,$$

by the Lipschitz continuity of \( \mathbf{E}^{(p)}(t, \beta) \) and condition C5. Then for \( \forall \beta \in B_i \) there exists a \( \beta_i, i = 1, \ldots, N(\epsilon, \mathcal{B}_i, \| \cdot \|) \) such that \( \| \beta_i - \beta \| < \epsilon \). Therefore, \( \forall f_{\beta}^{(p)} \in \mathcal{L}_t^{(p)} \) there exists a \( f_{\beta_i}^{(p)} \) such
that
\[ \| f^{(p)}_{\beta_i} - f^{(p)}_{\beta_i} \|_{Q,2} \leq K_0 \| \beta_i - \beta \| = \epsilon'. \]

Therefore, \( \mathcal{L}_t^{(p)} \) can be covered by \( N(\epsilon, B_t, \| \cdot \|) L_2(Q) \) \( \epsilon' \)-balls centered at \( f^{(p)}_{\beta_i} \). Thus, the covering number for \( \mathcal{L}_t^{(p)} \) is \( N(\epsilon', \mathcal{L}_t^{(p)}, L_2(Q)) \leq N(\epsilon, B_t, \| \cdot \|) \), where \( B_t \) is a Donsker class as a consequence of condition C2. In addition, using similar arguments to those in page 142 in Kosorok (2008), the class \( \mathcal{L}_t^{(p)} \) is pointwise measurable. Consequently, \( \mathcal{L}_t^{(p)} \) is Donsker and, thus, also Glivenko-Cantelli. Therefore, \( \sup_{\beta \in B_t} \| E_{n,t}(\beta) \| \to a.s. \) 0 and, thus,
\[
\sup_{\beta \in B_t} \| G_{n,t}(\beta, \gamma_n) - G_t(\beta) \| \to a.s. \) 0.
\]
This concludes the proof that \( \| \hat{\beta}_{n,t} - \beta_{0,t} \| \to a.s. \) 0.

Next, we prove the uniform consistency of \( \hat{\Lambda}_1(t) \). We define the supremum norm \( \| f(t) \|_\infty = \sup_{t \in [0, T]} |f(t)| \). Equality (2.3), conditions C5 and C7, and calculations similar to those in (2.1) lead to the conclusion that
\[
\Lambda_{0,t}(t) = \int_0^t \frac{P\{ \frac{1}{M} \sum_{j=1}^M d\tilde{N}_j(t; \gamma_0) \}}{P\{ \frac{1}{M} \sum_{j=1}^M \gamma_j(t) \exp(\beta_{0,t}^T Z_j) \}} dt.
\]
Now, after some algebra, we have that
\[
\hat{\Lambda}_{n,t}(t) - \Lambda_{0,t}(t) = A_{n,t}^*(t) + B_{n,t}^*(t),
\]
where
\[
A_{n,t}^*(t) = \mathbb{P}_n \left[ \frac{1}{M} \sum_{j=1}^M (1 - R_j) \{ \pi_i(W_j, \gamma_n) - \pi_i(W_j, \gamma_0) \} \int_0^t \frac{d\tilde{N}_j(t)}{\mathbb{P}_n \{ \frac{1}{M} \sum_{j=1}^M \gamma_j(t) \exp(\beta_{n,t}^T Z_j) \}} \right],
\]
\[
B_{n,t}^*(t) = \left[ \int_0^t \frac{\mathbb{P}_n \{ \frac{1}{M} \sum_{j=1}^M d\tilde{N}_j(t; \gamma_0) \}}{\mathbb{P}_n \{ \frac{1}{M} \sum_{j=1}^M \gamma_j(t) \exp(\beta_{n,t}^T Z_j) \}} - \int_0^t \frac{P\{ \frac{1}{M} \sum_{j=1}^M d\tilde{N}_j(t; \gamma_0) \}}{P\{ \frac{1}{M} \sum_{j=1}^M \gamma_j(t) \exp(\beta_{n,t}^T Z_j) \}} dt \right].
\]
By conditions C3, C4 and C5, \( \| A_{n,t}^*(\cdot) \|_\infty \to a.s. \) 0. By a similar expansion and arguments in Kosorok (2008), \( \| B_{n,t}^*(\cdot) \|_\infty \to a.s. \) 0. Therefore, \( \| \hat{\Lambda}_{n,t} - \Lambda_{0,t} \|_\infty \to a.s. \) 0.
2.2 Proof of Theorem 2.2

The estimator $\hat{\beta}_{n,l}$ satisfies

$$0 = \sqrt{n}G_{n,l}(\hat{\beta}_{n,l}, \hat{\gamma}_n)$$

$$= \sqrt{n}\{G_{n,l}(\hat{\beta}_{n,l}, \hat{\gamma}_n) - G_{n,l}(\hat{\beta}_{n,l}, \gamma_0)\} + \sqrt{n}G_{n,l}(\hat{\beta}_{n,l}, \gamma_0).$$

The first term in the right side of (2.5) can be expressed as

$$\sqrt{n}\{G_{n,l}(\hat{\beta}_{n,l}, \hat{\gamma}_n) - G_{n,l}(\hat{\beta}_{n,l}, \gamma_0)\} = A'_{n,l} - B'_{n,l} - C'_{n,l} + D'_{n,l},$$

where

$$A'_{n,l} = \sqrt{n}(P_n - P) \left[ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j)Z_jN_j(\tau)(1 - R_j)\{\pi_l(W_j, \hat{\gamma}_n) - \pi_l(W_j, \gamma_0)\} \right],$$

$$B'_{n,l} = \sqrt{n}(P_n - P) \left[ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j)(1 - R_j)\{\pi_l(W_j, \hat{\gamma}_n) - \pi_l(W_j, \gamma_0)\} \int_0^{\tau} E(t, \beta_{0,l})dN_j(t) \right],$$

$$C'_{n,l} = \sqrt{n}(P_n - P) \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^{\tau} I(M \geq j)\{E_n(t, \beta_l) - E(t, \beta_{0,l})\}d\{\hat{\gamma}_n(t) - \hat{\gamma}_0(t)\} \right],$$

and

$$D'_{n,l} = P \left[ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j)(1 - R_j) \int_0^{\tau} \{Z_j - E_n(t, \beta_{0,l})\} dN_j(t)\hat{\gamma}_l(W_j, \gamma_0)^T \right] \times \sqrt{n}(\hat{\gamma}_n - \gamma_0).$$

By conditions C3, C5 and the continuous mapping theorem it follows that $A'_{n,l} \xrightarrow{p} 0$ and $B'_{n,l} \xrightarrow{p} 0$. Using Lemma 4.2 of Kosorok (2008), it follows that $C'_{n,l} \xrightarrow{p} 0$. Therefore, the first term in the right side of (2.5) is

$$\sqrt{n}\{G_{n,l}(\hat{\beta}_{n,l}, \hat{\gamma}_n) - G_{n,l}(\hat{\beta}_{n,l}, \gamma_0)\} = A'_{n,l} - B'_{n,l} - C'_{n,l} + D'_{n,l} = D'_{n,l} + o_p(1).$$

The second term in the right side of (2.5) can be expressed as

$$\sqrt{n}G_{n,l}(\hat{\beta}_{n,l}, \gamma_0) = \sqrt{n}G_{n,l}(\beta_{0,l}, \gamma_0) - H_l(\beta_{0,l})\sqrt{n}(\hat{\beta}_{n,l} - \beta_{0,l}) + o_p(1 + \sqrt{n}\|\hat{\beta}_{n,l} - \beta_{0,l}\|).$$

By condition C6, $H_l(\beta_{0,l})$ is invertible and, thus, there exists a constant $K_0 > 0$ such that for any $\beta_l \in B_{l}$, $\|H_l(\beta_{0,l})(\beta_l - \beta_{0,l})\| \geq K_0\|\beta_l - \beta_{0,l}\|$. Therefore, by Taylor expansion around $\beta_{0,l},$
\[
\|G_{n,t}(\beta, \gamma_0) - G_{t}(\beta_{0,t}, \gamma_0)\| \geq K_0\|\beta_t - \beta_{0,t}\| + o_p(\|\beta_t - \beta_{0,t}\|).
\]
Now,
\[
\sqrt{n}\{G_{n,t}(\hat{\beta}_{n,t}, \gamma_0) - G_{t}(\beta_{0,t}, \gamma_0)\}
\]
\[
= \sqrt{n}(P_n - P) \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \{Z_j - E(t, \beta_{0,t})\} d\hat{N}_{jl}(t; \gamma_0) \right]
\]
\[
+ o_p(1 + \sqrt{n}\|\beta_{n,t} - \beta_{0,t}\|) + o_p(1)
\]
\[
= O_p(1) + o_p(1 + \sqrt{n}\|\beta_{n,t} - \beta_{0,t}\|) + o_p(1).
\]
Consequently, \(\{K_0 + o_p(1)\} \sqrt{n}\|\hat{\beta}_{n,t} - \beta_{0,t}\| \leq O_p(1) + o_p(1 + \sqrt{n}\|\hat{\beta}_{n,t} - \beta_{0,t}\|) + o_p(1)\), and thus
\[
\sqrt{n}\|\hat{\beta}_{n,t} - \beta_{0,t}\| = O_p(1).
\]
This leads to the conclusion that \(\sqrt{n}G_{n,t}(\hat{\beta}_{n,t}, \gamma_0) = \sqrt{n}G_{n,t}(\beta_{0,t}, \gamma_0) - H_{t}(\beta_{0,t})\sqrt{n}(\hat{\beta}_{n,t} - \beta_{0,t}) + o_p(1)\). Recalling that \(\hat{M}_{jt}(t; \beta_{0,t}, \gamma_0) = \hat{N}_{jt}(t; \gamma_0) - \int_0^t I(X_j \geq u) \exp(\beta_{0,t}^T Z_j) d\Lambda_{0,t}(u)\)
we have
\[
\sqrt{n}G_{n,t}(\beta_{0,t}, \gamma_0) = \sqrt{n}P_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \{Z_j - E(t, \beta_{0,t})\} d\hat{M}_{jt}(t; \beta_{0,t}, \gamma_0) \right] + o_p(1).
\]
Taking all the pieces together we have that
\[
0 = \sqrt{n}G_{n,t}(\hat{\beta}_{n,t}, \gamma_n)
\]
\[
= \sqrt{n}\{G_{n,t}(\hat{\beta}_{n,t}, \gamma_n) - G_{n,t}(\hat{\beta}_{n,t}, \gamma_0)\} + \sqrt{n}G_{n,t}(\hat{\beta}_{n,t}, \gamma_0)
\]
\[
= P \left[ \frac{1}{M} \sum_{j=1}^{m_0} I(M \geq j)(1 - R_j) \int_0^\tau [Z_j - E_n(t, \beta_{0,t})] dN_j(t) \hat{\pi}_l(W_j, \gamma_0)^T \right] \times \sqrt{n}(\gamma_n - \gamma_0)
\]
\[
+ \sqrt{n}P_n \left[ \frac{1}{M} \sum_{j=1}^{m_0} \int_0^\tau I(M \geq j) \{Z_j - E(t, \beta_{0,t})\} d\hat{M}_{jt}(t; \beta_{0,t}, \gamma_0) \right]
\]
\[
- H_{t}(\beta_{0,t})\sqrt{n}(\hat{\beta}_{n,t} - \beta_{0,t}) + o_p(1).
\]
Rearranging the terms and according to conditions C4 and C6 leads to
\[
\sqrt{n}(\hat{\beta}_{n,t} - \beta_{0,t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{M} \sum_{j=1}^{M_i} (\psi_{ij} + R_0(\omega_{ij})) \right\} + o_p(1),
\]
where
\[
\psi_{ij} = H_{t}^{-1}(\beta_{0,t}) \int_0^\tau \{Z_{ij} - E(t, \beta_{0,t})\} d\hat{M}_{ij}(t; \beta_{0,t}, \gamma_0),
\]
By similar analysis to that provided in Page 57 of Kosorok (2008), the second term in (2.4) can be written as

\[
R_i = H_t^{-1}(\beta_{0,i})E \left[ \frac{1}{M} \sum_{j=1}^{M} (1 - R_j) \int_0^r \{Z_j - E(t, \beta_{0,i})\} dN_j(t) \hat{\Lambda}_i(W_j, \gamma_0) \right].
\]

2.3 Proof of Theorem 2.3

By Taylor expansion and the consistency of \( \hat{\beta}_{n,l} \) and \( \hat{\gamma}_n \), the first term in the right side of expansion (2.4) can be written as

\[
A_{n,l}^*(t) = P_n \left[ \frac{1}{M} \sum_{j=1}^{M} (1 - R_j) \{ \pi_l(W_j, \hat{\gamma}_n) - \pi_l(W_j, \gamma_0) \} \int_0^t \frac{dN_j(s)}{P_n \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta_{0,i}^T Z_j) \right\}} \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{M_i} \sum_{j=1}^{M_i} R_j^*(t) \omega_{ij} \right\} + o_p(n^{-1/2}),
\]

where

\[
R_j^*(t) = E \left[ \frac{1}{M} \sum_{j=1}^{M} (1 - R_j) \hat{\pi}_j(W_j, \gamma_0) \int_0^t \frac{dN_j(s)}{E \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(s) \exp(\beta_{0,i}^T Z_j) \right\}} \right]^T.
\]

By similar analysis to that provided in Page 57 of Kosorok (2008), the second term in (2.4) can be written as

\[
B_{n,l}^*(t) = \left[ \int_0^t \frac{P_n \left\{ \frac{1}{M} \sum_{j=1}^{M} d\tilde{N}_j(t; \gamma_0) \right\}}{P_n \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta_{0,i}^T Z_j) \right\}} - \int_0^t \frac{P \left\{ \frac{1}{M} \sum_{j=1}^{M} d\tilde{N}_j(t; \gamma_0) \right\}}{P \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(t) \exp(\beta_{0,i}^T Z_j) \right\}} \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{M_i} \sum_{j=1}^{M_i} \phi_{ijl}(t) \right\} + o_p(n^{-1/2}),
\]

where

\[
\phi_{ijl}(t) = \int_0^t \frac{d\tilde{M}_{ijl}(s; \beta_{0,i}, \gamma_0)}{E \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(s) \exp(\beta_{0,i}^T Z_j) \right\}} - (\psi_{ijl} + R_i^*(t) \omega_{ij})^T \int_0^t E(s, \beta_{0,i}) d\Lambda_{0,i}(s).
\]

Therefore,

\[
\sqrt{n} \left\{ \hat{\Lambda}_{n,l}(t) - \Lambda_{0,l}(t) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} (\phi_{ijl}(t) + R_i^*(t) \omega_{ij}) \right] + o_p(1).
\]
By conditions C1, C4, and C5, and lemmas 1 and 2 in the supporting information of Bakoyannis (2021), the class of functions

\[
\left\{ \frac{1}{M} \sum_{j=1}^{M} \phi_{jl}(t) + \mathbf{R}_l^*(t) \omega_j \right\} : t \in [0, \tau]
\]

is Donsker. We now show that conditional on the data, the estimated multiplier process

\[
\tilde{W}_{n,l}(\cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} \phi_{ijl}(\cdot) + \hat{R}_l^*(\cdot) \omega_{ij} \right] \xi_i
\]

converges weakly to the same limiting process as \( W_{n,l}(\cdot) = \sqrt{n} \{ \hat{\Lambda}_{n,l}(\cdot) - \Lambda_{n,l}(\cdot) \} \). Define

\[
\tilde{W}_{n,l}(\cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} \phi_{ijl}(\cdot) + \hat{R}_l^*(\cdot) \omega_{ij} \right] \xi_i.
\]

By the Donsker property of the class of influence functions and the conditional multiplier central limit theorem (van der Vaart and Wellner, 1996), \( \tilde{W}_{n,l}(\cdot) \) converges weakly, conditionally on the data, to the same limiting process as \( W_{n,l}(\cdot) \). To complete the proof, we need to show

\[
\| \tilde{W}_{n,l}(\cdot) - \hat{W}_{n,l}(\cdot) \|_\infty = o_p(1),
\]

unconditionally. After some algebra

\[
\| \tilde{W}_{n,l}(\cdot) - \hat{W}_{n,l}(\cdot) \|_\infty \leq A''_{n,l} + B''_{n,l} + C''_{n,l},
\]

where

\[
A''_{n,l} = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} \left( \hat{\phi}_{ijl}(\cdot) - \phi_{ijl}(\cdot) \right) \right] \xi_i \right\|_\infty,
\]

\[
B''_{n,l} = \sup_{t \in [0, \tau]} \left\| \hat{R}_l^*(t) - R_l^*(t) \right\| \times \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} (\hat{\omega}_{ij} - \omega_{ij}) \right] \xi_i \right\| + \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{1}{M_i} \sum_{j=1}^{M_i} \omega_{ij} \right) \xi_i \right\|,
\]

and

\[
C''_{n,l} = \sup_{t \in [0, \tau]} \| R_l^*(t) \| \times \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{M_i} \sum_{j=1}^{M_i} (\hat{\omega}_{ij} - \omega_{ij}) \right] \xi_i \right\|.
\]

Using the same arguments to those used in the proof of Theorem 4 in Spiekerman and Lin (1998) and regularity conditions C3 and C4, \( A''_{n,l} = o_p(1) \).
For $B_{n,t}^{m}$, using the same arguments to those used in the proof of Lemma A.3 in Spiekerman and Lin (1998) leads to the conclusion that
\[
\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{M} \sum_{j=1}^{M} (\bar{\omega}_{ij} - \omega_{ij}) \right\} \xi_{i} \right\| = o_{p}(1).
\]
Next, by condition C4 and the central limit theorem
\[
\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{1}{M} \sum_{j=1}^{M} \omega_{ij} \right) \xi_{i} \right\| = O_{p}(1).
\]
Also, after some algebra, we have that
\[
\sup_{t \in [0, \tau]} \left\| \hat{\beta}_{n,t}^{*}(t) - \beta_{n,t}^{*}(t) \right\| \leq A_{n,t}^{m} + B_{n,t}^{m} + C_{n,t}^{m},
\]
where
\[
A_{n,t}^{m} = \sup_{t \in [0, \tau]} \left\| \mathbb{P}_{n} \left[ \frac{1}{M} \sum_{j=1}^{M} \left\{ \hat{\pi}_{l}(W_{j}, \hat{\gamma}_{n}) - \hat{\pi}_{l}(W_{j}, \gamma_{0}) \right\} \right] \right\|,
\]
\[
B_{n,t}^{m} = \sup_{t \in [0, \tau]} \left\| \mathbb{P}_{n} \left[ \frac{1}{M} \sum_{j=1}^{M} \hat{\pi}_{l}(W_{j}, \gamma_{0}) \right] \right\|,
\]
and
\[
C_{n,t}^{m} = \sup_{t \in [0, \tau]} \left\| (\mathbb{P}_{n} - P) \left[ \frac{1}{M} \sum_{j=1}^{M} \hat{\pi}_{l}(W_{j}, \gamma_{0}) \right] \right\|.
\]
By conditions C3, C4, C5, and the continuous mapping theorem,
\[
\max_{ij} \left\| \hat{\pi}_{l}(W_{ij}, \bar{\gamma}_{n}) - \hat{\pi}_{l}(W_{ij}, \gamma_{0}) \right\| = o_{as}(1).
\]
By Theorem 2.2 and conditions C1, C2, and C5,
\[
\left\| \mathbb{P}_{n} \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_{j}(\cdot) \exp(\hat{\beta}_{n,t}^{T}Z_{j}) \right\} \right\|_{\infty} = \left\| \mathbb{P} \left\{ \frac{1}{M} \sum_{j=1}^{M} Y_{j}(\cdot) \exp(\beta_{0,t}^{T}Z_{j}) \right\} + o_{as}(1) \right\|_{\infty} = O_{as}(1).
\]
Therefore, $A_{n,t}^{m} = o_{as}(1)$. Next, by conditions C3 and C5, $\max_{ij} \left\| \hat{\pi}_{l}(W_{ij}, \gamma_{0}) \right\| = O_{as}(1)$. Also, by conditions C2, C5, and the Donsker property of $\{Y_{j}(t) : t \in [0, \tau]\}$,
\[
\left\| \mathbb{P}_{n} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{M} Y_{j}(\cdot) \exp(\hat{\beta}_{n,t}^{T}Z_{j}) \right\} - \mathbb{P} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{M} Y_{j}(\cdot) \exp(\beta_{0,t}^{T}Z_{j}) \right\} \right\|_{\infty} = o_{as}(1),
\]
and, thus, $B_{n,l}'' = o_{as^*}(1)$. For $C_{n,l}'''$, consider the classes of functions $\mathcal{F} = \{N_j(s) : t \in [0, \tau]\}$ and
\[
\mathcal{L}_{l,1} = \left\{ f_{t,l} = \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}(M \geq j) \pi_l(W_j, \gamma_0) \int_0^t \frac{1}{P\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(s) e^{\beta_{0,l}^T z_j} \}} dN_j(s), t \in [0, \tau] \right\}.
\]
For any finitely discrete probability measure $Q$ and any $t_1, t_2 \in [0, \tau]$ we have that
\[
\|f_{t_1,l} - f_{t_2,l}\|_{Q,2} \leq \left\| \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}(M \geq j) \pi_l(W_j, \gamma_0) \int_{t_1}^{t_2} \frac{dN_j(s)}{P\{ \frac{1}{M} \sum_{j=1}^{M} Y_j(s) \exp(\beta_{0,l}^T z_j) \}} \right\|_{Q,2} \leq K_0 \|N_j(t_2) - N_j(t_1)\|_{Q,2}.
\]
Therefore, $\mathcal{L}_{l,1}$ can be covered by $N (\epsilon, \mathcal{F}, L_2(Q))$ $L_2(Q)$ $\epsilon'$-balls centered at $f_{t_1,l}$, where $\mathcal{F}$ is a Donsker class by lemma 4.1 in Kosorok (2008). In addition, using similar arguments to those in page 142 in Kosorok (2008), the class $\mathcal{L}_{l,1}$ is pointwise measurable. Consequently, the class $\mathcal{L}_{l,1}$ is Donsker and, thus, also Glivenko-Cantelli, which leads to the conclusion that $C_{n,l}''' = o_{as^*}(1)$.

Therefore, $\sup_{t \in [0, \tau]} \|\hat{R}_l^*(t) - R_l^*(t)\| = o_p(1)$ and, thus, $B_{n,l}'' = o_p(1)$. Similar arguments lead to the conclusion that $C_{n,l}''' = o_p(1)$. Thus,
\[
\left\| \hat{W}_{n,l}(\cdot) - \bar{W}_{n,l}(\cdot) \right\|_{\infty} = o_p(1),
\]
which completes the proof of the last statement in Theorem 2.3.

2.4 Proof of Theorem 2.4

It is easy to show that
\[
\hat{W}_{n,l}(t; z_0) = \sqrt{n} \{ \hat{A}_{n,l}(t; z_0) - \Lambda_{0,l}(t; z_0) \} = \sqrt{n} \{ \hat{A}_{n,l}(t) \exp(\beta_{n,l}^T z_0) - \Lambda_{0,l}(t) \exp(\beta_{0,l}^T z_0) \} = \sqrt{n} \{ \hat{A}_{n,l}(t) - \Lambda_{0,l}(t) + z_0^T (\beta_{n,l} - \beta_{0,l}) \Lambda_{0,l}(t) \exp(\beta_{0,l}^T z_0) + o_p(1) \}
\]
\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j=1}^{M_i} \{ z_0^T (\psi_{ijl} + R_l^*(t) \omega_{ij}) \Lambda_{0,l}(t) + \phi_{ijl}(t) + R_l^*(t) \omega_{ij} \} \exp(\beta_{0,l}^T z_0) + o_p(1)
\]
\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j=1}^{M_i} \phi_{ijl}(t; z_0) + o_p(1).
\]
Similarly to the decomposition in Cheng and others (1998),
\[
\sqrt{n} \left\{ \hat{F}_{n,t}(t; z_0) - F_{0,t}(t; z_0) \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{M_i} \sum_{j=1}^{M_i} \phi^F_{ij}(t; z_0) \right\} + o_p(1),
\]
where
\[
\phi^F_{ij}(t; z_0) = \int_0^t \exp \left\{ - \sum_{l=1}^{k} \Lambda_{0,l}(s-; z_0) \right\} d\phi^A_{ij}(s; z_0)
- \int_0^t \left\{ \exp \left\{ - \sum_{l=1}^{k} \Lambda_{0,l}(s-; z_0) \right\} \right\} d\Lambda_{0,l}(s; z_0),
\]
and \(\phi^A_{ij}(t; z_0) = \{ z_i^T (\psi_{ij} + R_i(\omega_{ij}) \Lambda_{0,l}(t) + \phi_{ij}(t) + R_i(t) \omega_{ij}) \} \exp(\beta_0^T z_0)\). The class of functions \(\{\phi^F_{ij}(t; z_0 : t \in [0, \tau]\} \) is Donsker by conditions C1, C4, and C5, lemmas 1 and 2 in the supporting information of Bakoyannis (2021), and corollary 9.32 in Kosorok (2008).

To conclude the proof of Theorem 2.4, we show that, conditionally on the data,
\[
\hat{W}_{n,t}^F(\cdot; z_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j=1}^{M_i} \hat{\phi}^F_{ij}(\cdot; z_0) \xi_i
\]
converges weakly to the same limiting process as \(W_{n,t}^F(\cdot; z_0) = \sqrt{n}(\hat{F}_{n,t}(\cdot; z_0) - F_{0,t}(\cdot; z_0))\). Now, define
\[
\tilde{W}_{n,t}^F(\cdot; z_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{j=1}^{M_i} \phi^F_{ij}(\cdot; z_0) \xi_i.
\]
By the Donsker property of the class of influence functions and the conditional multiplier central limit theorem (van der Vaart and Wellner, 1996), \(\tilde{W}_{n,t}^F(\cdot; z_0)\) converges weakly, conditionally on the data, to the same limiting process as \(W_{n,t}^F(\cdot; z_0)\). To complete the proof, we need to show that
\[
\|\hat{W}_{n,t}^F(\cdot; z_0) - \tilde{W}_{n,t}^F(\cdot; z_0)\|_{\infty} = o_p(1),
\]
unconditionally. Some algebra leads to the following bound
\[
\|\hat{W}_{n,t}^F(\cdot; z_0) - \tilde{W}_{n,t}^F(\cdot; z_0)\|_{\infty} \leq A_{n,t} + B_{n,t} + C_{n,t} + D_{n,t} + E_{n,t} + F_{n,t},
\]
where
\[
A_{n,t} = \left\| \int_0^t \exp \left\{ - \sum_{l=1}^{k} \hat{\Lambda}_{n,l}(\cdot-; z_0) \right\} d\left[ \sqrt{n} \sum_{j=1}^{M_i} \left\{ \hat{\phi}^A_{ij}(\cdot; z_0) - \phi^A_{ij}(\cdot; z_0) \right\} \right] \right\|_{\infty},
\]
The simulation results for the pointwise estimates of the infinite-dimensional parameters \( \Lambda \) which completes the proof of the last statement in Theorem 2.4.

By integration by parts, Theorem 2.3, Lemma A.3 in Spiekerman and Lin (1998) and the boundedness conditions, \( A''''_{n,l} = o_p(1) \). By Theorem 2.1, the Donsker property of the class \( \{ \phi_{jl}(t; z_0) : t \in [0, \tau] \} \), and arguments similar to those used in the proof of proposition 7.27 in Kosorok (2008), \( B''''_{n,l} = o_p(1) \). For \( C''''_{n,l} \), the integrand converges uniformly to 0 in probability and, thus, by Theorem 2.3 and arguments similar to those used in the proof of proposition 7.27 in Kosorok (2008), it follows that \( C''''_{n,l} = o_p(1) \). Using similar arguments, it can be shown that \( D''''_{n,l} = o_p(1) \). For \( E''''_{n,l} \), the integrand converges uniformly to 0 in probability and, thus, by condition C1 it follows that \( E''''_{n,l} = o_p(1) \). Finally, by Theorem 2.1, condition C1, and the central limit theorem, it follows that \( F''''_{n,l} = o_p(1) \). Therefore,

\[
\left\| \hat{W}_{n,l}^F (\cdot; z_0) - \tilde{W}_{n,l}^F (\cdot; z_0) \right\|_\infty = o_p(1),
\]

which completes the proof of the last statement in Theorem 2.4.

3. Additional Simulation Results

The simulation results for the pointwise estimates of the infinite-dimensional parameters \( \Lambda_{0,t}(t) \) and \( F_{0,t}(t) \) under scenario 1 are provided in Table 1. Simulation results under scenario 2 are
presented in Table 2, 3, and 4.

4. Additional HIV Data Analysis Results

Descriptive characteristics of EA-IeDEA study sample are presented in Table 5. The estimated residual process for the parametric model for the probability of death among those who were lost to clinic

\[
\text{logit}\{\pi_1(W, \gamma_0)\} = \gamma_0 + \gamma_1 T + \gamma_2 I(Sex = \text{"male"}) + \gamma_3 Age + \gamma_4 CD4 + \gamma_5 I(HIV\text{status} = \text{"yes"}),
\]

along with the corresponding 95% goodness-of-fit band, are depicted in Figure 1.

REFERENCES

BAKOYANNIS, G. (2021). Nonparametric analysis of nonhomogeneous multistate processes with clustered observations. *Biometrics* 77, 533–546.

BAKOYANNIS, G., ZHANG, Y. AND YIANNOUTSOS, C. T. (2019). Nonparametric inference for Markov processes with missing absorbing state. *Statistica Sinica* 29, 2083–2104.

BAKOYANNIS, G., ZHANG, Y. AND YIANNOUTSOS, C. T. (2020). Semiparametric regression and risk prediction with competing risks data under missing cause of failure. *Lifetime Data Analysis* 26, 659–684.

CHENG, S. C., FINE, J. P. AND WEI, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* 54, 219–228.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* New York: Springer.

PAN, Z. AND LIN, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics* 61, 1000–1009.
Table 1. Simulation results for the infinite-dimensional parameters \( \Lambda_{0,1}(t) \) and \( F_{0,1}(t) \) at selected time points under Scenario 1. Results from the proposed approach and the approach by Bakoyannis and others (2020) (BZY20) which ignores the within-cluster dependence.

| n   | \( p_m(\%) \) | t  | Bias | MCSD | ASE | CP | Bias | MCSD | ASE | CP |
|-----|----------------|----|------|------|-----|----|------|------|-----|----|
| 50  | 0.1            | 0.001 | 0.084 | 0.080 | 0.915 | 0.090 | 0.104 | 0.025 | 0.252 |
|     | 0.2            | 0.000 | 0.105 | 0.101 | 0.917 | 0.120 | 0.127 | 0.034 | 0.259 |
|     | 0.4            | 0.004 | 0.135 | 0.130 | 0.934 | 0.160 | 0.156 | 0.049 | 0.284 |
|     | 0.8            | 0.011 | 0.185 | 0.176 | 0.947 | 0.203 | 0.201 | 0.081 | 0.385 |
| 43  | 0.1            | 0.000 | 0.086 | 0.081 | 0.915 | 0.091 | 0.106 | 0.031 | 0.306 |
|     | 0.2            | -0.001 | 0.108 | 0.104 | 0.926 | 0.121 | 0.130 | 0.042 | 0.318 |
|     | 0.4            | 0.003 | 0.139 | 0.134 | 0.935 | 0.160 | 0.160 | 0.058 | 0.355 |
|     | 0.8            | 0.011 | 0.188 | 0.181 | 0.942 | 0.203 | 0.203 | 0.093 | 0.441 |
| 200 | 0.1            | 0.001 | 0.040 | 0.040 | 0.954 | 0.090 | 0.049 | 0.013 | 0.074 |
|     | 0.2            | 0.001 | 0.051 | 0.051 | 0.951 | 0.122 | 0.062 | 0.017 | 0.055 |
|     | 0.4            | 0.003 | 0.066 | 0.066 | 0.955 | 0.159 | 0.077 | 0.024 | 0.055 |
|     | 0.8            | 0.006 | 0.093 | 0.088 | 0.950 | 0.200 | 0.101 | 0.040 | 0.096 |
| 43  | 0.1            | 0.001 | 0.040 | 0.041 | 0.952 | 0.091 | 0.051 | 0.015 | 0.096 |
|     | 0.2            | 0.002 | 0.053 | 0.052 | 0.952 | 0.123 | 0.064 | 0.021 | 0.081 |
|     | 0.4            | 0.003 | 0.069 | 0.068 | 0.952 | 0.161 | 0.079 | 0.029 | 0.077 |
|     | 0.8            | 0.007 | 0.096 | 0.091 | 0.947 | 0.199 | 0.103 | 0.046 | 0.138 |

\[ F_{0,1}(t) \]

| n   | \( p_m(\%) \) | t  | Bias | MCSD | ASE | CP | Bias | MCSD | ASE | CP |
|-----|----------------|----|------|------|-----|----|------|------|-----|----|
| 50  | 0.1            | -0.002 | 0.049 | 0.047 | 0.924 | 0.036 | 0.055 | 0.014 | 0.300 |
|     | 0.2            | -0.003 | 0.051 | 0.049 | 0.926 | 0.033 | 0.057 | 0.015 | 0.343 |
|     | 0.4            | -0.002 | 0.053 | 0.051 | 0.933 | 0.025 | 0.058 | 0.017 | 0.388 |
|     | 0.8            | -0.002 | 0.053 | 0.051 | 0.938 | 0.015 | 0.058 | 0.018 | 0.429 |
| 43  | 0.1            | -0.002 | 0.049 | 0.047 | 0.924 | 0.036 | 0.056 | 0.018 | 0.381 |
|     | 0.2            | -0.003 | 0.053 | 0.051 | 0.924 | 0.033 | 0.059 | 0.020 | 0.427 |
|     | 0.4            | -0.002 | 0.055 | 0.053 | 0.934 | 0.026 | 0.060 | 0.023 | 0.493 |
|     | 0.8            | -0.002 | 0.056 | 0.054 | 0.936 | 0.015 | 0.060 | 0.024 | 0.555 |
| 200 | 0.1            | 0.000 | 0.023 | 0.024 | 0.954 | 0.038 | 0.027 | 0.007 | 0.177 |
|     | 0.2            | 0.000 | 0.025 | 0.025 | 0.950 | 0.035 | 0.028 | 0.008 | 0.198 |
|     | 0.4            | 0.000 | 0.026 | 0.026 | 0.952 | 0.028 | 0.029 | 0.008 | 0.287 |
|     | 0.8            | 0.000 | 0.027 | 0.026 | 0.945 | 0.016 | 0.029 | 0.009 | 0.402 |
| 43  | 0.1            | 0.000 | 0.024 | 0.024 | 0.952 | 0.038 | 0.028 | 0.009 | 0.204 |
|     | 0.2            | 0.000 | 0.026 | 0.026 | 0.953 | 0.036 | 0.029 | 0.010 | 0.276 |
|     | 0.4            | 0.000 | 0.028 | 0.027 | 0.949 | 0.028 | 0.030 | 0.011 | 0.368 |
|     | 0.8            | 0.000 | 0.028 | 0.027 | 0.945 | 0.017 | 0.030 | 0.012 | 0.503 |

Note: \( n \): number of clusters with cluster size \( M \in [30, 60] \); \( p_m \): percentage of missingness; \( t \): selected time points;

MCSD: Monte Carlo standard deviation; ASE: average estimated standard error; CP: coverage probability of 95% pointwise confidence interval
Table 2. Simulation results for the regression coefficient $\beta_1$ under Scenario 2 for the proposed approach and the approach by Bakoyannis and others (2020) (BZY20) which ignores the within-cluster dependence.

| $n$ | $p_m$ (%) | Proposed |        |        |        | BZY20 |        |        |        |
|-----|-----------|----------|--------|--------|--------|-------|--------|--------|--------|
|     |           | Bias     | MCSD   | ASE    | CP     | Bias  | MCSD   | ASE    | CP     |
| 50  | 26        | -0.002   | 0.035  | 0.033  | 0.938  | 0.008 | 0.035  | 0.023  | 0.781  |
| 36  |           | 0.000    | 0.036  | 0.035  | 0.936  | 0.011 | 0.036  | 0.025  | 0.803  |
| 44  |           | 0.002    | 0.037  | 0.036  | 0.936  | 0.013 | 0.037  | 0.027  | 0.832  |
| 100 | 26        | 0.002    | 0.024  | 0.024  | 0.950  | 0.012 | 0.023  | 0.016  | 0.740  |
| 36  |           | 0.004    | 0.025  | 0.025  | 0.937  | 0.014 | 0.024  | 0.018  | 0.771  |
| 44  |           | 0.005    | 0.026  | 0.026  | 0.937  | 0.016 | 0.025  | 0.019  | 0.782  |
| 200 | 26        | 0.003    | 0.017  | 0.017  | 0.945  | 0.013 | 0.016  | 0.011  | 0.681  |
| 36  |           | 0.005    | 0.017  | 0.018  | 0.943  | 0.016 | 0.017  | 0.012  | 0.672  |
| 44  |           | 0.007    | 0.018  | 0.018  | 0.940  | 0.018 | 0.018  | 0.014  | 0.663  |

Note: $n$: number of clusters with cluster size $M \in [30, 60]$; $p_m$: percentage of missingness; MCSD: Monte Carlo standard deviation; ASE: average estimated standard error; CP: coverage probability of 95% confidence interval.

Fig. 1. Plot for the cumulative residual process of the parametric model $\pi_1(W, \gamma_0)$ for death while in care among those lost to clinic based on the HIV data along with the 95% goodness-of-fit band (grey area) and the corresponding $p$-value.
Table 3. Simulation results for the infinite-dimensional parameters $\Lambda_{0,1}(t)$ and $F_{0,1}(t)$ at selected time points under Scenario 2. Results from the proposed approach and the approach by Bakoyannis and others (2020) (BZY20) which ignores the within-cluster dependence.

| n  | $p_m$ (%) | $t$ | Proposed  | BZY20     |
|----|-----------|----|-----------|-----------|
|    |           |    | $\Lambda_{0,1}(t)$ |          |
|    |           |    | Bias | MCSD | ASE | CP | Bias | MCSD | ASE | CP |
| 50 | 26        | 0.1 | -0.004 | 0.088 | 0.083 | 0.911 | 0.084 | 0.106 | 0.028 | 0.308 |
|    |           | 0.2 | -0.013 | 0.109 | 0.104 | 0.907 | 0.104 | 0.129 | 0.037 | 0.317 |
|    |           | 0.4 | -0.014 | 0.141 | 0.134 | 0.922 | 0.139 | 0.161 | 0.052 | 0.339 |
|    |           | 0.8 | 0.004  | 0.193 | 0.182 | 0.939 | 0.200 | 0.210 | 0.083 | 0.405 |
| 44 |           | 0.1 | -0.008 | 0.088 | 0.083 | 0.907 | 0.078 | 0.106 | 0.034 | 0.376 |
|    |           | 0.2 | -0.024 | 0.110 | 0.105 | 0.897 | 0.090 | 0.130 | 0.044 | 0.403 |
|    |           | 0.4 | -0.028 | 0.142 | 0.136 | 0.911 | 0.121 | 0.162 | 0.061 | 0.422 |
|    |           | 0.8 | -0.001 | 0.196 | 0.186 | 0.945 | 0.194 | 0.214 | 0.095 | 0.461 |
| 200| 26        | 0.1 | -0.004 | 0.041 | 0.042 | 0.955 | 0.085 | 0.050 | 0.014 | 0.101 |
|    |           | 0.2 | -0.012 | 0.053 | 0.052 | 0.931 | 0.106 | 0.062 | 0.018 | 0.109 |
|    |           | 0.4 | -0.015 | 0.068 | 0.067 | 0.934 | 0.139 | 0.078 | 0.026 | 0.111 |
|    |           | 0.8 | -0.001 | 0.096 | 0.091 | 0.948 | 0.196 | 0.104 | 0.041 | 0.111 |
| 44 |           | 0.1 | -0.007 | 0.041 | 0.042 | 0.942 | 0.079 | 0.050 | 0.017 | 0.163 |
|    |           | 0.2 | -0.021 | 0.053 | 0.053 | 0.916 | 0.092 | 0.063 | 0.022 | 0.201 |
|    |           | 0.4 | -0.029 | 0.069 | 0.068 | 0.916 | 0.121 | 0.079 | 0.030 | 0.205 |
|    |           | 0.8 | -0.006 | 0.098 | 0.093 | 0.940 | 0.188 | 0.106 | 0.047 | 0.168 |
| 200| 26        | 0.1 | 0.001  | 0.045 | 0.043 | 0.928 | 0.024 | 0.051 | 0.013 | 0.368 |
|    |           | 0.2 | -0.002 | 0.048 | 0.046 | 0.925 | 0.017 | 0.054 | 0.015 | 0.409 |
|    |           | 0.4 | -0.003 | 0.051 | 0.049 | 0.931 | 0.008 | 0.056 | 0.016 | 0.425 |
|    |           | 0.8 | -0.001 | 0.052 | 0.050 | 0.932 | -0.002 | 0.057 | 0.017 | 0.448 |
| 44 |           | 0.1 | 0.004  | 0.046 | 0.044 | 0.929 | 0.026 | 0.052 | 0.018 | 0.452 |
|    |           | 0.2 | -0.002 | 0.050 | 0.048 | 0.922 | 0.017 | 0.055 | 0.020 | 0.483 |
|    |           | 0.4 | -0.004 | 0.053 | 0.051 | 0.927 | 0.008 | 0.058 | 0.022 | 0.533 |
|    |           | 0.8 | -0.001 | 0.054 | 0.052 | 0.933 | -0.001 | 0.058 | 0.023 | 0.567 |
| 200| 26        | 0.1 | 0.003  | 0.022 | 0.022 | 0.957 | 0.025 | 0.025 | 0.007 | 0.270 |
|    |           | 0.2 | 0.000  | 0.024 | 0.024 | 0.947 | 0.019 | 0.027 | 0.007 | 0.347 |
|    |           | 0.4 | -0.002 | 0.025 | 0.025 | 0.943 | 0.009 | 0.028 | 0.008 | 0.420 |
|    |           | 0.8 | -0.001 | 0.026 | 0.025 | 0.944 | -0.001 | 0.029 | 0.009 | 0.442 |
| 44 |           | 0.1 | 0.006  | 0.022 | 0.023 | 0.952 | 0.029 | 0.026 | 0.009 | 0.306 |
|    |           | 0.2 | 0.000  | 0.025 | 0.024 | 0.942 | 0.019 | 0.028 | 0.010 | 0.434 |
|    |           | 0.4 | -0.002 | 0.026 | 0.026 | 0.945 | 0.009 | 0.029 | 0.011 | 0.517 |
|    |           | 0.8 | 0.000  | 0.027 | 0.027 | 0.939 | 0.000 | 0.030 | 0.012 | 0.554 |

Note: $n$: number of clusters with cluster size $M \in [30, 60]$; $p_m$: percentage of missingness; $t$: selected time points; MCSD: Monte Carlo standard deviation; ASE: average estimated standard error; CP: coverage probability of 95% pointwise confidence interval.
Table 4. Simulation results for the coverage probabilities of 95% simultaneous confidence bands for the infinite-dimensional parameters $\Lambda_{0,1}(t)$ and $F_{0,1}(t)$ under Scenario 2. Results from the proposed approach and the approach by Bakoyannis and others (2020) (BZY20) which ignores the within-cluster dependence.

| n  | $p_m(\%)$ | $\Lambda_{0,1}(t)$ | $F_{0,1}(t)$ |
|----|----------|-------------------|--------------|
|    |          | EP    | HW    | Proposed | EP    | HW    | Proposed | BZY20 | EP    | HW    | Proposed | BZY20 |
| 50 | 26       | 0.768 | 0.914 | 0.078 | 0.152 | 0.803 | 0.924 | 0.087 | 0.171 |
| 36 | 0.684    | 0.908 | 0.073 | 0.166 | 0.731 | 0.923 | 0.081 | 0.169 |
| 44 | 0.598    | 0.915 | 0.078 | 0.163 | 0.661 | 0.927 | 0.078 | 0.172 |
| 100| 26       | 0.755 | 0.942 | 0.038 | 0.098 | 0.796 | 0.935 | 0.035 | 0.113 |
| 36 | 0.568    | 0.938 | 0.026 | 0.089 | 0.609 | 0.931 | 0.024 | 0.089 |
| 44 | 0.421    | 0.941 | 0.013 | 0.069 | 0.452 | 0.927 | 0.010 | 0.060 |
| 200| 26       | 0.684 | 0.946 | 0.009 | 0.045 | 0.717 | 0.940 | 0.013 | 0.065 |
| 36 | 0.348    | 0.942 | 0.002 | 0.035 | 0.358 | 0.929 | 0.001 | 0.034 |
| 44 | 0.169    | 0.922 | 0.000 | 0.012 | 0.179 | 0.914 | 0.000 | 0.013 |

Note: $n$: number of clusters with cluster size $M \in [30, 60]$; $p_m$: percentage of missingness; EP: equal precision bands; HW: Hall-Wellner-type bands.

Table 5. Descriptive characteristics for the EA-IeDEA study sample

| Variable                          | Right censoring | Cause of failure | Missing |
|-----------------------------------|-----------------|------------------|---------|
|                                   | (N=8082)        | (N=1951$^1$)     | (N=11099) |
| Gender                            |                 |                  |         |
| Female                            | 5334 (66.0)     | 974 (49.9)       | 7362 (66.3) |
| Male                              | 2748 (34.0)     | 977 (50.1)       | 3737 (33.7) |
| HIV status disclosed              |                 |                  |         |
| Yes                               | 5116 (63.3)     | 1417 (72.6)      | 6916 (62.3) |
| No                                | 2966 (36.7)     | 534 (27.4)       | 4183 (37.7) |
| Median (IQR)                      |                 |                  |         |
| Age$^2$                           |                 |                  |         |
| CD4$^3$                           |                 |                  |         |

Note: $^1$: Included 84 reported deaths and 1867 unreported deaths confirmed through outreach; $^2$: Age at ART initiation in years; $^3$: CD4 count at ART initiation in cells/μl
REFERENCES

SPIEKERMAN, C. F. AND LIN, D. Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93**, 1164–1175.

VAN DER VAART, A. W. AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.

[Received August 1, 2010; revised October 1, 2010; accepted for publication November 1, 2010]