Semiparametric Regression Analysis of Multiple Right- and Interval-Censored Events

Fei Gao, Donglin Zeng, David Couper & D. Y. Lin

To cite this article: Fei Gao, Donglin Zeng, David Couper & D. Y. Lin (2018): Semiparametric Regression Analysis of Multiple Right- and Interval-Censored Events, Journal of the American Statistical Association, DOI: 10.1080/01621459.2018.1482756

To link to this article: https://doi.org/10.1080/01621459.2018.1482756

View supplementary material

Accepted author version posted online: 05 Jun 2018.

Submit your article to this journal

Article views: 73

View Crossmark data
Semiparametric Regression Analysis of Multiple Right- and Interval-Censored Events

Fei Gao, Donglin Zeng, David Couper, and D. Y. Lin

Abstract

Health sciences research often involves both right- and interval-censored events because the occurrence of a symptomatic disease can only be observed up to the end of follow-up, while the occurrence of an asymptomatic disease can only be detected through periodic examinations. We formulate the effects of potentially time-dependent covariates on the joint distribution of multiple right- and interval-censored events through semiparametric proportional hazards models with random effects that capture the dependence both within and between the two types of events. We consider nonparametric maximum likelihood estimation and develop a simple and stable EM algorithm for computation. We show that the resulting estimators are consistent and the parametric components are asymptotically normal and efficient with a covariance matrix that can be consistently estimated by profile likelihood or nonparametric bootstrap. In addition, we leverage the joint modelling to provide dynamic prediction of disease incidence based on the evolving event history. Furthermore, we assess the performance of the proposed methods through extensive simulation studies. Finally, we provide an application to a major epidemiological cohort study.

Keywords: Dynamic prediction; Joint models; Nonparametric likelihood; Proportional hazards; Random effects; Semiparametric efficiency.
1. INTRODUCTION

Many clinical and epidemiological studies are concerned with multiple types of diseases, which may be symptomatic or asymptomatic. Time to the development of a symptomatic disease is right-censored if the disease does not occur during the follow-up, whereas time to the development of an asymptomatic disease is typically interval-censored because the disease occurrence can only be monitored periodically using biomarkers. In the Atherosclerosis Risk in Communities (ARIC) study (The ARIC Investigators 1989), for instance, subjects were followed for up to 27 years for symptomatic cardiovascular diseases, such as myocardial infarction (MI) and stroke, through reviews of hospital records; they were also examined over five clinic visits, with the first four at approximately 3-year intervals, for occurrences of asymptomatic diseases, such as diabetes and hypertension.

There is a large body of literature on the joint analysis of correlated right-censored events (Kalbfleisch and Prentice 2002, chap. 10; Hougaard 2012), as well as a growing body of literature on correlated interval-censored events (Goggins and Finkelstein 2000; Kim and Xue 2002; Wen and Chen 2013; Chen et al. 2014; Zeng et al. 2017). In addition, there is a considerable amount of literature on competing risks and semi-competing risks (Fine and Gray 1999; Fine et al. 2001; Kalbfleisch and Prentice 2002, chap. 8). However, the existing literature has treated right-censored and interval-censored events separately. Joint modelling of the two kinds of data would allow investigators to evaluate the effects of covariates on both kinds of events and to predict the occurrence of a symptomatic disease given the history of asymptomatic diseases.

In this paper, we relate potentially time-dependent covariates to the joint distribution of multiple types of right- and interval-censored event times through semiparametric proportional hazards models with random effects. Specifically, we assume a shared random effect for the interval-censored events, which affects the right-censored events with unknown
coefficients. We assume an additional shared random effect for the right-censored events to capture their own dependence. The proposed models allow semi-competing risks and are reminiscent of selection models for joint modeling of survival and longitudinal data (Hogan and Laird 1997).

We estimate the model parameters through nonparametric maximum likelihood estimation, under which the baseline hazard functions are completely nonparametric. We develop a simple EM algorithm that converges stably for arbitrary sample sizes, even with time-dependent covariates. We show that the resulting estimators are consistent and the parametric components are asymptotically normal and asymptotically efficient. We also show that the covariance matrix of the parametric components can be estimated consistently with profile likelihood or nonparametric bootstrap. We pay special attention to the estimation of the conditional distribution function given the event history, which can be used to predict disease occurrence dynamically. Finally, we assess the performance of the proposed numerical and inferential procedures through extensive simulation studies and provide a substantive application to the ARIC data on diabetes, hypertension, stroke, MI, and death.

2. METHODS

2.1. Data, Models, and Likelihood

Suppose that there are $K_1$ types of asymptomatic events occurring at times $T_1, \ldots, T_{K_1}$ and $K_2$ types of symptomatic events occurring at times $T_{K_1+1}, \ldots, T_K$, where $K = K_1 + K_2$. Let $X_k(\cdot)$ be a $p$-vector of possibly time-dependent external covariates for the event time $T_k$. For $k = 1, \ldots, K_1$, the hazard function of $T_k$ conditional on covariate $X_k$ and random effect $b_1$ is given by

$$
\lambda_k(t; X_k, b_1) = e^{\beta^T X_k(t) + b_1} \lambda_k(t),
$$

(1)
where $\beta$ is a set of unknown regression parameters, $\lambda_k(\cdot)$ is an arbitrary baseline hazard function, and $b_1$ is a latent normal random variable with mean zero and variance $\sigma_1^2$. For $k = K_1 + 1, \ldots, K$, the hazard function of $T_k$ conditional on covariates $X_k$ and random effects $b_1$ and $b_2$ is given by

$$
\lambda_k(t; X_k, b_1, b_2) = e^{\beta^T X_k(t)} + \gamma_k b_1 + b_2 \lambda_k(t),
$$

(2)

where $\lambda_k(\cdot)$ is an arbitrary baseline hazard function, $\gamma \equiv (\gamma_{K_1+1}, \ldots, \gamma_K)^T$ is a set of unknown coefficients, and $b_2$ is a latent normal random variable with mean zero and variance $\sigma_2^2$. Write $\Sigma = (\sigma_1^2, \sigma_2^2)$. By letting $X_k$ depend on $k$, models (1) and (2) allow the regression parameters to be different among the $K$ events by appropriate definitions of dummy variables; see Lin (1994).

We implicitly assume that $K_1$ and $K_2$ are greater than one; otherwise, some of the parameters need to be fixed to ensure identifiability. For example, if $K_1 = K_2 = 1$, we require $\sigma_2^2 = 0$ and $\gamma_1 = 1$; if $K_1 > 1$ and $K_2 = 1$, we require $\sigma_2^2 = 0$; and if $K_1 = 1$ and $K_2 > 1$, we require one of the $\gamma_k$’s to be 1. In the last scenario, we may set different $\gamma_k$ to 1 and choose the model that yields the largest value of the likelihood function.

Remark 1. The random effects $b_1$ and $b_2$ characterize the underlying health conditions for the asymptomatic and symptomatic events, respectively. The random effect for the asymptomatic events affects the $k$th symptomatic event through the unknown coefficient $\gamma_k$. For example, in the ARIC study, $b_1$ represents the common pathways for diabetes and hypertension, such as obesity, inflammation, oxidative stress, and insulin resistance, which also serve as potential risk factors for MI, stroke, and death. The random effect $b_2$ represents the underlying propensity for major cardiovascular diseases and death.

Suppose that the asymptomatic event time $T_k$ ($k = 1, \ldots, K_1$) is monitored at a sequence of positive time points $U_{k1} < \cdots < U_{k,M_k}$ and is known to lie in the interval $(L_k, R_k]$, where $L_k = \max\{U_{kl} : U_{kl} < T_k, l = 0, \ldots, M_k\}$, and $R_k = \min\{U_{kl} : U_{kl} \geq T_k, l = \ldots, M_k\}$.
1, \ldots, M_k + 1 \}, \text{ with } U_{k0} = 0 \text{ and } U_{k,M_k + 1} = \infty. \text{ Let } C_k \text{ denote the censoring time on the }
\text{symptomatic event time } T_k (k = K_1 + 1, \ldots, K) \text{ such that we observe } Y_k = \min(T_k, C_k) \text{ and } \Delta_k = I(T_k \leq C_k), \text{ where } I(\cdot) \text{ is the indicator function. For a random sample of } n \text{ subjects, the data consist of } \{O_i : i = 1, \ldots, n\}, \text{ where }
\begin{align*}
O_i = \{L_{ik}, R_{ik}, X_{ik}(\cdot) : k = 1, \ldots, K_1\} \cup \{Y_{ik}, \Delta_{ik}, X_{ik}(\cdot) : k = K_1 + 1, \ldots, K\}.
\end{align*}
\text{ We assume that } \{U_{ikl} : k = 1, \ldots, K_1; l = 1, \ldots, M_{ik}\} \text{ and } \{C_{ik} : k = K_1 + 1, \ldots, K\} \text{ are independent of } \{T_{ik} : k = 1, \ldots, K\} \text{ and } b_i \equiv (b_{i1}, b_{i2}) \text{ conditional on } \{X_{ik}(\cdot) : k = 1, \ldots, K\}. \text{ Then, the likelihood concerning the parameters } \theta \equiv (\beta, \gamma, \Sigma) \text{ and } A \equiv (\Lambda_1, \ldots, \Lambda_K) \text{ is }
\begin{align*}
\prod_{i=1}^{n} \int_{b_{i1}}^{B_{i1}} \prod_{k=1}^{K_1} \left[ \exp \left\{ - \int_0^{L_{ik}} e^{\beta^T X_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} - \exp \left\{ - \int_0^{R_{ik}} e^{\beta^T X_{ik}(s) + b_{i1}} d\Lambda_k(s) \right\} \right] \\
\times \prod_{k=K_1+1}^{K} \left\{ e^{\beta^T X_{ik}(Y_{ik}) + \gamma b_{i1} + b_{i2}} \lambda_k(Y_{ik}) \right\}^{\Delta_{ik}} \exp \left\{ - \int_0^{Y_{ik}} e^{\beta^T X_{ik}(s) + \gamma b_{i1} + b_{i2}} d\Lambda_k(s) \right\} \psi(b_i; \Sigma) \psi(b_i; \Sigma) \psi(b_i; \Sigma) = 0.
\end{align*}
\text{ In some studies, one of the symptomatic events is terminal (e.g., death), such that we have a semi-competing risks set-up (Fine et al. 2001), where the occurrence of the terminal event precludes the development of the other events but not vice versa. Without loss of generality, suppose that the } K \text{th event is terminal. Then the monitoring times for } T_k (k \leq K_1) \text{ consist of the } U_{kl} \text{'s that are smaller than } T_K, \text{ and the censoring time for } T_k (k = K_1 + 1, \ldots, K - 1) \text{ is } \min(C_k, T_K). \text{ Conditional on } (b_1, b_2), \text{ the event times } T_1, \ldots, T_{K-1} \text{ are mutually independent and are independent of the monitoring times and censoring times. Thus, for any set } S_k \text{ that may depend on the monitoring times and censoring times, the joint probability of } T_1 \in S_1, \ldots, T_{K-1} \in S_{K-1} \text{ conditional on } (b_1, b_2) \text{ is equal to } \prod_{k=1}^{K-1} P(T_k \in S_k | b_1, b_2) \text{ with } S_k \text{ as a deterministic set. Therefore, the likelihood remains the same as before.}
2.2. Estimation Procedure

We adopt the nonparametric maximum likelihood estimation approach. For $k = 1, \ldots, K$, let $0 = t_{k0} < t_{k1} < t_{k2} < \cdots < t_{km_k} < \infty$ be the ordered sequence of all $L_{ik}$ and $R_{ik}$ with $R_{ik} < \infty$. For $k = K_1 + 1, \ldots, K$, let $0 = t_{k0} < t_{k1} < t_{k2} < \cdots < t_{km_k} < \infty$ be the ordered sequence of all $Y_{ik}$ with $\Delta_{ik} = 1$. The estimator for $\Lambda_k$ ($k = 1, \ldots, K$) is a step function that jumps only at $t_{k1}, \ldots, t_{km_k}$ with respective jump sizes $\lambda_k \equiv (\lambda_{k1}, \ldots, \lambda_{km_k})$.

We maximize the objective function

\[
L_n(\theta, \Lambda) = \prod_{i=1}^{n} \left\{ \prod_{k=1}^{K_1} g_{ik}^{(1)}(b_{i1}; \beta, \lambda_k) \right\} \left\{ \prod_{k=K_1+1}^{K} g_{ik}^{(2)}(b_i; \beta, \lambda_k) \right\} \psi(b_i; \Sigma) db_i,
\]

over $\theta$ and $\lambda_1, \ldots, \lambda_K$, where

\[
g_{ik}^{(1)}(b_{i1}; \beta, \lambda_k) = \exp \left( - \sum_{t_{kl} \leq L_{ik}} e^{\beta^T X_{ikl} + b_{i1} \lambda_{kl}} \right) - I(R_{ik} < \infty) \exp \left( - \sum_{t_{kl} \leq R_{ik}} e^{\beta^T X_{ikl} + b_{i1} \lambda_{kl}} \right),
\]

\[
g_{ik}^{(2)}(b_i; \beta, \lambda_k) = \left[ \Lambda_k \{Y_{ik} \} e^{\beta^T X_{ik} (Y_{ik} + 2 \gamma_{k1} b_{i1} + b_{i2})} \right]^{\Delta_{ik}} \exp \left( - \sum_{t_{kl} \leq Y_{ik}} e^{\beta^T X_{ikl} + 2 \gamma_{k1} b_{i1} + b_{i2} \lambda_{kl}} \right),
\]

$X_{ikl} = X_{ik}(t_{kl})$ for $k = 1, \ldots, K$ and $l = 1, \ldots, m_k$, and $\Lambda_k \{Y_{ik}\}$ is the jump size of $\Lambda_k$ at $Y_{ik}$.

Direct maximization of the objective function is difficult due to the lack of analytical expressions for $\lambda_1, \ldots, \lambda_K$. We introduce latent Poisson random variables to form a likelihood equivalent to the objective function such that the maximum likelihood estimators can be easily obtained via a simple EM algorithm. For $k = 1, \ldots, K_1$, we denote $R^*_{ik} = I(R_{ik} = \infty) L_{ik} + I(R_{ik} < \infty) R_{ik}$ and introduce independent Poisson random variables $W_{ikl}$ ($l = 1, \ldots, m_k, t_{kl} \leq R^*_{ik}$) with means $\lambda_{kl} \exp(\beta^T X_{ikl} + b_{i1})$. Conditional on $b_{i1}$, the likelihood function of $\{W_{ikl}; l = 1, \ldots, m_k, t_{kl} \leq R^*_{ik}\}$ is

\[
\prod_{l=1, t_{kl} \leq R^*_{ik}}^{m_k} \left\{ \frac{1}{W_{ikl}} \left( \lambda_{kl} e^{\beta^T X_{ikl} + b_{i1}} \right)^{W_{ikl}} \exp\left( -\lambda_{kl} e^{\beta^T X_{ikl} + b_{i1}} \right) \right\}.
\]
Let $A_{ik} = \sum_{tkl \leq L_{ik}} W_{ikl}$ and $B_{ik} = I(R_{ik} < \infty) \sum_{L_{ik} < tkl \leq R_{ik}} W_{ikl}$. The observed-data likelihood for $A_{ik} = 0$ and $B_{ik} > 0$ given $b_{i1}$ is equal to
\[
\exp\left( - \sum_{tkl \leq L_{ik}} e^{\theta^T X_{ikl} + b_{i1}} \lambda_{kl} \right) - I(R_{ik} < \infty) \exp\left( - \sum_{tkl \leq R_{ik}} e^{\theta^T X_{ikl} + b_{i1}} \lambda_{kl} \right),
\]
which is the same as $g_{ik}^{(1)}(b_{i1}; \beta, \lambda_k)$. Therefore, the objective function $L_n(\theta, \mathcal{A})$ can be viewed as the observed-data likelihood for $\{A_{ik} = 0, B_{ik} > 0 : i = 1, \ldots, n; k = 1, \ldots, K_1\} \cup \{Y_{ik}, \Delta_{ik} : i = 1, \ldots, n; k = K_1 + 1, \ldots, K\}$ with $(W_{ikl}, b_i)$ ($i = 1, \ldots, n; k = 1, \ldots, K_1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*)$ as latent variables. In view of the foregoing results, we propose an EM algorithm treating $W_{ikl}$ and $b_i$ as missing data.

In the M-step, we maximize the conditional expectation of the complete-data log-likelihood given the observed data so as to update the parameters. In particular, the conditional expectation of the complete-data log-likelihood is
\[
\sum_{i=1}^{n} \tilde{E}\left( \sum_{k=1}^{K_1} \left[ \sum_{l=1}^{m_k} I(t_{kl} \leq R_{ik}^*) \{W_{ikl} (\log \lambda_{kl} + \beta^T X_{ikl} + b_{i1}) - \lambda_{kl} \exp(\beta^T X_{ikl} + b_{i1})\} \right]
+ \sum_{k=K_1+1}^{K} \left[ \Delta_{ik} \left\{ \log \Lambda_k \{Y_{ik}\} + \beta^T X_{ik}(Y_{ik}) + \gamma_k b_{i1} + b_{i2} \right\}
- \sum_{t_{kl} \leq Y_{ik}} \lambda_{kl} \exp\left( \beta^T X_{ikl} + \gamma_k b_{i1} + b_{i2} \right) \right] \right),
\]
where $\tilde{E}(\cdot)$ denotes the conditional expectation given the observed data $\overline{O}_i$ ($i = 1, \ldots, n$), with $\overline{O}_i = \{A_{ik} = 0, B_{ik} > 0, X_{ik}(\cdot) : k = 1, \ldots, K_1\} \cup \{Y_{ik}, \Delta_{ik}, X_{ik}(\cdot) : k = K_1 + 1, \ldots, K\}$. To update the parameters, we first differentiate (3) with respect to $\lambda_{kl}$ ($k = 1, \ldots, K; l = 1, \ldots, m_k$) to obtain the updating formulas for $\lambda_k$:
\[
\lambda_{kl} = \frac{\sum_{i=1}^{n} I(t_{kl} \leq R_{ik}^*) \tilde{E}(W_{ikl})}{\sum_{i=1}^{n} I(t_{kl} \leq R_{ik}^*) \tilde{E}\{\exp(\beta^T X_{ikl} + b_{i1})\}}
\]
for $k = 1, \ldots, K_1$ and $l = 1, \ldots, m_k$ and
\[
\lambda_{kl} = \frac{\sum_{i=1}^{n} \Delta_{ik} I(Y_{ik} = t_{kl})}{\sum_{i=1}^{n} I(Y_{ik} \geq t_{kl}) \tilde{E}\{\exp(\beta^T X_{ikl} + \gamma_k b_{i1} + b_{i2})\}}
\]
for $k = K_1 + 1, \ldots, K$ and $l = 1, \ldots, m_k$. We then update $\beta$ by solving the equation

$$
\sum_{i=1}^n \left\{ \sum_{k=1}^{K} \sum_{l=1}^{m_k} \hat{E}(W_{ikl}) I(t_{kl} \leq R_{ik}^*) \left[ X_{ikl} - \frac{\sum_{j=1}^n X_{jkl} I(t_{kl} \leq R_{ijl}^*) \hat{E} \{ \exp(\beta^T X_{jkl} + b_{j1}) \}}{\sum_{j=1}^n I(t_{kl} \leq R_{ijl}^*) \hat{E} \{ \exp(\beta^T X_{jkl} + b_{j1}) \}} \right] \right\} + \sum_{k=K+1}^K \Delta_{ik} \left( X_{ik}(Y_{ik}) - \frac{\sum_{j=1}^n I(Y_{jk} \geq Y_{ik}) X_{jk}(Y_{ik}) \hat{E} \{ \exp(\beta^T X_{jk}(Y_{ik}) + \gamma_k b_{j1} + b_{j2}) \}}{\sum_{j=1}^n I(Y_{jk} \geq Y_{ik}) \hat{E} \{ \exp(\beta^T X_{jk}(Y_{ik}) + \gamma_k b_{j1} + b_{j2}) \}} \right) = 0
$$

and update $\gamma_k$ by solving the equation

$$
\sum_{i=1}^n \Delta_{ik} \left( \hat{E}(b_{i1}) - \frac{\sum_{j=1}^n I(Y_{jk} \geq Y_{ik}) \hat{E} \{ \exp(\beta^T X_{jk}(Y_{ik}) + \gamma_k b_{j1} + b_{j2}) \}}{\sum_{j=1}^n I(Y_{jk} \geq Y_{ik}) \hat{E} \{ \exp(\beta^T X_{jk}(Y_{ik}) + \gamma_k b_{j1} + b_{j2}) \}} \right) = 0.
$$

The two equations are obtained by differentiating (3) with respect to $\beta_k$ or $\gamma_k$ and replacing $\lambda_{kl}$ by the right hand side of (4) or (5). Finally, we update $\sigma_j^2$ by $\sigma_j^2 = \frac{\sum_{i=1}^n \hat{E}(b_{ij}^2)}{n}$ for $j = 1, 2$.

In the E-step, we evaluate the conditional expectation of $W_{ikl}$ ($k = 1, \ldots, K_1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) and the other terms of $b_i$ given the observed data $\tilde{O}_i$ for $i = 1, \ldots, n$. Specifically, the conditional expectation of $W_{ikl}$ ($k = 1, \ldots, K_1; l = 1, \ldots, m_k, t_{kl} \leq R_{ik}^*$) given $\tilde{O}_i$ and $b_i$ is

$$
I(L_{ik} < t_{kl} \leq R_{ik} < \infty) \frac{\lambda_{kl} \exp(\beta^T X_{ikl} + b_{i1})}{1 - \exp\left(-\sum_{L_{ik} < t_{kl} \leq R_{ik}} \lambda_{kl} \exp(\beta^T X_{ikl} + b_{i1})\right)}.
$$

Note that the density of $b_i$ given $\tilde{O}_i$ is proportional to $\{\prod_{k=1}^{K_1} g_{ik}^{(1)}(b_{i1}; \beta, \lambda_k)\} \times \{\prod_{k=K_1+1}^K g_{ik}^{(2)}(b_i; \beta, \lambda_k)\}$. We evaluate the conditional expectation of $W_{ikl}$ and the other terms through numerical integration over $b_i$ with Gauss-Hermite quadratures.

We iterate between the E-step and M-step until convergence. In the M-step, the high-dimensional nuisance parameters $\lambda_{kl}$ ($k = 1, \ldots, K; l = 1, \ldots, m_k$) are calculated explicitly, such that inversion of high-dimensional matrices is avoided. We denote the final estimators for $\theta$ and $A$ as $\hat{\theta} \equiv (\hat{\beta}, \hat{\gamma}, \hat{\Sigma})$ and $\hat{A} \equiv (\hat{\Lambda}_1, \ldots, \hat{\Lambda}_K)$. 

7
2.3. Asymptotic Theory

We establish the asymptotic properties of \((\hat{\theta}, \hat{A})\) under the following regularity conditions.

Condition 1. The true value of \(\theta\), denoted by \(\theta_0 \equiv (\beta_0, \gamma_0, \Sigma_0)\), belongs to the interior of a known compact set \(\Theta \equiv \mathcal{B} \times \mathcal{G} \times \mathcal{S}\), where \(\mathcal{B} \subset \mathbb{R}^p\), \(\mathcal{G} \subset \mathbb{R}^{K_2}\), and \(\mathcal{S} \subset (0, \infty) \times (0, \infty)\).

Condition 2. For \(k = 1, \ldots, K\), the true value \(\Lambda_{k0}(\cdot)\) of \(\Lambda_k(\cdot)\) is strictly increasing and continuously differentiable in \([0, \tau_k]\) with \(\Lambda_{k0}(0) = 0\).

Condition 3. For \(k = 1, \ldots, K_1\), the monitoring times have finite support \(U_k\) with the least upper bound \(\tau_k\). The number of potential monitoring times \(M_k\) is positive with \(E(M_k) < \infty\). There exists a positive constant \(\eta\) such that \(\Pr\{\min_{1 \leq k \leq K_1, 0 \leq m < M_k} (U_{k,m + 1} - U_{km}) \geq \eta | M_k, X_k\} = 1\). In addition, there exists a probability measure \(\mu_k\) in \(U_k\) such that the bivariate distribution function of \((U_{km}, U_{k,m + 1})\) conditional on \((M_k, X_k)\) is dominated by \(\mu_k \times \mu_k\) and its Radon-Nikodym derivative, denoted by \(\tilde{f}_{km}(u,v; M_k, X_k)\), can be expanded to a positive and twice-continuously differentiable function in the set \(\{(u,v) : 0 \leq u \leq \tau_k, 0 \leq v \leq \tau_k, v - u \geq \eta\}\).

Condition 4. For \(k = K_1 + 1, \ldots, K\), let \(\tau_k\) denote the study duration time and \(U_k = [0, \tau_k]\). There exists a positive constant \(\delta\) such that \(\Pr(C_k \geq \tau_k | X_k) = \Pr(C_k = \tau_k | X_k) \geq \delta\) almost surely.

Condition 5. With probability 1, \(X_k(\cdot)\) has bounded total variation in \(U_k\). If there exists a constant vector \(a_1\) and a deterministic function \(a_{2k}(t)\) such that \(a_1^T X_k(t) + a_{2k}(t) = 0\) for any \(t \in U_k\) and any \(k \in \{1, \ldots, K\}\) with probability 1, then \(a_1 = 0\) and \(a_{2k}(t) = 0\) for any \(t \in U_k\) and any \(k \in \{1, \ldots, K\}\).

Remark 2. Conditions 1, 2, and 5 are standard conditions for failure time regression with time-dependent covariates. Condition 3 pertains to the joint distribution of monitoring times of the asymptomatic events. It requires that two adjacent monitoring times are separated by at least \(\eta\); otherwise, the data may contain exact observations, which require a different theoretical treatment. The dominating measure \(\mu_k\) is chosen as the Lebesgue
measure if the monitoring times are continuous random variables and as the counting measure if monitorings occur only at a finite number of time points. The number of potential monitoring times $M_k$ can be fixed or random, is possibly different among study subjects and event types, and is allowed to depend on covariates. Condition 4 implies that there is a positive probability for the $k$th symptomatic event to be observed in the time interval $[0, \tau_k]$.

We state the strong consistency of $(\hat{\theta}, \hat{A})$ and the weak convergence of $\hat{\theta}$ in two theorems.

**Theorem 1.** Under Conditions 1−5, $\|\hat{\theta} - \theta_0\| \to_{a.s.} 0$, and $\|\hat{\Lambda}_k - \Lambda_{k0}\|_{l^\infty(U_k)} \to_{a.s.} 0$, where $\|\cdot\|_{l^\infty(U_k)}$ denotes the supremum norm on $U_k$ for $k = 1, \ldots, K$.

**Theorem 2.** Under Conditions 1−5, $n^{1/2}(\hat{\theta} - \theta_0)$ converges weakly to a $(p + K_2 + 2)$-dimensional zero-mean normal random vector with a covariance matrix that attains the semiparametric efficiency bound.

The proofs of all theorems are provided in the Appendix.

We propose two approaches to estimate the covariance matrix of $\hat{\theta}$. The first approach makes use of the profile likelihood (Murphy and Van der Vaart 2000). Specifically, we define the profile log-likelihood function

$$pl_n(\theta) = \max_{A \in C_1 \times \cdots \times C_K} \log L_n(\theta, A),$$

where $C_k$ is the set of step functions with non-negative jumps at $t_{kl}$ ($k = 1, \ldots, K; l = 1, \ldots, m_k$). We estimate the covariance matrix of $\hat{\theta}$ by the inverse of

$$\sum_{i=1}^n \left( \begin{array}{c} pl_i(\hat{\theta} + h_n e_1) - pl_i(\hat{\theta}) \\ \vdots \\ pl_i(\hat{\theta} + h_n e_{p+K_2+2}) - pl_i(\hat{\theta}) \\ \end{array} \right)^\otimes 2,$$
where \(pl_i\) is the \(i\)th subject’s contribution to \(pl_n\), \(e_j\) is the \(j\)th canonical vector in \(\mathbb{R}^{p+K_2+2}\), \(a^\otimes 2 = aa^T\), and \(h_n\) is a constant of order \(n^{-1/2}\). To evaluate the profile likelihood, we use the EM algorithm of Section 2.2 but only update \(\Lambda_1, \ldots, \Lambda_K\) in the M-step.

Alternatively, we approximate the asymptotic distribution of \(\hat{\theta}\) by bootstrapping the observations. In particular, we draw a simple random sample of size \(n\) with replacement from the observed data \(\{O_i : i = 1, \ldots, n\}\). Let \(\hat{\theta}^*\) be the estimator of \(\theta\) in the bootstrap sample. The empirical distribution of \(\hat{\theta}^*\) can be used to approximate the distribution of \(\hat{\theta}\). Confidence intervals for \(\theta_0\) can be constructed by the Wald method (with the variance of \(\hat{\theta}^*\)) or from the empirical percentiles of \(\hat{\theta}^*\). The following theorem states the asymptotic properties of \(\hat{\theta}^*\), thereby validating the bootstrap procedure.

**Theorem 3.** Under Conditions 1–5, the conditional distribution of \(n^{1/2}(\hat{\theta}^* - \bar{\theta})\) given the data converges weakly to the asymptotic distribution of \(n^{1/2}(\bar{\theta} - \theta_0)\).

### 2.4. Dynamic Prediction

Given the fitted joint model, we can predict future events by updating the event history. For a subject with covariates \(X\), let \(O(t)\) denote the event history at time \(t > 0\), which includes the interval-censored observations of the asymptomatic events \(\{L_k(t), R_k(t) : k = 1, \ldots, K_1\}\), and the right-censored observations of the symptomatic events \(\{Y_k(t), \Delta_k(t) : k = K_1 + 1, \ldots, K\}\).

If no event history is available, the density of the random effect \(b\) can be estimated by \(\psi(b; \hat{\Sigma})\). We estimate the survival function of \(T_k\), denoted by \(P(T_k \geq t | X)\), by

\[
\int_b s_k(t; X, b) \psi(b; \hat{\Sigma}) db,
\]

where

\[
s_k(t; X, b) = \begin{cases} 
\exp \left\{ - \int_0^t e^{b^T X_k(u) + b_1 d\hat{\Lambda}_k(u)} \right\} & k = 1, \ldots, K_1 \\
\exp \left\{ - \int_0^t e^{b_1 X_k(u) + b_2 d\hat{\Lambda}_k(u)} \right\} & k = K_1 + 1, \ldots, K 
\end{cases}
\]
and the integral is evaluated by numerical integration with Gauss-Hermite quadratures. Here, the function $s_k(t; X, b)$ can be interpreted as the conditional survival probability of $T_k$ at time $t$ given $b$ and $X$.

In the semi-competing risks set-up, where one of the symptomatic events is terminal, it is more meaningful to use the cumulative incidence function to predict the event time of interest. Without loss of generality, we assume the $K$th event is terminal. The cumulative incidence function of the $k$th event ($k = 1, \ldots, K - 1$) is given by

$$P(T_k \leq t, T_k \leq T_K | X) = \int_b \left\{ P(T_k \leq t \leq T_K | X, b) + P(T_k \leq T_K < t | X, b) \right\} \psi(b; \Sigma) db$$

which can be estimated by

$$\int_b \left[ \{ 1 - s_k(t; X, b) \} s_K(t; X, b) \right. + \int_0^t \{ 1 - s_k(u; X, b) \} s_K(u; X, b)e^{\hat{\beta}X_k(u) + \hat{\gamma}b_1 + b_2} \Delta_K(u) \left. \right\} \psi(b; \hat{\Sigma}) db.$$

Here, the function $s_k(t; X, b)$ can be interpreted as the conditional survival probability of $T_k$ at time $t$ given $T_K \geq t$, $b$, and $X$.

At time $t_0 > 0$, we update the posterior density of $b$ given the event history $O(t_0)$ so as to perform dynamic prediction. Note that the posterior density of $b$ is proportional to

$$J(b; t_0, X) \equiv \prod_{k=1}^{K_1} \left\{ s_k(L_k(t_0); X, b) - s_k(R_k(t_0); X, b) \right\}$$

$$\times \prod_{k=K_0+1}^{K} \left( s_k(Y_k(t_0); X, b) \left[ \Delta_k(Y_k(t_0))e^{\hat{\beta}X_k(Y_k(t_0)) + \hat{\gamma}b_1 + b_2} \right] \right) \psi(b; \hat{\Sigma}).$$
If the subject has not developed the $k$th event or the terminal event by time $t_0$, i.e., $Y_k(t_0) = Y_K(t_0) = t_0$ and $\Delta_k(t_0) = \Delta_K(t_0) = 0$, we estimate the conditional cumulative incidence function of the $k$th event, $P(T_k \leq t, T_k \leq T_K | O(t_0), X)$, by

$$\int_b \frac{J(b; t_0, X)}{s_k(t_0; X, b)s_K(t_0; X, b)} \int_{b'} J(b'; t_0, X) db' \left[ \{s_k(t_0; X, b) - s_k(t; X, b)\} s_K(t; X, b) + \int_{t_0}^t \{s_k(t_0; X, b) - s_k(u; X, b)\} s_K(u; X, b)e^{\hat{b}^T x_k(u) + \hat{\gamma}_k b_1 + b_2 \Delta K(u)} \right] db.$$ 

In practice, it is desirable to identify subjects who are at increased risk as the event history is accumulating. In the same vein as the risk score under the standard proportional hazards model, we use the risk score $\hat{\beta}^T x_k(t_0) + \hat{\gamma}_k b_1(t_0) + \hat{b}_2(t_0)$ to dynamically predict the $k$th event ($k = K_1 + 1, \ldots, K$), where $\hat{b}(t_0) \equiv (\hat{b}_1(t_0), \hat{b}_2(t_0))$ is a suitable estimator of $b$ given the event history $O(t_0)$. The estimator $\hat{b}(t_0)$ can be the posterior mean or mode of $b$ or an imputed value from the posterior distribution. For example, the risk score using the posterior mean is given by

$$\hat{\beta}^T x_k(t_0) + \frac{\int_b (\hat{\gamma}_k b_1 + b_2) J(b; t_0, X) db}{\int_b J(b; t_0, X) db}.$$ 

The risk score quantifies the subject-specific risk and can be very useful to both individual patients and clinicians when making decisions about lifestyle modifications and preventive medical treatments.

### 3. SIMULATION STUDIES

We conducted simulation studies to assess the performance of the proposed methods. We considered one time-independent covariate $X_1 \sim \text{Unif}(0, 1)$ and one time-dependent covariate $X_2(t) = I(t \leq V)B_1 + I(t > V)B_2$, where $B_1$ and $B_2$ are independent Bernoulli(0.5), $V \sim \text{Unif}(0, \tau)$, and $\tau = 4$. We considered two asymptomatic events and two symptomatic events. We set $X_k = e_k \otimes (X_1, X_2)^T$, where $e_k$ is the $k$th canonical vector in $\mathbb{R}^4$, and $\otimes$ denotes the Kronecker product. We set $\beta = (0.5, 0.4, 0.5, -0.2, -0.5, 0.5, -0.5, 0.5)^T$. 

12
\[ \Lambda_1(t) = 0.5t, \quad \Lambda_k(t) = \log\{1 + t/(k - 1)\} \]

for \( k = 2, 3, 4 \), \( \gamma_3 = \gamma_4 = 0.25 \), and \( \sigma_1^2 = \sigma_2^2 = 1 \).

Both symptomatic events were censored by \( C \sim Unif(2\tau/3, \tau) \), such that the censoring rates are 33\% and 39\%, respectively. The series of monitoring times were generated sequentially, with \( U_m = U_{m-1} + 0.1 + Unif(0, 0.5) \) for \( m \geq 1 \) and \( U_0 = 0 \). The last monitoring time is the largest \( U_m \) that is smaller than \( C \). We set \( n = 100 \) or 200 and simulated 2000 replicates. For each dataset, we applied the proposed EM algorithm by setting the initial value of \( \beta \) to 0, the initial values of \( \gamma_k \) and \( \sigma_k^2 \) to 1 and the initial value of \( \lambda_{kl} \) to \( 1/m_k \).

We used 20 quadrature points for integration with respect to each random effect and set the convergence threshold to \( 10^{-3} \). For variance estimation, we set \( h_n = 5n^{-1/2} \) for profile likelihood and used 100 bootstrap samples.

Table 1 summarizes the simulation results. The biases for all parameter estimators are small, especially for \( n = 200 \). Both the profile-likelihood and bootstrap variance estimators for \( \hat{\beta} \) are accurate, especially for \( n = 200 \). Both variance estimators for \( \hat{\gamma} \) tend to overestimate the true variabilities, but the coverage probabilities of the confidence intervals get closer to the nominal level as sample size increases. The profile-likelihood variance estimators for \( \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \) overestimate the true variabilities, while the bootstrap variance estimators for \( \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \) accurately reflect the true variabilities. Figure S.1(a) of the Supplemental Materials shows the estimation of the baseline survival functions with sample size \( n = 200 \). The estimators are virtually unbiased.

We considered a second setup with an additional terminal event. We set \( X_k = e_k \otimes (X_1, X_2)^T \), where \( e_k \) is the \( k \)th canonical vector in \( \mathbb{R}^5 \). In addition, we set \( \beta = (0.5, 0.4, 0.5, -0.2, -0.5, 0.5, -0.5, 0.5, 0.3, -0.2)^T \), \( \Lambda_5(t) = \log(1 + t/4) \), and \( \gamma_5 = 0.25 \). The terminal event was also censored by \( C \). The censoring rates for the right-censored events are 51\%, 58\%, and 43\%, respectively. The results are shown in Table S.1 and Figure S.1(b) of the Supplemental Materials. The conclusions are similar to the case of no terminal event.

We assessed the performance of dynamic prediction based on the conditional cumulative
### Table 1. Summary Statistics for the Simulation Studies Without a Terminal Event

<table>
<thead>
<tr>
<th></th>
<th>$n = 100$</th>
<th></th>
<th>$n = 200$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>SE</td>
<td>SEE</td>
<td>CP</td>
</tr>
<tr>
<td></td>
<td>Beta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.006</td>
<td>0.585</td>
<td>0.597</td>
<td>0.961</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.029</td>
<td>0.327</td>
<td>0.321</td>
<td>0.941</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.015</td>
<td>0.623</td>
<td>0.609</td>
<td>0.946</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>-0.005</td>
<td>0.341</td>
<td>0.329</td>
<td>0.940</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>-0.022</td>
<td>0.617</td>
<td>0.635</td>
<td>0.957</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>-0.002</td>
<td>0.319</td>
<td>0.338</td>
<td>0.965</td>
</tr>
<tr>
<td>$\beta_{41}$</td>
<td>-0.012</td>
<td>0.623</td>
<td>0.651</td>
<td>0.969</td>
</tr>
<tr>
<td>$\beta_{42}$</td>
<td>0.004</td>
<td>0.330</td>
<td>0.348</td>
<td>0.967</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.012</td>
<td>0.227</td>
<td>0.252</td>
<td>0.979</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>-0.013</td>
<td>0.237</td>
<td>0.260</td>
<td>0.976</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>0.062</td>
<td>0.445</td>
<td>0.751</td>
<td>0.978</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>-0.102</td>
<td>0.413</td>
<td>0.510</td>
<td>0.993</td>
</tr>
</tbody>
</table>

NOTE: SE and SEE denote, respectively, the empirical standard error and mean standard error estimator. CP stands for the empirical coverage probability of the 95% confidence interval based on the Wald method for the profile-likelihood approach and the 95% symmetric confidence interval for the bootstrap approach. For $\gamma_3, \gamma_4, \sigma_1^2,$ and $\sigma_2^2$, bias and SEE are based on the median instead of the mean, and SE is based on the mean absolute deviation. For $\sigma_1^2$ and $\sigma_2^2$, the confidence intervals are based on the log transformation.
Figure 1. Estimation of the baseline cumulative incidence function conditional on the event history. The solid black curve, dotted blue curve, and dashed red curve pertain, respectively, to the true value and the mean estimates from the proposed method with \( n = 100 \) and \( n = 200 \).

incidence function in the setting with a terminal event. Suppose that at the first monitoring time \( t_0 = 1 \), event 2 has occurred but events 1, 3, and 4 have not. Figure 1 shows the estimation of the baseline cumulative incidence functions (pertaining to \( X = 0 \)) for events 3 and 4 given the event history at time \( t_0 = 1 \). The estimators slightly underestimate the true values at the right tail, but the biases get smaller as \( n \) increases.

To investigate the performance of the proposed dynamic prediction methods under misspecified models, we conducted another set of simulation studies where the event times were generated from the proportional odds, instead of the proportional hazards, models with random effects. As shown in Section S.1 of the Supplemental Materials, the dynamic prediction is still quite accurate.
4. ARIC STUDY

ARIC is a prospective epidemiological cohort study conducted in four U.S. communities: Forsyth County, NC; Jackson, MS; Minneapolis, MN; and Washington County, MD. A total of 15792 participants received a baseline examination between 1987 and 1989 and four subsequent examinations in 1990-1992, 1993-1995, 1996-1998, and 2011-2013. At each examination, medical data were collected, such that interval-censored observations for diabetes and hypertension were obtained. The participants were also followed for cardiovascular diseases through reviews of hospital records, such that potentially right-censored observations on MI, stroke, and death were collected.

We related the disease incidence to race, sex, and five baseline risk factors: age, body mass index (BMI), glucose level, systolic blood pressure, and smoking status. Since the Jackson cohort is composed of black subjects only, and neither Minneapolis nor Washington County cohorts contain black subjects, we included the cohort×race indicators as predictors. We excluded subjects with prevalent cases at baseline or missing covariate values to obtain a total of 8728 subjects. During the study, 17.3%, 46.8%, 8.3%, and 5.1% of the subjects developed diabetes, hypertension, MI, and stroke, respectively, while 28.7% died.

We jointly modeled the asymptomatic and symptomatic events in the ARIC study with equations (1) and (2). For variance estimation, we used the profile likelihood approach with $h_n = 5n^{-1/2}$. Tables 2 and 3 show the estimation results for the regression parameters. Several characteristics and baseline risk factors are found to be predictive of the events. Older subjects have higher risks of hypertension, MI, stroke, and death than younger subjects. Males have lower risk of hypertension but higher risks of MI, stroke, and death than females. Smokers have significantly higher risks for all events than non-smokers. In addition, higher baseline BMI increases the risks of diabetes, hypertension, and MI; higher baseline glucose level increases the risks of diabetes, stroke, and death; and higher baseline
value of systolic blood pressure increases the risks of all considered events.

The estimation results for the remaining parametric components are shown in Table S.2 of the Supplemental Materials. The variance components $\sigma_1^2$ and $\sigma_2^2$ are significantly larger than zero, indicating strong correlation among the asymptomatic events and among the symptomatic events. The parameters $\gamma_{\text{MI}}$, $\gamma_{\text{Stroke}}$, and $\gamma_{\text{Death}}$ are also significantly larger than zero, reflecting the strong positive dependence of the symptomatic events on the asymptomatic events. The Akaike information criterion (AIC) for the proposed model is 108852.8. For comparisons, we also fit a model with one random effect shared by all events. The corresponding AIC is 109000.6, and the $p$-value for the likelihood ratio test is less than 0.0001, indicating that the proposed model provides a much better fit to the data than the model with one shared random effect.

To evaluate the performance of the proposed prediction methods, we randomly divided the study cohort into training and testing sets with equal numbers of subjects. We analyzed the training set to obtain parameter estimates, based on which we calculated the risk scores for subjects in the testing set, where the posterior means of the random effects were used. Specifically, at examinations 2, 3, and 4, we calculated the risk scores of MI or stroke for subjects who have not developed the disease. We evaluated the performance of the prediction using C-index (Uno et al. 2011) and compared it with that of the risk scores based on the standard models. In particular, for MI and stroke, we considered the univariate model of Fine and Gray (1999) with death as a competing risk; for death, we considered the standard proportional hazards model. The values of the C-index based on twenty randomly divided training/test tests are shown in Figure 2. The proposed risk score performs better than the risk score of the standard model at all examinations for all symptomatic events.

Figure 3 shows the estimated conditional cumulative incidence functions of MI and stroke for two smokers and two non-smokers who have different event histories at year 3 but with the same values of other risk factors. The risks of MI and stroke are consid-
erably higher for the smokers than the non-smokers with the same event history. The estimated conditional probabilities for the subjects who have developed both diabetes and hypertension are higher than those who have not developed diabetes or hypertension.

Table 2. Estimation Results for the Regression Parameters of the Asymptomatic Events in the ARIC Study

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Diabetes</th>
<th>Hypertension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std error</td>
</tr>
<tr>
<td>Forsyth County, white</td>
<td>−0.5332</td>
<td>0.1817</td>
</tr>
<tr>
<td>Jackson, black</td>
<td>−0.1356</td>
<td>0.1806</td>
</tr>
<tr>
<td>Minneapolis, white</td>
<td>−0.9415</td>
<td>0.1802</td>
</tr>
<tr>
<td>Washington County, white</td>
<td>−0.3778</td>
<td>0.1778</td>
</tr>
<tr>
<td>Age</td>
<td>−0.0093</td>
<td>0.0057</td>
</tr>
<tr>
<td>Male</td>
<td>−0.0655</td>
<td>0.0593</td>
</tr>
<tr>
<td>BMI</td>
<td>0.0911</td>
<td>0.0059</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.1075</td>
<td>0.0033</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>0.0096</td>
<td>0.0026</td>
</tr>
<tr>
<td>Smoker</td>
<td>0.4576</td>
<td>0.0674</td>
</tr>
</tbody>
</table>

NOTE: The blacks in Forsyth County form the reference group for the cohort × race variables.
Table 3. Estimation Results for the Regression Parameters of the Symptomatic Events in the ARIC Study

<table>
<thead>
<tr>
<th>Covariate</th>
<th>MI</th>
<th></th>
<th></th>
<th>Stroke</th>
<th></th>
<th></th>
<th>Death</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std error</td>
<td>p-value</td>
<td>Estimate</td>
<td>Std error</td>
<td>p-value</td>
<td>Estimate</td>
<td>Std error</td>
<td>p-value</td>
<td>Estimate</td>
<td>Std error</td>
<td>p-value</td>
</tr>
<tr>
<td>Forsyth County, white</td>
<td>0.0467</td>
<td>0.2477</td>
<td>0.8504</td>
<td>0.1308</td>
<td>0.3688</td>
<td>0.7228</td>
<td>−0.2475</td>
<td>0.1049</td>
<td>0.0183</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackson, black</td>
<td>−0.3121</td>
<td>0.2681</td>
<td>0.2444</td>
<td>0.6622</td>
<td>0.3755</td>
<td>0.0778</td>
<td>0.1871</td>
<td>0.1118</td>
<td>0.0941</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minneapolis, white</td>
<td>−0.1052</td>
<td>0.2476</td>
<td>0.6710</td>
<td>0.0507</td>
<td>0.3688</td>
<td>0.8907</td>
<td>−0.3262</td>
<td>0.1040</td>
<td>0.0017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington County, white</td>
<td>0.1953</td>
<td>0.2457</td>
<td>0.4266</td>
<td>0.5013</td>
<td>0.3653</td>
<td>0.1700</td>
<td>−0.1194</td>
<td>0.1032</td>
<td>0.2471</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0805</td>
<td>0.0078</td>
<td>&lt;0.0001</td>
<td>0.1121</td>
<td>0.0099</td>
<td>&lt;0.0001</td>
<td>0.1465</td>
<td>0.0054</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.9279</td>
<td>0.0901</td>
<td>&lt;0.0001</td>
<td>0.4050</td>
<td>0.1071</td>
<td>0.0002</td>
<td>0.6108</td>
<td>0.0545</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.0273</td>
<td>0.0101</td>
<td>0.0068</td>
<td>−0.0010</td>
<td>0.0123</td>
<td>0.9356</td>
<td>0.0080</td>
<td>0.0060</td>
<td>0.1847</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>0.0059</td>
<td>0.0046</td>
<td>0.2007</td>
<td>0.0215</td>
<td>0.0057</td>
<td>0.0002</td>
<td>0.0104</td>
<td>0.0030</td>
<td>0.0006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>0.0135</td>
<td>0.0036</td>
<td>0.0002</td>
<td>0.0192</td>
<td>0.0047</td>
<td>&lt;0.0001</td>
<td>0.0089</td>
<td>0.0022</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoker</td>
<td>1.2378</td>
<td>0.0888</td>
<td>&lt;0.0001</td>
<td>1.0023</td>
<td>0.1127</td>
<td>&lt;0.0001</td>
<td>1.3045</td>
<td>0.0599</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: See the Note to Table 2.
Figure 2. Boxplots of the estimates of the C-index at each examination in the ARIC study. The red boxes pertain to the univariate model of Fine and Gray (1999) for MI and stroke and the standard proportional hazards model for death. The blue boxes pertain to the proposed joint model.

Figures S.2(a) and S.2(b) in the Supplementary Materials illustrate the estimation of the conditional cumulative incidence functions of stroke given different event histories. We estimated the cumulative incidence functions at time zero when only baseline covariates are available and then updated them at two examinations at year 3 and year 6 using the event histories. The development of diabetes, hypertension, and MI substantially increases the incidence of stroke, whereas the history of no diabetes, hypertension, or MI over the first six years entails lower incidence of stroke. For comparison, we show in Figures S.2(c) in the Supplementary Materials the estimated cumulative incidence function of stroke under the univariate model of Fine and Gray (1999), which does not condition on the event history and thus reflects the population average. This estimate lies between the two previous conditional estimates, as expected.
Figure 3. Estimation of the conditional cumulative incidence functions of MI and stroke for a 50-year-old white female residing in Forsyth County, NC, with BMI 40 kg/m$^2$, glucose 98 mg/dl, and systolic blood pressure 113 mmHg. The solid curves pertain to smokers, while the dashed curves pertain to non-smokers. The black curves pertain to subjects who have not developed diabetes or hypertension by year 3. The red curves pertain to subjects who have developed both diabetes and hypertension by year 3.

5. DISCUSSION

In this paper, we formulated the joint distribution of multiple right- and interval-censored events with proportional hazards models with random effects. We characterized the correlation structure of the asymptomatic and symptomatic events through two independent random effects and used unknown coefficients to capture the effects of the asymptomatic events on the symptomatic events. To our knowledge, no such modelling approach has been previously adopted.

We studied efficient nonparametric maximum likelihood estimation of the proposed joint model and established the asymptotic properties of the estimators through innovative use of modern empirical process theory. We showed the Glivenko-Cantelli and Donsker
properties for the classes of functions of interest by carefully evaluating their bracketing numbers. The estimators of the cumulative baseline hazard functions for the symptomatic and asymptomatic events converge at different \((n^{1/2} \text{ and } n^{1/3})\) rates, such that separate treatments were required in the proofs.

The proposed EM algorithm performed well in both the simulation studies and the real example. There was no occurrence of nonconvergence in any of the simulated or real dataset. It took 2.5 or 12 minutes to analyze a simulated dataset with \(K = 5\) events and sample sizes \(n = 100\) or \(200\), respectively. It took ten days to analyze the ARIC data, which involves 8728 subjects with 10 covariates and 2232, 2291, 701, 431, and 2130 distinct jump times for diabetes, hypertension, MI, stroke, and death, respectively. We can alleviate the computational burden for such large studies by grouping or subsampling the examination times so as to reduce the number of distinct time points. In particular, the computing time was shortened to two days when the distinct values were reduced to 154, 162, 276, 229, and 311 by rounding the examination times to the nearest months in the ARIC data.

We proposed nonparametric bootstrap for variance estimation as an alternative to the conventional profile-likelihood approach. We established the validity of the bootstrap procedure and showed through simulation studies that bootstrap yields more accurate estimators of the variabilities for the variance components. To our knowledge, bootstrap with interval-censored data has not been rigorously studied. In large studies, bootstrap may be overly time-consuming. It would be worthwhile to develop other versions of bootstrap, such as subsampling bootstrap, to reduce computational burden.

In models (1) and (2), we distinguish asymptomatic from symptomatic events when modeling the correlation structures because it is of particular interest to study the effects of asymptomatic diseases, which are typically interval-censored, on symptomatic diseases, which are typically right-censored, and to use the former to predict the latter. We show in Section S.2 of the Supplementary Materials that our framework can be modified to allow
any of the $K$ event times to be interval- or right-censored.

ARIC is one of many epidemiological cohort studies with multiple symptomatic and asymptomatic events. Such events are also available in electronic health records. Indeed, other types of outcomes, such as longitudinal repeated measures and recurrent events, may also be available. The proposed joint model can be extended to accommodate additional multivariate outcomes and improve dynamic prediction.

**APPENDIX: PROOFS OF ASYMPTOTIC RESULTS**

Let $P_n$ denote the empirical measure for $n$ independent subjects, $P$ denote the true probability measure, and $G_n \equiv \sqrt{n}(P_n - P)$ denote the empirical process. The proofs of Theorems 1, 2, and 3 make use of five lemmas, which are stated and proved in Section S.3 of the Supplemental Materials.

**A.1 Proof of Theorem 1**

We first show the existence of the estimator $(\hat{\theta}, \hat{A})$. Let $\hat{M} = \sum_{k=1}^{K} \sup_{t \in U_k} \sup_{X_k(t), \beta} |\beta^T X_k(t)| + \sum_{k=K_1+1}^{K} |\gamma_k|$. For any $(\theta, A)$ in the parameter space, the integrand in the $i$th term of $l_n(\theta, A)$ is bounded by

$$O(1) \prod_{k=K_1+1}^{K} \left[ \left( \Lambda_k \{Y_{ik}\} e^{\hat{M}(b_i)} |\Delta_{ik}| \right)^{\Delta_{ik}} \left\{ 1 + \int_0^{Y_{ik}} e^{\beta^T X_{ik}(s) + \gamma_k b_{i1} + b_{i2} d \Lambda_k(s)} \right\}^{\Delta_{ik}} \psi(b_i; \Sigma) \right].$$

Thus, $l_n(\theta, A)$ attains the maximum for finite values of $\Lambda_k$ for $k = K_1 + 1, \ldots, K$, so the estimator $(\hat{\theta}, \hat{A})$ exists by allowing $\hat{\Lambda}_k(\tau_k) = \infty$ for $k = 1, \ldots, K_1$.

We shall prove that $\limsup_n \hat{\Lambda}_k(\tau_k - \epsilon) < \infty$ with probability 1 for any $\epsilon > 0$ and $k = 1, \ldots, K_1$ and that $\limsup_n \hat{\Lambda}_k(\tau_k) < \infty$ with probability 1 for $k = K_1 + 1, \ldots, K$. By definition, $l_n(\hat{\theta}, \hat{A}) - l_n(\theta, A) \geq 0$ for any $(\theta, A)$ in the parameter space. We wish to show that if $\limsup_n \hat{\Lambda}_k(\tau_k - \epsilon) = \infty$ for some $\epsilon > 0$ for $k = 1, \ldots, K_1$ or $\limsup_n \hat{\Lambda}_k(\tau_k) = \infty$
for $k = K_1 + 1, \ldots, K$, then this difference must be negative, which is a contradiction. The key is to construct a suitable function in the parameter space that converges uniformly to $A_0$.

For $k = 1, \ldots, K_1$, we define the step function $\widetilde{\Lambda}_k$ with $\widetilde{\Lambda}_k(t) = \Lambda_{k0}(t)$ for $t = t_{k1}, \ldots, t_{km}$ such that it converges uniformly to $\Lambda_{k0}$. For $k = K_1 + 1, \ldots, K$, we construct function $\widetilde{\Lambda}_k$ by imitating $\Lambda_k$. Specifically, by differentiating $l_n(\theta, A)$ with respect to $\Lambda_k\{Y_{ik}\}$ and setting the derivative to 0, we find that $\widetilde{\Lambda}_k$ satisfies the equation

$$\Delta_{ik} = \sum_{j=1}^{n} \frac{J_1(b, O_j; \beta, \gamma, \tilde{A}) J_2k(Y_{ik}, b, O_j; \beta, \gamma_k) \phi(b; \Sigma) db}{\int_b J_1(b, O_j; \beta, \gamma, \tilde{A}) \phi(b; \Sigma) db},$$

(A.1)

where

$$J_1(b, O_j; \beta, \gamma, \tilde{A}) = \prod_{k=1}^{K_1} \left[ \exp \left\{ - \int_0^{L_k} e^{\beta^T X_k(s) + b_1} d\Lambda_k(s) \right\} \right] - \exp \left\{ - \int_0^{R_k} e^{\beta^T X_k(s) + b_1} d\Lambda_k(s) \right\} \right]$$

and

$$J_{2k}(Y_{ik}, b, O_j; \beta, \gamma_k) = I(Y_k \geq t) e^{\beta^T X_k(t) + \gamma_k b_1 + b_2}.$$  

We replace $\tilde{\theta}$ and $\tilde{A}$ on the right side of equation (A.1) by $\theta_0$ and $A_0$, respectively, to obtain a similar function. We denote the solution as $\tilde{\Lambda}_k$. By the Glivenko-Cantelli result in Lemma 1, $\widetilde{\Lambda}_k$ converges uniformly to $\Lambda_{k0}$ in $U_k$ for $k = K_1 + 1, \ldots, K$. We denote $\tilde{A} = (\tilde{\Lambda}_1, \ldots, \tilde{\Lambda}_K)$.

Clearly, $n^{-1} \left\{ l_n(\tilde{\theta}, \tilde{A}) - l_n(\theta_0, \tilde{A}) \right\} \geq 0$. Let $\delta_{ikm} = I(U_{ikm} < T_{ik} \leq U_{ik,m+1})$ for $i = 1, \ldots, n$, $k = 1, \ldots, K_1$, and $m = 0, \ldots, M_{ik}$, where $U_{ik,M_{ik}+1} = \infty$. By the fact that $e^{-|x|}(1 + y) \leq 1 + e^{xy} \leq e^{x|x|}(1 + y)$, we obtain

$$0 \leq n^{-1} l_n(\tilde{\theta}, \tilde{A}) - n^{-1} l_n(\theta_0, \tilde{A}) \leq O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K_1+1}^{K} \log \left( n\tilde{\Lambda}_k\{Y_{ik}\} \right)$$

24
\[ + n^{-1} \sum_{i=1}^{n} \left\lfloor \log \int_{b}^{k=K+1} \frac{e^{b^T X_{ik}(Y_{ik}) + \gamma_b b_1 + b_2}}{1 + \int_{t}^{Y_{ik}} e^{b^T X_{ik}(t) + \gamma_b b_1 + b_2} d\Lambda_k(t)} \right \rfloor_{\Delta k} \phi \left( b; \tilde{\Sigma} \right) db \]

\[ \leq O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K+1}^{K} \log \left( n\Lambda_k \{ Y_{ik} \} \right) \]

\[ + n^{-1} \sum_{i=1}^{n} \left( \log \int_{b}^{k=K+1} \frac{e^{\tilde{M} \| b \|}}{e^{-M \| b \|} \left\{ 1 + \Lambda_k (Y_{ik}) \right\} \right)_{\Delta k} \phi \left( b; \tilde{\Sigma} \right) db \]

\[ \leq O(1) + n^{-1} \sum_{i=1}^{n} \sum_{k=K+1}^{K} \log \left( n\Lambda_k \{ Y_{ik} \} \right) - n^{-1} \sum_{i=1}^{n} \sum_{k=K+1}^{K} \left[ \Delta k \log \left\{ 1 + \Lambda_k (Y_{ik}) \right\} \right] \]

We first show that \( \limsup_n \Lambda_k (\tau_k) < \infty \) using the partitioning idea of Murphy (1994). Specifically, we construct a sequence \( u_k = \tau_k > u_{k+1} > \cdots > u_{k,Q_k} = 0 \). Then,

\[ n^{-1} \sum_{i=1}^{n} \sum_{k=K+1}^{K} \log \left( n\Lambda_k \{ Y_{ik} \} \right) - n^{-1} \sum_{i=1}^{n} \sum_{k=K+1}^{K} \left[ \Delta k \log \left\{ 1 + \Lambda_k (Y_{ik}) \right\} \right] \]

\[ \leq O(1) + \sum_{k=K+1}^{K} \sum_{q=0}^{Q_k} \sum_{i=1}^{n} I(Y_{ik} = \tau_k) \log \left( n\Lambda_k \{ Y_{ik} \} \right) - \sum_{k=K+1}^{K} \sum_{q=0}^{Q_k} \sum_{i=1}^{n} \Delta k I(Y_{ik} = \tau_k) \log \left\{ 1 + \Lambda_k (\tau_k) \right\} \]

\[ - \sum_{k=K+1}^{K} \sum_{q=0}^{Q_k} \sum_{i=1}^{n} \Delta k I(Y_{ik} \in [u_{k,q+1}, u_{k,q}]) \log \left\{ 1 + \Lambda_k (u_{k,q+1}) \right\} \]

which is further bounded by

\[ - (2n)^{-1} \sum_{i=1}^{n} \sum_{k=K+1}^{K} \Delta k I(Y_{ik} = \tau_k) \log \left\{ 1 + \Lambda_k (\tau_k) \right\} \]

\[ - \sum_{k=K+1}^{K} \left\{ (2n)^{-1} \sum_{i=1}^{n} \Delta k I(Y_{ik} = \tau_k) - n^{-1} \sum_{i=1}^{n} \Delta k I(Y_{ik} \in [u_1, u_0]) \right\} \log \left\{ 1 + \Lambda_k (\tau_k) \right\} \]

\[ - \sum_{k=K+1}^{K} \sum_{q=1}^{Q_k} \left\{ n^{-1} \sum_{i=1}^{n} \Delta k I(Y_{ik} \in [u_{k,q}, u_{k,q-1}]) - n^{-1} \sum_{i=1}^{n} \Delta k I(Y_{ik} \in [u_{k,q+1}, u_{k,q}]) \right\} \]

\[ \times \log \left\{ 1 + \Lambda_k (u_{k,q}) \right\} \].
Note that \( u_{kq} \) is chosen such that the coefficients in front of \( \log \{1 + \hat{\Lambda}_k(u_{kq})\} \) are all negative when \( n \) is large enough. Thus, the corresponding terms cannot diverge to \( \infty \). However, if \( \hat{\Lambda}_k(\tau_k) \) diverges to \( \infty \), then the first term diverges to \( -\infty \). We conclude that there exists some \( M^* < \infty \) such that \( \max_{K_1+1 \leq k \leq K} \limsup_n \hat{\Lambda}_k(\tau_k) \leq M^* \) for \( k = K_1 + 1, \ldots, K \).

We denote \( \tilde{\Lambda}^* = (\tilde{\Lambda}_1, \ldots, \tilde{\Lambda}_{K_1}, \tilde{\Lambda}_{K_1+1}, \ldots, \tilde{\Lambda}_K) \). Then,

\[
0 \leq n^{-1} l_n(\tilde{\theta}, \tilde{A}) - n^{-1} l_n(\theta_0, \tilde{A}^*) \\
\leq O(1) + n^{-1} \sum_{i=1}^{K_1} \left( \log \int_{b_k=1}^{K_1} \left[ \exp \left\{ -e^{-\hat{\gamma}_k(b_kl_{ik,M_{ik}})} \right\} \delta_{ik,M_{ik}} \phi \left( b; \hat{\Sigma} \right) db \right) \right) \\
\leq O(1) + n^{-1} \sum_{i=1}^{K_1} \left( \log \int_{||b|| \leq 1} \prod_{k=1}^{K_1} \left[ \exp \left\{ -e^{-\hat{\gamma}_k(b_kl_{ik,M_{ik}})} \right\} \delta_{ik,M_{ik}} \phi \left( b; \hat{\Sigma} \right) db \right) \right) \\
+n^{-1} \sum_{i=1}^{n} \left\{ \log \int_{||b|| > 1} \phi \left( b; \hat{\Sigma} \right) db \right\} \\
\leq O(1) - n^{-1} \sum_{i=1}^{K_1} \sum_{k=1}^{K_1} \delta_{ik,M_{ik}} e^{\hat{\gamma}_k(U_{ik,M_{ik}})}.
\]

Therefore, for \( k = 1, \ldots, K_1 \), \( \limsup_n \hat{\Lambda}_k(\tau_k - \epsilon) < \infty \) with probability 1 for any \( \epsilon > 0 \). By choosing a sequence of \( \epsilon \) decreasing to 0, it then follows from Helly’s selection lemma that along a subsequence, \( \hat{\Lambda}_k \to \Lambda_{k*} \) pointwise on any interior set of \( \mathcal{U}_k \) and \( \hat{\theta} \to \theta_* \equiv (\beta_*, \gamma_*) \).

We denote \( \mathcal{A}_* = (\Lambda_{1*}, \ldots, \Lambda_{K*}) \).

We now show that \( \theta_* = \theta_0 \) and \( \mathcal{A}_* = \mathcal{A}_0 \). First, we consider the differentiability of \( \Lambda_{k*} \) for \( k = K_1 + 1, \ldots, K \). By the definition of \( \tilde{\Lambda}_k \), \( \tilde{\Lambda}_k(t) \) is absolutely continuous with respect to \( \tilde{\Lambda}(t) \), and

\[
\tilde{\Lambda}_k(t) = \int_0^t \frac{P_n \nu_k(s, \mathcal{O}; \theta_0, \mathcal{A}_0)}{P_n \nu_k(s, \mathcal{O}; \tilde{\theta}, \tilde{A})} d\tilde{\Lambda}_k(s), \quad (A.2)
\]

where

\[
\nu_k(t, \mathcal{O}; \theta, \mathcal{A}) = \frac{\int_{b} J_1(b, \mathcal{O}; \beta, \gamma, \mathcal{A}) J_{2k}(t, b, \mathcal{O}; \beta, \gamma, \mathcal{A}) \psi(b; \Sigma) db}{\int_{b} J_1(b, \mathcal{O}; \beta, \gamma, \mathcal{A}) \psi(b; \Sigma) db}.
\]

To take limits on the two sides of equation (A.2), we first show that the denominator of the integrand is uniformly bounded away from zero. It follows from the Glivenko-Cantelli
property in Lemma 1 that
\[
\sup_{t \in U_k} \left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_0, \mathcal{A}_0) - \mathbb{P} \nu_k(t, \mathcal{O}; \theta_0, \mathcal{A}_0) \right| \to_{a.s.} 0
\]
and
\[
\sup_{t \in U_k} \left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \hat{\theta}, \hat{\mathcal{A}}) - \mathbb{P} \nu_k(t, \mathcal{O}; \theta_*, \mathcal{A}_*) \right| \to_{a.s.} 0.
\]

Note that for any \( \epsilon > 0 \),
\[
\lim_{n \to \infty} \Lambda_k(t_k) \geq \int_{0}^{t_k} \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_0, \mathcal{A}_0) \frac{\epsilon}{\epsilon + \left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_0, \mathcal{A}_0) \right|} d\Lambda_{k0}(s).
\]

Let \( \epsilon \to 0 \). By the Monotone Convergence Theorem,
\[
\int_{0}^{t_k} \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_0, \mathcal{A}_0) \frac{\epsilon}{\epsilon + \left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_0, \mathcal{A}_0) \right|} d\Lambda_{k0}(t) < \infty.
\]

We claim that \( \min_{t \in U_k} \left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_*, \mathcal{A}_*) \right| > 0 \). If this inequality does not hold, then there exists some \( t_* \in U_k \) such that \( \mathbb{P}_n \nu_k(t_*, \mathcal{O}; \theta_*, \mathcal{A}_*) = 0 \). The function \( \mathbb{P}_n \nu_k(t_*, \mathcal{O}; \theta_*, \mathcal{A}_*) \) is right-differentiable almost everywhere. Thus, there exists \( \delta > 0 \) such that for \( t \in (t_*, t_* + \delta) \),
\[
\left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_*, \mathcal{A}_*) \right| = \left| \mathbb{P}_n \nu_k(t, \mathcal{O}; \theta_*, \mathcal{A}_*) - \mathbb{P}_n \nu_k(t_*, \mathcal{O}; \theta_*, \mathcal{A}_*) \right| \leq O(1) |t - t_*|
\]
almost everywhere. Hence,
\[
\int_{t_*}^{t_* + \delta} \frac{1}{|t - t_*|} d\Lambda_{k0}(t) < \infty,
\]
which is a contradiction. By taking the limits on both sides of (A.2), we conclude that \( \Lambda_{k*}(t) \) is absolutely continuous with respect to \( \Lambda_{k0}(t) \), so that \( \Lambda_{k*}(t) \) is differentiable with respect to \( t \). In addition, \( d\Lambda_k(t)/d\Lambda_k(t) \) converges to \( d\Lambda_{k*}(t)/d\Lambda_{k0}(t) \) uniformly in \( t \).

Define
\[
m(\theta, \mathcal{A}) = \log \left\{ \frac{L(\theta, \mathcal{A}) + L(\theta_0, \mathcal{A})}{2} \right\}
\]
and
\[
\mathcal{M} = \{ m(\theta, \mathcal{A}) : \theta \in \Theta, \mathcal{A} \in \mathcal{D}_{1,\infty} \times \cdots \times \mathcal{D}_{K_1,\infty} \times \mathcal{D}_{K_1+1,\infty} \times \cdots \times \mathcal{D}_{K,M} \},
\]

27
where $L(\theta, A)$ is the objective function for a single subject, and $\mathcal{D}_{k,c} = \{ \Lambda : \Lambda$ is increasing with $\Lambda(0) = 0, \Lambda(\tau_k) \leq c \}$. By the concavity of the log function,

$$\mathbb{P}_nm(\hat{\theta}, \hat{A}) \geq \frac{1}{2} \left\{ \mathbb{P}_n \log L(\hat{\theta}, \hat{A}) + \mathbb{P}_n \log L(\theta_0, \hat{A}) \right\} \geq \mathbb{P}_nl(\theta_0, \hat{A}) = \mathbb{P}_nm(\theta_0, \hat{A}).$$

It follows from Lemma 1 that the class $\mathcal{M}$ is Glivenko-Cantelli. Thus,

$$0 \leq \mathbb{P}_nm\left(\hat{\theta}, \hat{A}\right) - \mathbb{P}_nm\left(\theta_0, \hat{A}\right) = \mathbb{P}\left\{ m\left(\hat{\theta}, \hat{A}\right) - m\left(\theta_0, \hat{A}\right) \right\} + o_P(1)$$

$$= \mathbb{P}\log\left[ \frac{L(\hat{\theta}, \hat{A}) + L(\theta_0, \hat{A})}{2L(\theta_0, \hat{A})} \right] + o_P(1)$$

$$= \mathbb{P}\log\left\{ \frac{1}{2} + \frac{\prod_{k=K_1+1}^K \Lambda_k(Y_k)^{\Delta k} \int_b J_1(b, \mathcal{O}; \hat{\beta}, \hat{\gamma}, \hat{A}) \psi(b; \hat{\Sigma})db}{\prod_{k=K_1+1}^K \Lambda_k(Y_k)^{\Delta k} \int_b J_1(b, \mathcal{O}; \beta_0, \gamma_0, \hat{A}) \psi(b; \Sigma_0)db} \right\} + o_P(1)$$

$$\rightarrow \mathbb{P}\left[ \log\left\{ \frac{1}{2} + \frac{\prod_{k=K_1+1}^K \Lambda_k^*(Y_k)^{\Delta k} \int_b J_1(b, \mathcal{O}; \beta^*, \gamma^*, A^*) \psi(b; \Sigma^*)db}{\prod_{k=K_1+1}^K \Lambda_k^0(Y_k)^{\Delta k} \int_b J_1(b, \mathcal{O}; \beta_0, \gamma_0, A_0) \psi(b; \Sigma_0)db} \right\} \right],$$

such that the negative Kullback-Leibler information is positive. The identifiability result in Lemma 3 implies that $\beta^* = \beta_0$, $\gamma^* = \gamma_0$, $\Sigma^* = \Sigma_0$, and $\Lambda_k^*(t_k) = \Lambda_k^0(t_k)$ for $k = 1, \ldots, K$ and $t_k \in \mathcal{U}_k$. We conclude that $||\hat{\theta} - \theta_0|| \to 0$ and $|\hat{\Lambda}_k(t_k) - \Lambda_k^0(t_k)| \to 0$ for any $t_k \in \mathcal{U}_k$. Because $A_0$ is continuous, $\hat{A}$ converges uniformly to $A_0$ on $\prod_k \mathcal{U}_k$.

### A.2 Proof of Theorem 2

Let

$$H_{1k}(t, u, v, b, \mathcal{O}; \theta, A) = \frac{J_1(b, \mathcal{O}; \beta, \gamma, A)Q_1(t, u, v, b_1, X_k; \beta, \Lambda_k)\psi(b; \Sigma)}{\int_{b'} J_1(b', \mathcal{O}; \beta, \gamma, A)\psi(b'; \Sigma)db'}$$

for $k = 1, \ldots, K_1$, and

$$H_{2k}(t, b, \mathcal{O}; \theta, A) = \frac{J_1(b, \mathcal{O}; \beta, \gamma, A)Q_2(t, Y_k, b, X_k; \beta, \gamma_k)\psi(b; \Sigma)}{\int_{b'} J_1(b', \mathcal{O}; \beta, \gamma, A)\psi(b'; \Sigma)db'}$$

for $k = K_1 + 1, \ldots, K$, where

$$Q_1(t, u, v, b_1, X_k; \beta, \Lambda_k)$$

28
Then, the score function for $\theta$ is

$$l_\theta(\theta, A) = (l_\beta(\theta, A)^T, l_{\gamma_1}(\theta, A)^T, \ldots, l_{\gamma_K}(\theta, A)^T, l_{\sigma_j^2}(\theta, A)^T)^T,$$

where

$$l_\beta(\theta, A) = \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} \int_b H_{1k}(t, U_{km}, U_{k, m+1}, b, \mathcal{O}; \theta, A) dB X_k(t) d\Lambda_k(t),$$

and

$$l_{\gamma_k}(\theta, A) = \Delta_k \int_b \int_{\tau_k} b_1 J_1(b, \mathcal{O}; \beta, \gamma, A) \psi(b; \Sigma) dB b_1 X_k(t) d\Lambda_k(t) - \int_0^{\tau_k} \int_b b_1 H_{2k}(t, b, \mathcal{O}; \theta, A) dB d\Lambda_k(t),$$

$$l_{\sigma_j^2}(\theta, A) = \frac{\int J_1(b, \mathcal{O}; \beta, \gamma, A) \phi_j'(b; \sigma_j^2) \phi(b_{\beta - j}; \sigma_j^2) dB}{\int J_1(b, \mathcal{O}; \beta, \gamma, A) \psi(b; \Sigma) dB}$$

for $j = 1, 2$, and $\phi_j'(b; \sigma_j^2)$ is the derivative of $\phi(b; \sigma_j^2)$ with respect to $\sigma_j^2$. The score operator for $A$ along the submodel $dA_{k,h} = (1 + \epsilon h_1) d\Lambda_1, \ldots, (1 + \epsilon h_K) d\Lambda_K)^T$ for $h = (h_1, \ldots, h_K)$ with $h_k \in L_2(\mu_k)$ for $k = 1, \ldots, K_1$ and $h_k \in BV_1(U_k)$ for $k = K_1 + 1, \ldots, K$ is

$$l_A(\theta, A)(h) = \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \delta_{km} \int_0^{\tau_k} \int_b H_{1k}(t, U_{km}, U_{k, m+1}, b, \mathcal{O}; \theta, A) dB h_k(t) d\Lambda_k(t)$$

and

$$l_{\sigma_j^2}(\theta, A) = \frac{\int J_1(b, \mathcal{O}; \beta, \gamma, A) \phi_j'(b; \sigma_j^2) \phi(b_{\beta - j}; \sigma_j^2) dB}{\int J_1(b, \mathcal{O}; \beta, \gamma, A) \psi(b; \Sigma) dB}$$

where $BV_1(B)$ denotes the set of functions on $B$ with total variation bounded by 1.
Clearly,
\[ G_n \left\{ l_\theta(\hat{\theta}, \hat{A}) \right\} = -\sqrt{n} \mathcal{P} \left\{ l_\theta(\hat{\theta}, \hat{A}) - l_\theta(\theta_0, A_0) \right\}, \]
and
\[ G_n \left\{ l_A(\hat{\theta}, \hat{A})(h) \right\} = -\sqrt{n} \mathcal{P} \left\{ l_A(\hat{\theta}, \hat{A})(h) - l_A(\theta_0, A_0)(h) \right\}. \]

We apply the Taylor series expansions at \((\theta_0, A_0)\) to the right sides of the above two equations. In light of Lemma 5, the second-order terms are bounded by
\[
O_P(1) \sqrt{n} E \left[ \sum_{k=1}^{K_1} \sum_{m=0}^{M_k} \left\{ \Lambda_k(U_{km}) - \Lambda_{k0}(U_{km}) \right\}^2 + \sum_{k=K_1+1}^{K} \left\{ \Lambda_k(Y_k) - \Lambda_{k0}(Y_k) \right\}^2 \right] 
+ \left\| \hat{\beta} - \beta_0 \right\|^2 + \left\| \hat{\gamma} - \gamma_0 \right\|^2 + \left\| \hat{\Sigma} - \Sigma_0 \right\|^2
\]
\[
= \sqrt{n} \left\{ O_P(n^{-2/3}) + O_P(1) \left\| \hat{\beta} - \beta_0 \right\|^2 + O_P(1) \left\| \hat{\gamma} - \gamma_0 \right\|^2 + O_P(1) \left\| \hat{\Sigma} - \Sigma_0 \right\|^2 \right\}
\]
\[
= O_P \left( n^{-1/6} + \sqrt{n} \left\| \hat{\beta} - \beta_0 \right\|^2 + \sqrt{n} \left\| \hat{\gamma} - \gamma_0 \right\|^2 + \sqrt{n} \left\| \hat{\Sigma} - \Sigma_0 \right\|^2 \right). \]

Thus,
\[ G_n \left\{ l_\theta(\hat{\theta}, \hat{A}) \right\} = -\sqrt{n} \mathcal{P} \left\{ l_{\theta\theta}(\hat{\theta} - \theta_0) + l_{\theta A}(\hat{A} - A_0) \right\} 
+ O_P \left( n^{-1/6} + \sqrt{n} \left\| \hat{\beta} - \beta_0 \right\|^2 + \sqrt{n} \left\| \hat{\gamma} - \gamma_0 \right\|^2 + \sqrt{n} \left\| \hat{\Sigma} - \Sigma_0 \right\|^2 \right), \]
and
\[ G_n \left\{ l_A(\hat{\theta}, \hat{A})(h) \right\} = -\sqrt{n} \mathcal{P} \left\{ l_{A\theta}(h)(\hat{\theta} - \theta_0) + l_{A A}(h, \hat{A} - A_0) \right\} 
+ O_P \left( n^{-1/6} + \sqrt{n} \left\| \hat{\beta} - \beta_0 \right\|^2 + \sqrt{n} \left\| \hat{\gamma} - \gamma_0 \right\|^2 + \sqrt{n} \left\| \hat{\Sigma} - \Sigma_0 \right\|^2 \right), \]

where \( l_{\theta\theta} \) is the second derivative of \( l(\theta, A) \) with respect to \( \theta \), \( l_{\theta A}(h) \) is the derivative of \( l_\theta \) along the submodel \( dA_{\epsilon,h} \), \( l_{A\theta}(h) \) is the derivative of \( l_A(h) \) with respect to \( \theta \), and \( l_{A A}(h, \hat{A} - A_0) \) is the derivative of \( l_A(h) \) along the submodel \( dA_0 + cd(\hat{A} - A_0) \). All derivatives are evaluated at \((\theta_0, A_0)\).
If the least favorable direction exists, we denote it as \( h^* = (h_1^*, \ldots, h_K^*) \), where \( h_k^* \) (\( k = 1, \ldots, K_1 \)) is \((p + K_2 + 2)\)-dimensional vector of functions in \( L_2(\mu_k) \), and \( h_k^* \) (\( k = K_1 + 1, \ldots, K \)) is \((p + K_2 + 2)\)-dimensional vector of functions in \( L_2(U_k) \). We first show the existence of \( h^* \), which is the solution to \( l_A^* l_A(h^*) = l_A^* \) with \( l_A^* \) as the adjoint operator of \( l_A \). Let \( Q = \prod_{k=1}^{K_1} L_2(\mu_k) \times \prod_{k=K_1+1}^K L_2(U_k) \). We equip \( Q \) with an inner product defined as

\[
<h^{(1)}, h^{(2)}> = \sum_{k=1}^{K_1} \int_{U_k} h_k^{(1)}(t) h_k^{(2)}(t) d\mu_k(t) + \sum_{k=K_1+1}^K \int_0^{t_k} h_k^{(1)}(t) h_k^{(2)}(t) d\Lambda_{k0}(t),
\]

where \( h^{(1)} = (h_1^{(1)}, \ldots, h_K^{(1)}) \) and \( h^{(2)} = (h_1^{(2)}, \ldots, h_K^{(2)}) \). On the same space, we define

\[
\|h\| = \mathbb{P}\{l_A(\theta, A_0)(h)^2\}^{1/2}
\]

for \( h = (h_1, \ldots, h_K) \). It is easy to show that \( \cdot \) is a seminorm on \( Q \). Furthermore, if \( \|h\| = 0 \), then \( \mathbb{P}\{l_A(\theta, A_0)(h)^2\} = 0 \). Thus, with probability 1, \( l_A(\theta, A_0)(h) = 0 \). By the arguments in the proof of Lemma 5, \( h_k(t_k) = 0 \) for \( t_k \in U_k \) for \( k = 1, \ldots, K \). Clearly, \( \|h\| \leq c \leq h, h > 1/2 \) for some constant \( c \) by the Cauchy-Schwarz inequality. According to the bounded inverse theorem in Banach spaces, we have \( <h, h>^{1/2} \leq \tilde{c}\|h\| \) for another constant \( \tilde{c} \). By the Lax-Milgram theorem (Zeidler 1995), \( h^* \) exists and for any \( t_k \in U_k \),

\[
\int_0^{t_k} \mathbb{P}\left\{ \sum_{m=0}^{M_k} \delta_{km} \int_b H_1k(t_k, U_{km}, U_{km+1}, b, \mathcal{O}; \theta, A_0) db \int_b H_1k(s, U_{km}, U_{km+1}, b, \mathcal{O}; \theta, A_0) db \right\} \times h_k^*(s) d\Lambda_{k0}(s)
\]

for \( k = 1, \ldots, K_1 \) and

\[
\int_0^{t_k} \mathbb{P}\left[ I(t_k \leq C_k) \exp\{-\Lambda_k(t_k)\} + \int_b H_2k(t_k, b, \mathcal{O}; \theta, A_0) db \right] \int_b H_2k(s, b, \mathcal{O}; \theta, A_0) db
\]

31
Let \( A \) be the estimator of \( A \) in the bootstrap sample. We denote \( \hat{P}_n \) as the bootstrap empirical distribution and \( \hat{G}_n = \sqrt{n}(\hat{P}_n - P_n) \) as the bootstrap empirical process. Using
arguments in the proof of Theorem 2, we can show that

\[
\hat{G}_n \{ l_\theta (\hat{\theta}, \hat{A}) \} = \sqrt{n} \hat{P}_n \left\{ l_\theta (\hat{\theta}, \hat{A}) - l_\theta (\hat{\theta}^*, \hat{A}^*) \right\} \\
= \sqrt{n} \hat{P}_n \left\{ l_{\theta\theta} (\hat{\theta} - \hat{\theta}^*) + l_{\theta A} (\hat{A} - \hat{A}^*) \right\} \\
+ O_P \left( n^{-1/6} + \sqrt{n} \| \hat{\beta} - \beta_0 \|^2 + \sqrt{n} \| \hat{\gamma} - \gamma_0 \|^2 + \sqrt{n} \| \hat{\Sigma} - \Sigma_0 \|^2 \right)
\]

and

\[
\hat{G}_n \{ l_A (\hat{\theta}, \hat{A}) (h) \} = \sqrt{n} \hat{P}_n \left\{ l_A (\hat{\theta}, \hat{A}) (h) - l_A (\hat{\theta}^*, \hat{A}^*) (h) \right\} \\
= \sqrt{n} \hat{P}_n \left\{ l_{A\theta} (h) (\hat{\theta} - \hat{\theta}^*) + l_{AA} (h, \hat{A} - \hat{A}^*) \right\} \\
+ O_P \left( n^{-1/6} + \sqrt{n} \| \hat{\beta} - \beta_0 \|^2 + \sqrt{n} \| \hat{\gamma} - \gamma_0 \|^2 + \sqrt{n} \| \hat{\Sigma} - \Sigma_0 \|^2 \right).
\]

Therefore,

\[
\hat{G}_n \{ l_\theta (\hat{\theta}, \hat{A}) \} - \hat{G}_n \{ l_A (\hat{\theta}, \hat{A}) (h^*) \} \\
= \sqrt{n} \hat{P}_n \left\{ l_{\theta\theta} (\hat{\theta} - \hat{\theta}^*) + l_{\theta\theta} (\hat{A} - \hat{A}^*) \right\} - \sqrt{n} \hat{P}_n \left\{ l_{\theta\theta} (h^*) (\hat{\theta} - \hat{\theta}^*) + l_{A\theta} (h^*, \hat{A} - \hat{A}^*) \right\} \\
+ O_P \left( n^{-1/6} + \sqrt{n} \| \hat{\beta} - \beta_0 \|^2 + \sqrt{n} \| \hat{\gamma} - \gamma_0 \|^2 + \sqrt{n} \| \hat{\Sigma} - \Sigma_0 \|^2 \right) \\
= \sqrt{n} \hat{P} \left[ \left\{ l_\theta (\hat{\theta}, \hat{A}) - l_A (\hat{\theta}, \hat{A}) (h^*) \right\} \otimes^2 \right] (\hat{\theta} - \hat{\theta}^*) \\
+ O_P \left( n^{-1/6} + \sqrt{n} \| \hat{\beta} - \beta_0 \|^2 + \sqrt{n} \| \hat{\gamma} - \gamma_0 \|^2 + \sqrt{n} \| \hat{\Sigma} - \Sigma_0 \|^2 \right).
\]

By the arguments in the proof of Theorem 2,

\[
\sqrt{n} (\hat{\theta} - \hat{\theta}^*) = \left( \mathbb{P} \left\{ l_\theta (\theta_0, A_0) - l_A (\theta_0, A_0) (h^*) \right\} \otimes^2 \right)^{-1} \hat{G}_n \left\{ l_\theta (\hat{\theta}, \hat{A}) - l_A (\hat{\theta}, \hat{A}) (h^*) \right\} + o_P(1) \\
= \left( \mathbb{P} \left\{ l_\theta (\theta_0, A_0) - l_A (\theta_0, A_0) (h^*) \right\} \otimes^2 \right)^{-1} \hat{G}_n \left\{ l_\theta (\hat{\theta}, \hat{A}) - l_A (\hat{\theta}, \hat{A}) (h^*) \right\} + o_P(1),
\]

where the last equality follows from Theorem 3.6.1 of van der Vaart and Wellner (1996). Therefore, \( \sqrt{n} (\hat{\theta} - \hat{\theta}^*) \) converges weakly to a zero-mean normal random vector, and \( \sqrt{n} (\hat{\theta} - \hat{\theta}^*) \) and \( \sqrt{n} (\hat{\theta} - \theta_0) \) have the same asymptotic distribution.
REFERENCES


