

Multistate analysis of multitype recurrent event and failure time data with event feedbacks in biomarkers

Chuoxin Ma | Jianxin Pan 

Department of Mathematics,
The University of Manchester,
Manchester, UK

Correspondence

Jianxin Pan, Department of Mathematics,
The University of Manchester, Oxford
Road, Manchester M13 9PL, UK.
Email: jianxin.pan@manchester.ac.uk

Abstract

In this paper we propose a class of multistate models for the analysis of multitype recurrent event and failure time data when there are past event feedbacks in longitudinal biomarkers. It can well incorporate various effects, including time-dependent and time-independent effects, of different event paths or the number of occurrences of events of different types. Asymptotic unbiased estimating equations based on polynomial splines approximation are developed. The consistency and asymptotic normality of the proposed estimators are provided. Simulation studies show that the naive estimators which either ignore the past event feedback or the measurement errors are biased. Our method has a better coverage probability of the time-varying/constant coefficients, compared to the naive methods. An application to the dataset from the Atherosclerosis Risk in Communities Study, which is also the motivating example to develop the method, is presented.

KEYWORDS

cardiovascular disease, measurement errors, multistate process, past event feedback, semiparametric coefficients

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

1 | INTRODUCTION

Characterizing the association between multiple ordered events and time-varying covariates, under a feedback mechanism in the event process, is of great interest in many areas. Steele (2011) studied women employment histories, where the time-varying covariate, presence of children, may affect employment transitions; and employment status may in turn affect the timing of childbearing. Another example occurs in cardiovascular disease (CVD) studies. Some clinical evidences (Amarenco & Labreuche, 2009; Elisaf et al., 1999; Khan et al., 2017; Woo et al., 1990) suggest that the biomarker data may carry information from the past CVD events, and may have further impact on the occurrences of the future ones.

In our motivating example, the Atherosclerosis Risk in Communities (ARIC) Study, researchers sought to evaluate the relationship between the levels of and/or changes in risk factors with CVD events in men and women from four U.S. communities (The ARIC Investigators, 1989). Various types of events were recorded and more than one type of event was observed from the same participant, including myocardial infarction (MI), stroke, and cardiovascular death. The existing analysis of the ARIC data, including but not limited to Wattanakit et al. (2005), Yatsuya et al. (2011), Kim et al. (2012), and Li and Lu (2015), focused on survival event or a single type of recurrent event but did not address the issue of event feedback mechanism. To properly reflect the overall cardiovascular burden of the risk factors, the multitype CVD events should be modeled simultaneously, because such events are very likely associated. Moreover, lack of consideration of the feedback mechanism may lead to biased estimate of the association between event times and biomarkers. For example, the profile of systolic blood pressure (SBP) can be associated with prior MI history and the levels of high-density lipoprotein cholesterol (HDL) may change after onset of stroke. The statistical estimation procedure is challenged by the fact that the biomarkers are only measured at 3-year intervals in the ARIC Study and are very likely subject to measurement errors as well.

Various approaches have been proposed for the analysis of time to multitype events. Chen et al. (2012) developed an additive marginal rate model for the analysis of multitype recurrent event data. Such model characterizes the rates of the event process without specifying the dependence structure among event occurrences. Frailty models are also popular choices to account for the association among different types of time-to-event process (Brown & Ezekowitz, 2019; Chen et al., 2005; Yu et al., 2018). The aforementioned methods, however, either fail to capture the competing risks structure of the multitype events or is not able to characterize disease progression, which is reflected by the impact of the previous event history on new event occurrences. For example for some chronic disease, a larger number of hospital admissions may be associated with a decrease in the times between hospital admissions in the future (Ieva et al., 2017). To have a natural formulation of all possible combinations of multitype events and reflect the progressive feature of a disease process, multistate models have also been employed for the analysis of recurrent event and terminal event data. Nathoo and Dean (2008) proposed a hierarchical multistate modeling framework for the analysis of longitudinal event data where the dependence between different transition rates can be well addressed. Their model and estimation method were applied to investigate the readmission rate and mortality for acute coronary syndrome. Ieva et al. (2017) proposed a flexible multistate representation of the event process. Repeated hospitalizations and death in heart failure patients were investigated. These multistate models are effective in exploring the dependence structure among event occurrences. But they only allow

for a single type of recurrent event. Therefore new statistical models need to be developed for the joint analysis of multitype recurrent events and terminal events, which is very common in practice.

When the past event history enters the model as a single covariate itself, it is defined as a dynamic covariate indicating a feedback mechanism in the event process (Gjessing et al., 2010). There have been many works dealing with the dynamic covariate in time-to-event analysis. Aalen et al. (2004) presented an approach for the analysis of multivariate counting process, in which the dynamic covariate comes into the intensity model in an additive manner. Several forms of the dynamic covariates were discussed, such as the number of previous events, time since last event and average number of prior events. Peña (2006) described a general class of models for recurrent events, where the number of past events was used as a covariate in the intensity function. Fosen et al. (2006) developed a method for the analysis of recurrent event data using information on previous occurrences of the event as a time-dependent covariate. Direct, indirect, and total effects of other baseline covariates were investigated based on the additive hazard model. Yiu et al. (2018) considered a clustered multistate model with dynamic covariates and analyzed the transition intensities and sojourn times in a study of psoriatic arthritis. These methods, however, are not adequate for the event history analysis when the past event feedbacks are associated with the longitudinal covariates. As mentioned in the motivating example, the trajectories of biomarkers are influenced by prior event history. Hence the past event impact is not simply served as a single covariate for the event intensity function but is actually associated with other time-dependent covariates in a complex way. Dai and Pan (2018) considered the joint models for survival and longitudinal data where the longitudinal process is related to the history of some random observation time points. Their models were developed under the framework of informative observation times for longitudinal data and hence is not appropriate for our analysis where primary interest is in the time-to-event modelling.

In this paper, we propose a new class of multistate models with past event feedbacks in the associated biomarkers. To our knowledge, this is the first attempt at modeling past event feedback in the longitudinal covariates in event history analysis. There is also limited literature that considered a multistate representation of the multitype recurrent event and terminal event data. Our proposed model and estimation procedure are novel and feature the following aspects. First, the competing risks feature of events of different types and the correlation structure of the consecutive events are well addressed within the multistate model framework. The functional forms of the event intensities are allowed to depend on the event order or event type or a combination of the both, reflecting disease progression in a flexible way. Second, the trajectories of the associated biomarkers are characterized by random effects models. Function of event paths enters the random effects model as a predictor variable to capture the past event feedback. The distributions of the random effects are left unspecified making our model robust against distribution misspecification. Asymptotically unbiased estimating equations are obtained based on an extension of the corrected score approach. Third, both time-dependent coefficients and time-independent regression parameters are incorporated in the multistate models, considering the fact that our motivating example is a long term study. The association between event time and biomarker may change along time and therefore the proportional assumption may be violated. The rest of the paper is structured as follows. Section 2 introduces the model and notation. Section 3 presents the estimation procedure and the large sample results. Simulation studies and real application are presented in Sections 4 and 5, respectively.

2 | MODEL AND NOTATION

Suppose that there are n individuals and the data from different individuals are independent. Each individual can possibly experience K types of recurrent events. Let M denote the maximum number of nonterminal events observed among all individuals. We represent the event process in terms of a multistate process. An individual moves between the intermediate states when the recurrent events occur and moves to the absorbing state when a terminal event occurs. A simple example showing state transition paths when $K = 2$ and $M = 4$ is shown in Figure 1. We use white boxes to represent intermediate states and use grey boxes for absorbing states. The intermediate states in the first row are defined by the type 1 recurrent events and those in the second row are defined by the type 2 recurrent events. The m th column of the intermediate states represents that this is the m th nonterminal event a subject has experienced. For example, consider stroke as a type 1 recurrent event, MI as a type 2 recurrent event and cardiovascular death as a terminal event. Transition to state 2 corresponds to the fact that an individual has experienced a second recurrent event and this event is a stroke. Transition to state 7 represents the case that a third recurrent event has occurred and it is a MI. The absorbing state 9 represents direct cardiovascular death and the absorbing state 10 represents death with previous stroke and/or MI history. If an individual experienced the following series of events: MI \rightarrow stroke \rightarrow MI \rightarrow cardiovascular death, then the multi-state transition path is $5 \rightarrow 2 \rightarrow 7 \rightarrow 10$ according to Figure 1. Under a general setup, we assume that the multistate process takes values in a finite state space $\{0, 1, \dots, KM, KM + 1, KM + 2\}$ and denote the counting process that tracks state transitions for subject i as $N_{hg}^i(t)$, $i = 1, \dots, n$. The possible pair of values of h and g can be defined according to the following two cases. For a terminal event, we have $h \in \{0, 1, \dots, KM\}$ and $g \in \{KM + 1, KM + 2\}$. For a recurrent event, we have $h \in \{0, 1, \dots, M - 1, M + 1, \dots, 2M - 1, \dots, (K - 1)M + 1, \dots, KM - 1\}$ and $g \in \{r_h + 1, r_h + M + 1, \dots, r_h + (K - 1)M + 1\}$, where r_h is the remainder of the division of $h + M$ by M .

Let $Y_h^i(t)$ be the at risk process which equals 1 if subject i is at risk for the transition out of state h at t . We assume that the multistate process is Markov and the intensity process can be written as

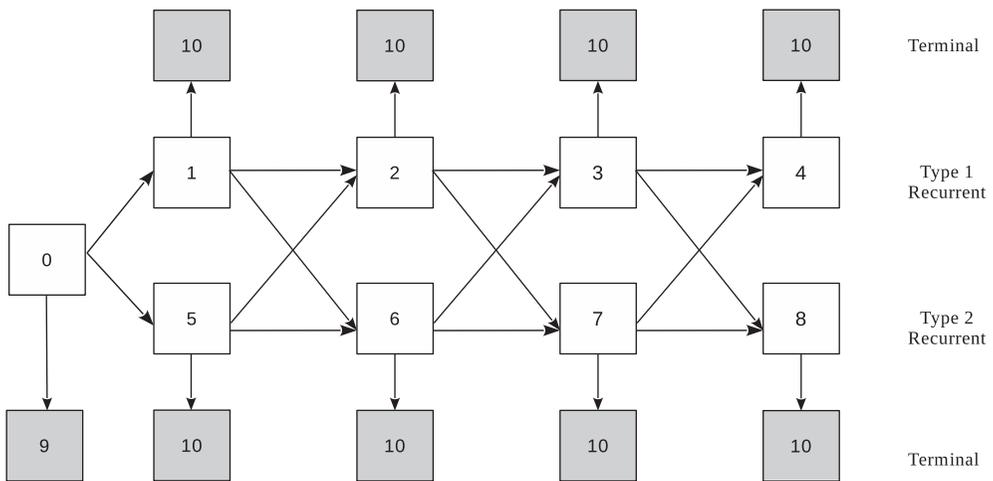


FIGURE 1 A simple example of state transition paths with two types of recurrent events ($K = 2$) and maximum four occurrences of nonterminal events ($M = 4$)

$$\lim_{\Delta t \downarrow 0} \frac{\text{pr} \left\{ N_{hg}^i(t + \Delta t) = 1 \mid Y_h^i(t) = 1, \mathcal{H}_i(t-) \right\}}{\Delta t} = Y_h^i(t) \lambda_{hg}^i(t). \tag{1}$$

To characterize the associations between the risk factors and state transition times, we consider the following multiplicative intensity model

$$\lambda_{hg}^i(t) = \lambda_{hg}^0(t) \exp \left(\boldsymbol{\eta}_{hg}^T(t) \mathbf{X}_i(t) + \boldsymbol{\gamma}_{hg}^T \mathbf{W}_i(t) \right), \tag{2}$$

where $\mathbf{W}_i(t)$ are p_w -dimensional covariates with constant effects on the event intensities and $\boldsymbol{\gamma}_{hg}$ are the time-independent coefficients capturing such effects. $\mathbf{X}_i(t)$ are p_x -dimensional covariates with effects that change over time and $\boldsymbol{\eta}_{hg}(t)$ are the time-dependent coefficients capturing such effects. $\lambda_{hg}^0(t)$ are baseline intensity functions. In practical situations some of the states may have very sparse data for estimation. One can consider a more parsimonious model assuming that several transitions share the same intensity model parameters in this case. For example in our real analysis, among those participants who had at least one recurrent event, the average number of type 1 recurrent events observed per subject is around 1.26 and the average number of type 2 recurrent events is about 1.19. It is expected that some intermediate states have sparse event data. Therefore we assume that all the type 1 recurrent event times share the same intensity model parameters and all the type 2 recurrent event times share the same parameters. Another way to tackle the sparse data problem is to choose a smaller M and treat the data from the subjects observed to have more than M intermediate events to be censored (Cook et al., 2004). Different applications may use different values of M and it depends on the model complexity and the background of the data. For example in Ieva et al. (2017) the repeated hospitalization and death in patients with heart failure were investigated using multistate modeling approach. In their research, a maximum of six hospital admissions are modeled, and subsequent admissions (but not deaths) are ignored, due to the sparsity of data from individuals with more than six admissions. We suggest trying different values of M and compute the AIC-type criterion using the formulas in section 3.1 in the Appendix S1. The number M that minimizes the AIC-type criterion can indicate a best fit of the data.

Estimating the regression coefficients is challenging by the fact that covariates $\mathbf{X}_i(t)$ and $\mathbf{W}_i(t)$ are not observed at each event time point and are possibly subject to measurement errors. The presence of past event feedback makes the problem even more complicated because we need to properly evaluate the event history information in the longitudinal observations. For each r th component in $\mathbf{X}_i(t)$, denote the longitudinal observations as X_{ijr} , $j = 1, \dots, m_{X_{ir}}$, $r = 1, \dots, p_x$. Similarly for each s th component in $\mathbf{W}_i(t)$ denote the observations as W_{ijs} , $j = 1, \dots, m_{W_{is}}$, $s = 1, \dots, p_w$. Let $\mathcal{H}_i(t-)$ represent the history of the multistate process of subject i before time t . Assume that the longitudinal observations of the covariates are generated from the following models with additive measurement errors.

$$\begin{aligned} X_{ijr} &= \boldsymbol{\beta}_{X_{ir}}^T f_{Xr} \left(t_{ij}, \mathcal{H}_i(t_{ij}-) \right) + \varepsilon_{X_{ijr}}, \\ W_{ijs} &= \boldsymbol{\beta}_{W_{is}}^T f_{Ws} \left(t_{ij}, \mathcal{H}_i(t_{ij}-) \right) + \varepsilon_{W_{ijs}}, \end{aligned} \tag{3}$$

where $f_{Xr}(\cdot)$ and $f_{Ws}(\cdot)$ are known functions of time and past event history. They should satisfy the condition that the intensity process of $N_{hg}^i(t)$ will not go to infinity over time (Gjessing et al., 2010). $\boldsymbol{\beta}_{X_{ir}}$ and $\boldsymbol{\beta}_{W_{is}}$ are vectors of random effects, of dimensions c_{Xr} and c_{Ws} , respectively. Note that we do not make any assumptions about the distributions of these random

effects. The measurement errors $\varepsilon_{X_{ijr}}$ and $\varepsilon_{W_{ijs}}$ are mutually independent with $\varepsilon_{X_{ijr}} \stackrel{iid}{\sim} N(0, \sigma_{X_r}^2)$ and $\varepsilon_{W_{ijs}} \stackrel{iid}{\sim} N(0, \sigma_{W_s}^2)$. The normal distribution assumption is a standard practice in measurement error theories. To assess this normal assumption, one can apply the common method such as QQ plot to check the residuals of the fitted models of the longitudinal data. For simplicity, the measurement errors of any two components of the covariates are assumed to be independent. We also assume that the observation time points t_{ij} are noninformative. The measurement errors are independent of the true covariates, the event times and the censoring time.

Model (3) extends the existing joint models for time-to-event data and longitudinal data, in the sense that it allows the covariates to depend on past event history. The linear combinations $\beta_{X_{ir}f_{X_r}}^T(t_{ij}, \mathcal{H}_i(t_{ij}-))$ and $\beta_{W_{is}f_{W_s}}^T(t_{ij}, \mathcal{H}_i(t_{ij}-))$ provide flexible modeling of the true covariate process, allowing various forms of time trajectories and past event feedbacks. Let $I(\cdot)$ represents an indicator function. The longitudinal model $\beta_{X_{ir}f_{X_r}}^T(t_{ij}, \mathcal{H}_i(t_{ij}-)) = \beta_{X_{ir0}} + \beta_{X_{ir1}}t_{ij} + \beta_{X_{ir2}} \sum_{h,g} \left(N_{hg}^i(t-)I(g \in \{1, \dots, M\}) \right)$ represents a linear trajectory of observation times and the type 1 recurrent event history. If the model of the true covariate process is chosen as $\beta_{W_{is}f_{W_s}}^T(t_{ij}, \mathcal{H}_i(t_{ij}-)) = \beta_{W_{is1}}\sqrt{t_{ij}} + \beta_{W_{is2}} \left(\sum_{h,g} N_{hg}^i(t-)I(g = h + 1) \right)$, it indicates that the levels of biomarker have a nonlinear relationship with the observation times and they are influenced by the successive occurrences of events of the same type. In practice one has to choose the functional forms of the longitudinal trajectory $f_{X_r}(\cdot)$ and $f_{W_s}(\cdot)$, and it mainly depends on the background and knowledge of the application field (Fisher & Lin, 1999). For example in our real data analysis, blood pressure and cholesterol levels are modeled as a linear trajectory of measurement times plus an indicator function of prior CVD history. The linear trajectory is chosen according to the previous medical research of the ARIC data (Barrett et al., 2019; Paige et al., 2017) and the indicator function is chosen with relevant medical theories (Amarenco et al., 2020; Khan et al., 2017, 2018). If there is no background knowledge of the functional forms, one may try several different functions and compute the AIC-type criterion using the formulas in section 3.1 in Appendix S1. The function with a smaller AIC-type value can indicate a better fit of the data. Model (3) is also applicable to several types of covariates. For example, a precisely measured time-dependent covariate $X_{ir}(t)$ can be accommodated by setting $\varepsilon_{X_{ijr}} = 0$, and a baseline covariate W_{is} without measurement errors can be handled by setting $m_{W_{is}} = 1$ and $\varepsilon_{W_{ijs}} = 0$.

3 | ESTIMATION PROCEDURE

When the true values of the biomarkers are known at all event time points, we can obtain an accurate estimator of the coefficients $\eta_{hg}(t)$ and γ_{hg} in Model (2) based on the local partial likelihood (Cai & Sun, 2003) or the spline-based partial likelihood (Nan et al., 2005). Even if the biomarkers are intermittently measured and possibly with errors, analysis focusing on a single event is still feasible using the conditional score approach (Tsiatis & Davidian, 2001) or the corrected score approach (Huang et al., 2016; Song & Wang, 2008, 2017). However in the case of multiple events with feedback mechanism, the existing methods cannot be easily adapted to obtain a valid estimate. In what follows, we develop an inference procedure based on an extension to the corrected score approach. The idea is that we first estimate the values of the time-varying covariates at each event time point using the least squares method. We also approximate the time-dependent coefficients with polynomial splines. Then we replace the corresponding

quantities in the score function of the log-partial likelihood with the least square estimates and the spline approximations, and derive the resulting bias due to the substitution. The asymptotically unbiased estimating equations are constructed.

3.1 | The estimating equations

To calculate the least square estimates of the covariate value at time t , we use only the longitudinal data up to and including time t . This is to ensure the measurable property of the covariate estimates, which is essential for the derivation of the theoretical results. Let $m_{X_{ir}}(t)$ be the number of longitudinal observations of the r th component of $\mathbf{X}_i(u)$ before time t and similarly define $m_{W_{is}}(t)$ for $\mathbf{W}_i(u)$. The estimated values of the covariates at time t are denoted as $\hat{X}_{ir}(t)$ and $\hat{W}_{is}(t)$. And the conditional variances of $\hat{X}_{ir}(t)$ and $\hat{W}_{is}(t)$ given the measurement time points and the variances of the measurement errors $\sigma_{X_r}^2$ and $\sigma_{W_s}^2$, are denoted as $\hat{\sigma}_{X_{ir}}^2(t)$ and $\hat{\sigma}_{W_{is}}^2(t)$. The detailed derivation is placed in section 1 of Appendix S1. The matrix notations can be easily written as $\hat{\mathbf{X}}_i(t) = (\hat{X}_{i1}(t), \dots, \hat{X}_{ip_x}(t))^T$ and $\hat{\mathbf{W}}_i(t) = (\hat{W}_{i1}(t), \dots, \hat{W}_{ip_w}(t))^T$. If we simply replace the unknown values of the true covariates with their least squares estimates, the “naive” score function of the log-partial likelihood is

$$1/n \sum_{i=1}^n \int_0^L Y_h^i(u) \left\{ \left(\hat{\mathbf{X}}_i^T(u), \hat{\mathbf{W}}_i^T(u) \right) - \frac{\sum_{j=1}^n Y_h^j(u) \left(\hat{\mathbf{X}}_j^T(u), \hat{\mathbf{W}}_j^T(u) \right) \exp \left(\boldsymbol{\eta}_{hg}^T(u) \hat{\mathbf{X}}_j(u) + \boldsymbol{\gamma}_{hg}^T \hat{\mathbf{W}}_j(u) \right)}{\sum_{j=1}^n Y_h^j(u) \exp \left(\boldsymbol{\eta}_{hg}^T(u) \hat{\mathbf{X}}_j(u) + \boldsymbol{\gamma}_{hg}^T \hat{\mathbf{W}}_j(u) \right)} \right\} dN_{hg}^i(u), \tag{4}$$

where the observation period is $[0, L]$. Let $\mathcal{F}_i(t-)$ be the filtration containing all the history information of the multi-state process and the covariate process before time t for individual i . Let $(A_i)_{i=1}^n$ denote a series of values from A_1 to A_n , and denote $\boldsymbol{\Sigma}_X^i(t) = \text{diag} \left\{ \left(\hat{\sigma}_{X_{ir}}^2(t) \right)_{r=1}^{p_x} \right\}$, $\boldsymbol{\Sigma}_W^i(t) = \text{diag} \left\{ \left(\hat{\sigma}_{W_{is}}^2(t) \right)_{s=1}^{p_w} \right\}$. Using the following property of a normal random variable X with mean μ and variance matrix Σ : $\mathbb{E} \{ (X - \Sigma \lambda) \exp(\lambda^T X - 1/2 \lambda^T \Sigma \lambda) \} = \mu \exp(\lambda^T \mu)$, it can be shown that the conditional expectation of the numerator in the score function is

$$\begin{aligned} & \mathbb{E} \left\{ Y_h^j(u) \left(\hat{\mathbf{X}}_j^T(u), \hat{\mathbf{W}}_j^T(u) \right)^T \exp \left(\boldsymbol{\eta}_{hg}^T(u) \hat{\mathbf{X}}_j(u) + \boldsymbol{\gamma}_{hg}^T \hat{\mathbf{W}}_j(u) \right) \middle| \mathcal{F}_j(t-) \right\} \\ &= Y_h^j(u) \left(\mathbf{X}_j^T(u) + \boldsymbol{\Sigma}_X^j(u) \boldsymbol{\eta}_{hg}(u), \mathbf{W}_j^T(u) + \boldsymbol{\Sigma}_W^j(u) \boldsymbol{\gamma}_{hg} \right)^T \exp \left(\boldsymbol{\eta}_{hg}^T(u) \mathbf{X}_j(u) + \boldsymbol{\gamma}_{hg}^T \mathbf{W}_j(u) \right) \\ & \quad + 1/2 \boldsymbol{\eta}_{hg}^T(u) \boldsymbol{\Sigma}_X^j(u) \boldsymbol{\eta}_{hg}(u) + 1/2 \boldsymbol{\gamma}_{hg}^T \boldsymbol{\Sigma}_W^j(u) \boldsymbol{\gamma}_{hg}. \end{aligned} \tag{5}$$

Therefore the biases are resulted from the terms involving $\boldsymbol{\Sigma}_X^j(u)$ and $\boldsymbol{\Sigma}_W^j(u)$.

To show the spline-based approximation of the time-dependent coefficients, we need the following notations. For the r th component in $\boldsymbol{\eta}_{hg}(t)$, let $n_{hg,r}$ be the number of knots of the spline function that approximates it and let d be the degree of the spline function. Let $\boldsymbol{\alpha}_{hg,r} = (\alpha_{hg,jr})_{j=1}^{n_{hg,r}+d}$ be the spline coefficients, and $\mathbf{B}_{hg,r}(t) = (B_{hg,jr}(t))_{j=1}^{n_{hg,r}+d}$ be the basis

functions. Then we have $\eta_{hg,r}(t) \approx \alpha_{hg,r}^T \mathbf{B}_{hg,r}(t)$. Denote $\alpha_{hg} = (\alpha_{hg,r})_{r=1}^{p_x}$, and the vector containing the spline coefficients and constant parameters as $\theta_{hg} = (\alpha_{hg}^T, \gamma_{hg}^T)^T$. The dimension of θ_{hg} is $D_{hg} = \sum_{r=1}^{p_x} (n_{hg,r} + d) + p_w$. Define $\text{blkdiag} \{ (A_i^1)_{i=1}^n, (A_i^2)_{i=1}^n \}$ as a block diagonal matrix with the main diagonal blocks $(A_i^1)_{i=1}^n$ and $(A_i^2)_{i=1}^n$. Let $\mathbf{B}_{hg}(t) = \text{blkdiag} \{ \mathbf{B}_{hg,1}(t), \dots, \mathbf{B}_{hg,p_x}(t) \}$ be a $p_x \times (\sum_{r=1}^{p_x} (n_{hg,r} + d))$ -dimensional matrix of the stacked basis functions. Let $\hat{\mathbf{Z}}_{hg}^i(t) = (\hat{\mathbf{X}}_i^T(t) \mathbf{B}_{hg}(t), \hat{\mathbf{W}}_i^T(t))^T$ be a D_{hg} -dimensional vector of the estimated covariates and let $\Sigma_{Z_{hg}}^i(t) = \text{blkdiag} \{ \mathbf{B}_{hg}^T(t) \Sigma_X^i(t) \mathbf{B}_{hg}(t) \Sigma_W^i(t) \}$ be a $D_{hg} \times D_{hg}$ -dimensional matrix of the conditional variances of the estimated covariates.

Denote C_i as the censoring time of subject i and we assume that it is noninformative. Let $\bar{Y}_h^i(t) = I(m_{Xir}(t) \geq c_{Xr}, m_{Wis}(t) \geq c_{Ws}, C_i \geq t, Y_h^i(t-) = 1)$ be the overall at risk process. Denote $A^{\otimes 0} = 1, A^{\otimes 1} = A$ and $A^{\otimes 2} = AA^T$ for a general vector A . The following functions are defined according to the results in (5), which aim to remove the bias due to the least square estimates of the covariates. For $l = 0, 1, 2$,

$$S_{hg}^{(l)}(\theta_{hg}, u) = 1/n \sum_{j=1}^n \bar{Y}_h^j(u) \left(\hat{\mathbf{Z}}_{hg}^j(u) - \Sigma_{Z_{hg}}^j(u) \theta_{hg} \right)^{\otimes l} \times \exp \left(\theta_{hg}^T \hat{\mathbf{Z}}_{hg}^j(u) - 1/2 \theta_{hg}^T \Sigma_{Z_{hg}}^j(u) \theta_{hg} \right).$$

Combining (4) and (5), it can be easily shown that the conditional expectations of the numerator and denominator in $S_{hg}^{(1)}(\theta_{hg}, u) / S_{hg}^{(0)}(\theta_{hg}, u)$ in the following estimating function equal to the corresponding quantities in the score function of the log-partial likelihood with true values of the covariates. Therefore as the sample size increases, the fraction approaches the corresponding fraction in (4) with true covariates. The asymptotically unbiased estimating equation of θ_{hg} is

$$U_{hg}(\theta_{hg}) = 1/n \sum_{i=1}^n \int_0^L \bar{Y}_h^i(u) \left\{ \hat{\mathbf{Z}}_{hg}^i(u) - \frac{S_{hg}^{(1)}(\theta_{hg}, u)}{S_{hg}^{(0)}(\theta_{hg}, u)} \right\} dN_{hg}^i(u). \tag{6}$$

Let $S_{\Sigma_{Z_{hg}}}(\theta_{hg}, u) = 1/n \sum_{j=1}^n \bar{Y}_h^j(u) \Sigma_{Z_{hg}}^j(u) \exp \left(\theta_{hg}^T \hat{\mathbf{Z}}_{hg}^j(u) - 1/2 \theta_{hg}^T \Sigma_{Z_{hg}}^j(u) \theta_{hg} \right)$. The Hessian function can be written as

$$\partial U_{hg}(\theta_{hg}) / \partial \theta_{hg}^T = 1/n \sum_{i=1}^n \int_0^L \bar{Y}_h^i(u) \left\{ \frac{S_{hg}^{(1)}(\theta_{hg}, u)^{\otimes 2}}{S_{hg}^{(0)}(\theta_{hg}, u)^2} - \frac{\left(S_{hg}^{(2)}(\theta_{hg}, u) - S_{\Sigma_{Z_{hg}}}(\theta_{hg}, u) \right)}{S_{hg}^{(0)}(\theta_{hg}, u)} \right\} dN_{hg}^i(u). \tag{7}$$

Given (6) and (7), the estimates of the coefficients can be found using Newton's method.

3.2 | Theoretical properties

We will next present the asymptotic results of the estimators, which are used to calculate the standard errors and confidence intervals. The proofs are given in Appendix S1. We first define the norm functions that are used in the theorems. For any matrix function $\mathbf{F}(t)$, denote the supremum

norm as $\|\mathbf{F}\|_\infty = \max_{i,j} \sup_t |f_{i,j}(t)|$. Let the empirical L_2 -norm be $\|\mathbf{F}(t)\|_n^2 = 1/n \sum_{i=1}^n \mathbf{F}^T(t)\mathbf{F}(t)$ and the corresponding theoretical norm be $\|\mathbf{F}(t)\|_2^2 = \mathbb{E}\{\mathbf{F}^T(t)\mathbf{F}(t)\}$. For any vector \mathbf{a} , denote the Euclidean norm as $\|\mathbf{a}\|$. For simplicity, we assume that the number of knots $n_{hg,r}$ are the same for each component in $\boldsymbol{\eta}_{hg}(t)$ and write $n_{hg,r}$ as n_{hg} . We consider orthonormal B-spline basis functions for $(\mathbf{B}_{hg,r}(t))_{r=1}^{p_x}$. Let $\mathbf{I}_{A \times B}$ be a A by B dimensional diagonal matrix with the first B diagonal elements to be 1 and the rest to be 0. Let

$$\mathbf{B}_{hg}^*(t) = \begin{bmatrix} \mathbf{B}_{hg}(t) & \mathbf{0}_{p_x \times p_w} \\ \mathbf{0}_{p_w \times (p_x(n_{hg}+d))} & \mathbf{I}_{p_w \times p_w} \end{bmatrix}$$

be a $(p_x + p_w) \times D_{hg}$ -dimensional matrix. Denote $\boldsymbol{\eta}_{hg,T}(t)$ as the true time-varying coefficients and $\boldsymbol{\gamma}_{hg,T}$ as the true time-fixed coefficients. Denote $\boldsymbol{\phi}_{hg,T}(t) = (\boldsymbol{\eta}_{hg,T}^T(t), \boldsymbol{\gamma}_{hg,T}^T)^T$ as the true values of all the coefficients. According to the result in p. 149 in De Boor (2001), there exists $\tilde{\boldsymbol{\alpha}}_{hg}$ such that for $\tilde{\boldsymbol{\eta}}_{hg}(t) = \mathbf{B}_{hg}(t)\tilde{\boldsymbol{\alpha}}_{hg}$, we have $\|\tilde{\boldsymbol{\eta}}_{hg}(t) - \boldsymbol{\eta}_{hg,T}(t)\|_\infty = O(n_{hg}^{-c})$, where c can be found in the below regularity condition (A5). Define $\boldsymbol{\theta}_{hg,T} = (\tilde{\boldsymbol{\alpha}}_{hg}^T, \boldsymbol{\gamma}_{hg,T}^T)^T$. To facilitate derivation of the asymptotic results, we define the following notations which resemble the $\mathbf{Z}_{hg}^j(u)$, $\boldsymbol{\Sigma}_{Z_{hg}}^j(u)$ and $S_{hg}^{(l)}(\boldsymbol{\theta}_{hg}, u)$ in the proposed estimating equation (6), except that some terms involving the product of spline basis function and covariate are replaced with the covariate itself. Let $\mathbf{R}_i(t) = (\mathbf{X}_i^T(t), \mathbf{W}_i^T(t))^T$ and $\boldsymbol{\Sigma}_R^i(t) = \text{blkdiag}\{\boldsymbol{\Sigma}_X^i(t), \boldsymbol{\Sigma}_W^i(t)\}$. The following functions are notated in a way that the letter I is used when the true values of the covariates are incorporated and the letter S is used when the least squares estimates of the covariates are incorporated. For $l = 0, 1, 2$, define

$$\begin{aligned} I_{hg}^{(l)}(\boldsymbol{\phi}_{hg}, u) &= 1/n \sum_{j=1}^n \bar{Y}_h^j(u) \mathbf{R}_j^{\otimes l}(u) \exp\left(\boldsymbol{\phi}_{hg}^T(u) \mathbf{R}_j(u)\right), \\ \tilde{I}_{hg}^{(l)}(\boldsymbol{\theta}_{hg}, u) &= 1/n \sum_{j=1}^n \bar{Y}_h^j(u) \mathbf{R}_j^{\otimes l}(u) \exp\left(\boldsymbol{\theta}_{hg}^T \mathbf{B}_{hg}^{*T}(u) \mathbf{R}_j(u)\right), \\ Q_{hg}^{(l)}(\boldsymbol{\phi}_{hg}, u) &= 1/n \sum_{j=1}^n \bar{Y}_h^j(u) \left(\hat{\mathbf{R}}_j(u) - \boldsymbol{\Sigma}_R^j(u) \boldsymbol{\phi}_{hg}(u)\right)^{\otimes l} \\ &\quad \times \exp\left(\boldsymbol{\phi}_{hg}^T(u) \hat{\mathbf{R}}_j(u) - 1/2 \boldsymbol{\phi}_{hg}^T(u) \boldsymbol{\Sigma}_R^j(u) \boldsymbol{\phi}_{hg}(u)\right), \end{aligned}$$

where we simply write $\boldsymbol{\phi}_{hg}(t)$ as $\boldsymbol{\phi}_{hg}$ for the argument of $I_{hg}^{(l)}$ and $Q_{hg}^{(l)}$. Denote $s_{hg}^{(l)}(\boldsymbol{\theta}_{hg}, u)$, $i_{hg}^{(l)}(\boldsymbol{\phi}_{hg}, u)$ and $\tilde{c}_{hg}^{(l)}(\boldsymbol{\phi}_{hg}, u)$ as the limit of $S_{hg}^{(l)}(\boldsymbol{\theta}_{hg}, u)$, $I_{hg}^{(l)}(\boldsymbol{\phi}_{hg}, u)$ and $\tilde{I}_{hg}^{(l)}(\boldsymbol{\theta}_{hg}, u)$, respectively. The following notations are defined for the theoretical properties of the proposed estimating equations, which will be used to derive the asymptotic variance-covariance matrix of the estimator of the time-fixed regression coefficient.

$$\hat{U}_{hg}^*(\boldsymbol{\phi}_{hg}) = 1/n \sum_{i=1}^n \int_0^L \bar{Y}_h^i(u) \mathbf{B}_{hg}^{*T}(u) \left\{ \frac{i_{hg}^{(2)}(\boldsymbol{\phi}_{hg}, t)}{i_{hg}^{(0)}(\boldsymbol{\phi}_{hg}, t)} - \frac{i_{hg}^{(1)}(\boldsymbol{\phi}_{hg}, t)^{\otimes 2}}{i_{hg}^{(0)}(\boldsymbol{\phi}_{hg}, t)^2} \right\} \mathbf{B}_{hg}^*(u) dN_{hg}^i(u),$$

$$\begin{aligned}
U_{hg}^*(\boldsymbol{\phi}_{hg}) &= 1/n \sum_{i=1}^n \int_0^L \bar{Y}_h^i(u) \mathbf{B}_{hg}^{*T}(u) \left\{ \hat{\mathbf{R}}_i(u) - \frac{i_{hg}^{(1)}(\boldsymbol{\phi}_{hg,T}, u)}{i_{hg}^{(0)}(\boldsymbol{\phi}_{hg,T}, u)} \right\} dN_{hg}^i(u) \\
&\quad - 1/n \sum_{i=1}^n \int_0^L \bar{Y}_h^i(u) \mathbf{B}_{hg}^{*T}(u) Q_{hg}^{(1)}(\boldsymbol{\phi}_{hg,T}, u) d\Lambda_{hg}^0(u) \\
&\quad + 1/n \sum_{i=1}^n \int_0^L \bar{Y}_h^i(u) \mathbf{B}_{hg}^{*T}(u) \left\{ \frac{i_{hg}^{(1)}(\boldsymbol{\phi}_{hg,T}, u)}{i_{hg}^{(0)}(\boldsymbol{\phi}_{hg,T}, u)} \right\} \times Q_{hg}^{(0)}(\boldsymbol{\phi}_{hg,T}, u) d\Lambda_{hg}^0(u).
\end{aligned}$$

Note that $\dot{U}_{hg}^*(\boldsymbol{\phi}_{hg})$ is some quantity derived from the negative of the Hessian matrix. Denote the limit of the variance-covariance matrix as $v_{hg}(\boldsymbol{\phi}_{hg}) = \mathbb{E} \left(U_{hg}^*(\boldsymbol{\phi}_{hg}) U_{hg}^{*T}(\boldsymbol{\phi}_{hg}) \right)$ and denote the limit of the negative Hessian matrix as $\dot{u}_{hg}^*(\boldsymbol{\phi}_{hg}) = \mathbb{E} \left(\dot{U}_{hg}^*(\boldsymbol{\phi}_{hg}) \right)$. Partitioning $\dot{u}_{hg}^*(\boldsymbol{\phi}_{hg,T})$ into four blocks,

$$\dot{u}_{hg}^*(\boldsymbol{\phi}_{hg,T}) = \begin{bmatrix} \dot{u}_{hg,\alpha\alpha}^* & \dot{u}_{hg,\alpha\gamma}^* \\ \dot{u}_{hg,\gamma\alpha}^* & \dot{u}_{hg,\gamma\gamma}^* \end{bmatrix},$$

and defining

$$h_{hg}(\boldsymbol{\phi}_{hg,T}) = \left(\dot{u}_{hg,\gamma\gamma}^* - \dot{u}_{hg,\gamma\alpha}^* \left(\dot{u}_{hg,\alpha\alpha}^* \right)^{-1} \dot{u}_{hg,\alpha\gamma}^* \right)^{-1} \left[-\dot{u}_{hg,\gamma\alpha}^* \left(\dot{u}_{hg,\alpha\alpha}^* \right)^{-1} \mathbf{I}_{p_w \times p_w} \right],$$

then we can write the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\gamma}}_{hg}$ as

$$\boldsymbol{\Gamma}_{hg}(\boldsymbol{\phi}_{hg,T}) = h_{hg}(\boldsymbol{\phi}_{hg,T}) v_{hg}(\boldsymbol{\phi}_{hg,T}) h_{hg}(\boldsymbol{\phi}_{hg,T})^T. \quad (8)$$

Assume that the following regularity conditions hold for Models (2.2) and (2.3).

- (A1) $P \left(\bar{Y}_h^i(t) = 1, \forall t \in [0, L] \right) > 0$ for $h \in \{0, 1, \dots, KM\}$.
- (A2) $\lambda_{hg}^0(t)$ are continuous and $\int_0^L \lambda_{hg}^0(u) du < \infty$.
- (A3) $\mathbf{X}(t)$ and $\mathbf{W}(t)$ are left continuous with right-hand limits.
- (A4) There exists a random variable B with bounded b th moments for some $b > 2$, such that $\sup_{t \in [0, L]} E \left\{ \exp \left(2\boldsymbol{\phi}_{hg,T}(t) B \right) \right\} < \infty$, $\|\mathbf{X}(t)\|_\infty \leq B$, $\|\mathbf{W}(t)\|_\infty \leq B$, $\|\boldsymbol{\varepsilon}_X(t)\|_\infty \leq B$, $\|\boldsymbol{\varepsilon}_W(t)\|_\infty \leq B$, $\|\boldsymbol{\sigma}_X^2(t)\|_\infty \leq B$ and $\|\boldsymbol{\sigma}_W^2(t)\|_\infty \leq B$.
- (A5) Each component of the true time-varying coefficient $\boldsymbol{\eta}_{hg,T}(t)$ belongs to the space of function with bounded c th derivatives in $[0, L]$ for some $c \geq 0.5$. The true value of the constant parameter $\boldsymbol{\gamma}_{hg,T}$ belongs to a bounded open subset of \mathbb{R}^{p_w} .
- (A6) The density function of event times is bounded away from zero and infinity on $[0, L]$.
- (A7) $i_{hg}^{(2)}(\boldsymbol{\phi}_{hg}, t) / i_{hg}^{(0)}(\boldsymbol{\phi}_{hg}, t) - i_{hg}^{(1)}(\boldsymbol{\phi}_{hg}, t)^{\otimes 2} / i_{hg}^{(0)}(\boldsymbol{\phi}_{hg}, t)^2$ is positive definite with bounded eigenvalues uniformly in $t \in [0, L]$.

The above conditions will be used to derive the theoretical properties of the parameter estimators $\hat{\gamma}_{hg}$ and the nonparametric estimators $\hat{\eta}_{hg}(t)$. Conditions (A1) and (A2) are standard assumptions for proportional intensity models. Condition (A3) ensures that the counting process martingale properties are valid in the case of intensity models with time-dependent covariates. Condition (A5) ensures that the Lindeberg's condition is satisfied and also provides conditions under which the functional law of large numbers can be applied. Condition (A5) and Condition (A6) are standard regularity conditions in time-to-event data analysis with spline approximations. Condition (A6) implies that the limit of the Hessian matrix is positive definite.

Theorem 1. For Model (2), under the regularity conditions (A1)–(A6) and assume that $n_{hg} = O(n^v)$ with $0 < v < 0.5$, as $n \rightarrow \infty$ we have $\left\| \hat{\eta}_{hg}(t) - \eta_{hg,T}(t) \right\|_2 = O_p(n^{(v-1)/2} + n^{v(1/2-c)})$, $\left\| \hat{\gamma}_{hg} - \gamma_{hg,T} \right\|_\infty = O_p(n^{(v-1)/2} + n^{v(1/2-c)})$, and $\left\| \hat{\gamma}_{hg} - \gamma_{hg,T} \right\| = O_p(n^{(v-1)/2} + n^{v(1/2-c)})$.

Theorem 2. For Model (2), under the regularity conditions (A1)–(A6) and assume that $n_{hg} = O(n^v)$ with $1/(2c) < v < 0.5$, as $n \rightarrow \infty$ we have $n^{1/2}(\hat{\gamma}_{hg} - \gamma_{hg,T}) \xrightarrow{d} N(\mathbf{0}, \Gamma_{hg}(\phi_{hg,T}))$ as $n \rightarrow \infty$.

4 | NUMERICAL STUDY

We conducted several groups of simulation studies to assess the finite sample performance of the proposed estimation methods. The following scenarios were considered. Scenario 1: Random coefficients of the time-dependent covariate process were assumed to be normally distributed. Different sample size, different censoring percentages and different strength of feedback were considered: (a) sample size $n = 500$, censoring rate 50% (Censoring times were generated from a uniform distribution $U[0, 3]$ and truncated at 1), (b) $n = 1000$, censoring rate 50%, (c) $n = 500$, censoring rate 35% (Censoring times were all chosen as 1), (d) greater impact of past event feedback. Scenario 2: Random coefficients of the time-dependent covariate process were generated from other distributions: (a) mixture of normal distributions (b) mixture of normal distributions and smaller variances of the measurement errors (c) Dirichlet distributions. Scenario 3: The true models that generate the longitudinal data did not contain past event feedbacks. For all scenarios, we generated two types of recurrent events. Let S_1 denote the set of possible transition paths for type 1 recurrent events and $S_1 = \{h \in \{0, 1, \dots, M - 1, M + 1, \dots, 2M - 1\}, g \in \{\{1, \dots, M\} \cap \{r_h + 1, r_h + M + 1, \dots, r_h + (K - 1)M + 1\}\}\}$. For type 2 recurrent events, let $S_2 = \{h \in \{0, 1, \dots, M - 1, M + 1, \dots, 2M - 1\}, g \in \{\{M + 1, \dots, 2M\} \cap \{r_h + 1, r_h + M + 1, \dots, r_h + (K - 1)M + 1\}\}\}$. For terminal events, let $S_3 = \{h \in \{0, \dots, 2M\}, g \in \{2M + 1, 2M + 2\}\}$. The multistate models we considered here are

$$\begin{aligned} \lambda_{hg}^i(t) &= \lambda_{0,1}(t) \exp(\eta_1(t)X_i(t) + \gamma_{1,1}W_{i1} + \gamma_{1,2}W_{i2}(t)), h, g \in S_1, \\ \lambda_{hg}^i(t) &= \lambda_{0,2}(t) \exp(\eta_2(t)X_i(t) + \gamma_{2,1}W_{i1} + \gamma_{2,2}W_{i2}(t)), h, g \in S_2, \\ \lambda_{hg}^i(t) &= \lambda_{0,3}(t) \exp(\eta_3(t)X_i(t) + \gamma_{3,1}W_{i1} + \gamma_{3,2}W_{i2}(t)), h, g \in S_3. \end{aligned} \tag{9}$$

Here we assume that events of the same type share the same baseline intensity functions and regression coefficients, to mimic the real analysis. The estimating equations can be similarly derived with some straightforward algebra and are presented in section 2 in Appendix S1. W_{i1} is a baseline covariate generated from a Bernoulli distribution with probability 0.5. The observed data of $X_i(t)$ and $W_{i2}(t)$ are generated from the following models

$$\begin{aligned}
 W_{ij2} &= \beta_{w_{i0}} + \beta_{w_{i1}} t_{ij} + \beta_{w_{i2}} I \left(\left(\sum_{h,g} N_{hg}^i(t-) I(g \in \{1, \dots, M\}) \right) > 0 \right) + \varepsilon_{w_{ij}}, \\
 X_{ij} &= \beta_{x_{i0}} + \beta_{x_{i1}} t_{ij}^{1/2} + \beta_{x_{i2}} \min \left(\sum_{h,g} N_{hg}^i(t-), 5 \right) + \varepsilon_{x_{ij}}.
 \end{aligned} \tag{10}$$

The function of past event history in the first model implies that the intercept of $W_{i2}(t)$ will change after the first occurrence of type 1 recurrent event and the function of event history in the second model implies that the intercept of $X_i(t)$ will change after each occurrence of recurrent events of any kind and will remain the same after the fifth recurrent event. Particularly, the functional form of the past event feedback in $W_{i2}(t)$ mimic the setting of the real analysis, where we assume that the blood pressure profile is related to prior MI (type 1 recurrent event) and the levels of HDLC is associated with previous stroke (type 2 recurrent event). $\text{var}(\varepsilon_{w_{ij}}) = \text{var}(\varepsilon_{x_{ij}}) = 0.2$. The longitudinal measurement times are $t_{ij} = (0, 0.02, 0.05, 0.09, 0.15, 0.2, 0.3, 0.4, 0.5, 0.9)$. The maximum follow-up time was chosen as 1. The baseline intensity functions are $\lambda_{0,1}(t) = \lambda_{0,2}(t) = 2$ and $\lambda_{0,3}(t) = 0.1 + t^{1/2}$. The true values of the unknown coefficients are $\eta_1(t) = -0.8 \sin(\pi t)$, $\eta_2(t) = -0.6(1.5t - 0.75)^3 - 0.5$, $\eta_3(t) = -0.1 \sin(1.8\pi t + 1.3) - 0.4$ and $\gamma_1 = (0.4, 0.5)$, $\gamma_2 = (0.5, 0.4)$, $\gamma_3 = (0.6, 0.5)$. In each scenario, the sample size was chosen as 500 and simulation results were based on 500 Monte Carlo replications. Cubic splines were used to estimate the time-varying coefficients. The number of knots for each spline function was chosen by minimizing a AIC-type criterion based on the corrected log-partial likelihood function. The detailed formulas can be found in section 3.1 in Appendix S1. The time-varying coefficients and constant coefficients in all scenarios were estimated in four ways: (1) our proposed estimator (DC) where D is short for dynamic feedbacks and C is short for bias correction; (2) the naive estimator that ignores the dynamic feedbacks (NaivD); (3) the naive estimator using least squares estimate of the random coefficients without correcting bias (NaivC); and (4) the ideal estimator (I) that assumes the random coefficients to be known at each time point. For each simulation, we calculated the following summary statistics for the constant parameters: the difference between the average of the estimates and the true parameter (Bias), the Monte Carlo SD, the average of the estimated SE, and the 95% empirical CP. As for the time-varying coefficients, the average of the spline estimates, the 95% pointwise confidence band and the pointwise coverage probability were calculated. To save space, the simulation results under Scenario 1, setting (a) and Scenario 2, setting (a) are presented here and the rest are shown in Appendix S1.

4.1 | Scenario 1, setting (a)

In this simulation, the random coefficients of the time-dependent covariates were generated from normal distributions. In settings (a)–(c), the time-to-event data and longitudinal data are generated from a system with moderate feedback effects. The mean of β_{x_i} was chosen as $(1, -0.2, 0.3)$ and the variance was $(0.5, 0.1, 0.1)$ with correlations $\text{corr}(\beta_{x_{i0}}, \beta_{x_{i1}}) = 0.2$, $\text{corr}(\beta_{x_{i0}}, \beta_{x_{i2}}) = 0.8$, $\text{corr}(\beta_{x_{i1}}, \beta_{x_{i2}}) = 0.6$. β_{w_i} were sampled from a normal distribution with mean $(1, -0.5, 0.2)$, variance $(0.4, 0.2, 0.1)$ and correlations $\text{corr}(\beta_{w_{i0}}, \beta_{w_{i1}}) = 0.2$, $\text{corr}(\beta_{w_{i0}}, \beta_{w_{i2}}) = 0.8$, $\text{corr}(\beta_{w_{i1}}, \beta_{w_{i2}}) = 0.6$. In setting (d), we generated data from a stochastic processes with stronger feedback by setting $\mathbb{E}(\beta_{x_{i2}}) = 0.5$ and $\mathbb{E}(\beta_{w_{i2}}) = 0.8$ with the other parameters remaining the same.

Table 1 presents the estimation result of the constant parameters when sample size $n = 500$ and the censoring rate is around 50%. We can see that our proposed estimator is noticeably

TABLE 1 Estimation of the constant parameters in Simulation 1 setting (a)

	$\gamma_{1,1}^{DC}$	$\gamma_{1,2}^{DC}$	$\gamma_{1,1}^{NaivC}$	$\gamma_{1,2}^{NaivC}$	$\gamma_{1,1}^{NaivD}$	$\gamma_{1,2}^{NaivD}$	$\gamma_{1,1}^I$	$\gamma_{1,2}^I$
Bias	-0.0031	0.0158	-0.0298	-0.1515	-0.0146	-0.0330	-0.0052	-0.0050
SE	0.0885	0.0621	0.0867	0.0420	0.0832	0.0535	0.0767	0.0446
SD	0.0886	0.0659	0.0867	0.0448	0.0823	0.0592	0.0761	0.0451
CP	0.9589	0.9302	0.9384	0.0719	0.9528	0.8850	0.9630	0.9569
	$\gamma_{2,1}^{DC}$	$\gamma_{2,2}^{DC}$	$\gamma_{2,1}^{NaivC}$	$\gamma_{2,2}^{NaivC}$	$\gamma_{2,1}^{NaivD}$	$\gamma_{2,2}^{NaivD}$	$\gamma_{2,1}^I$	$\gamma_{2,2}^I$
Bias	0.0077	0.0120	-0.0147	-0.1212	-0.0009	-0.0246	0.0026	-0.0042
SE	0.0910	0.0631	0.0873	0.0424	0.0850	0.0542	0.0808	0.0470
SD	0.0953	0.0648	0.0946	0.0437	0.0909	0.0558	0.0856	0.0465
CP	0.0971	0.9343	0.9322	0.1807	0.9343	0.9220	0.9343	0.9671
	$\gamma_{3,1}^{DC}$	$\gamma_{3,2}^{DC}$	$\gamma_{3,1}^{NaivC}$	$\gamma_{3,2}^{NaivC}$	$\gamma_{3,1}^{NaivD}$	$\gamma_{3,2}^{NaivD}$	$\gamma_{3,1}^I$	$\gamma_{3,2}^I$
Bias	0.0155	0.0203	-0.0150	-0.1332	0.0036	-0.0305	0.0068	0.0029
SE	0.1352	0.0899	0.1290	0.0611	0.1330	0.0819	0.1288	0.0714
SD	0.1309	0.0956	0.1242	0.0648	0.1274	0.0831	0.1229	0.0743
CP	0.9589	0.9322	0.9548	0.4312	0.9569	0.8994	0.9569	0.9343

unbiased and the Monte Carlo SDs agree with the estimated standard errors. The empirical coverage probabilities are very close to their nominal values 95%. On the other hand, the NaivC estimator is seriously biased with poor coverage percentage. The NaivD estimator is also biased with coverage probability below the nominal level. The four estimators perform almost the same in terms of the coverage probabilities of the coefficients of the baseline covariates. This is expected since the baseline covariates were assumed to be precisely measured and did not contain any past event feedback. Therefore ignoring the past event feedback or the bias resulted from the measurement errors in the time-dependent covariates have less impact on the estimation of the coefficients of the baseline covariates.

Figure 2 displays the true and fitted curves of the time-varying coefficients. The estimates based on the proposed method are close to the true curves. Our proposed estimator shows better performance than the naive ones in terms of the pointwise coverage probabilities. It can be seen that the coverage rates of the time-dependent coefficients calculated via the naive methods are lower comparing to the proposed method at most of the time points.

It can be found in Figure 2 that the coverage probabilities of $\eta_2(t)$ and $\eta_3(t)$ were closer to the nominal levels comparing to those of $\eta_1(t)$. As discussed in Cummins et al. (2001), nonparametric curve estimates based on a single global smoothing parameter tend to be most highly biased at points of sharp curvature. And because bias in nonparametric estimation is not uniform, they do not hold the desired level of coverage probabilities uniformly at all design points. In our estimation procedure, there is only one global smoothing parameter which is the number of knots. And the curvature of $\eta_1(t)$ is greater than that of $\eta_2(t)$ and $\eta_3(t)$, especially around the minimum point of $\eta_1(t)$. Therefore it is expected that the coverage probabilities of $\eta_1(t)$ around the minimum point would be lower. The simulations based on a larger sample size and a lower censoring rate presented in Appendix S1 show that the performance of our proposed estimator can be further improved with a larger effective sample size. The simulation results also display a more

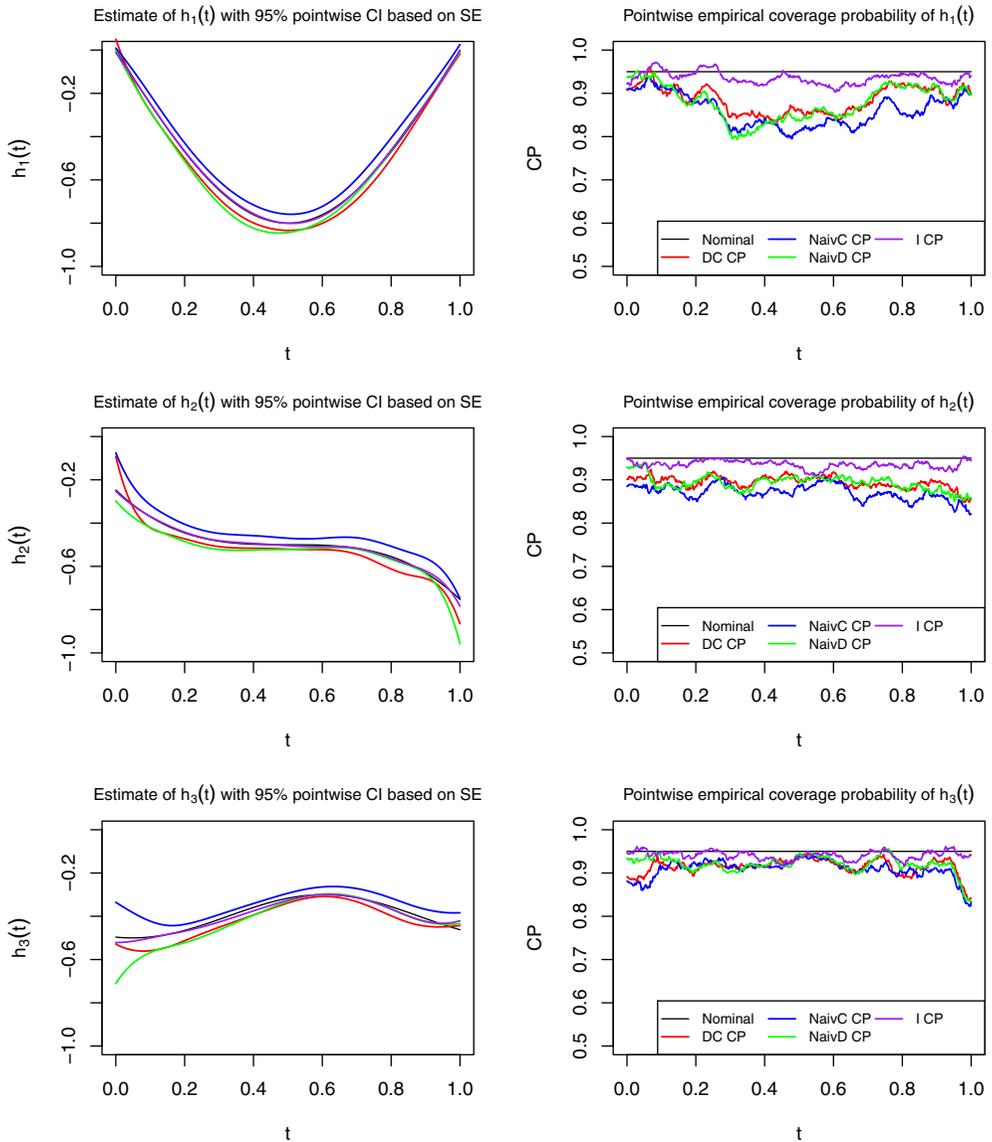


FIGURE 2 Estimation of the time-varying coefficients in Simulation 1 setting (a) [Colour figure can be viewed at wileyonlinelibrary.com]

remarkable difference between the results based on the proposed estimation method and those based on the naive methods.

4.2 | Scenario 2, setting (a)

In this simulation, the random effects of the longitudinal covariates were generated from a mixture of two normal distributions. This simulation aims to show that our estimation method is robust against random effect distributions. The mixing proportion was chosen as 0.3. The means

TABLE 2 Estimation of the constant parameters in Simulation 2 setting (a)

	$\gamma_{1,1}^{DC}$	$\gamma_{1,2}^{DC}$	$\gamma_{1,1}^{NaivC}$	$\gamma_{1,2}^{NaivC}$	$\gamma_{1,1}^{NaivD}$	$\gamma_{1,2}^{NaivD}$	$\gamma_{1,1}^I$	$\gamma_{1,2}^I$
Bias	0.0012	0.0212	-0.0281	-0.1306	-0.0152	-0.0207	0.0002	-0.0004
SE	0.0889	0.0514	0.0863	0.0365	0.0842	0.0452	0.0750	0.0371
SD	0.0897	0.0575	0.0863	0.0388	0.0876	0.0489	0.0737	0.0387
CP	0.9457	0.9144	0.9395	0.0605	0.9395	0.8914	0.9374	0.9457
	$\gamma_{2,1}^{DC}$	$\gamma_{2,2}^{DC}$	$\gamma_{2,1}^{NaivC}$	$\gamma_{2,2}^{NaivC}$	$\gamma_{2,1}^{NaivD}$	$\gamma_{2,2}^{NaivD}$	$\gamma_{2,1}^I$	$\gamma_{2,2}^I$
Bias	0.0047	0.0135	-0.0181	-0.1092	-0.0053	-0.0149	0.0000	-0.0017
SE	0.0880	0.0507	0.0863	0.0368	0.0840	0.0450	0.0799	0.0392
SD	0.0913	0.0477	0.0886	0.0343	0.0879	0.0417	0.0806	0.0361
CP	0.9395	0.9562	0.9248	0.1441	0.9290	0.9562	0.9624	0.9749
	$\gamma_{3,1}^{DC}$	$\gamma_{3,2}^{DC}$	$\gamma_{3,1}^{NaivC}$	$\gamma_{3,2}^{NaivC}$	$\gamma_{3,1}^{NaivD}$	$\gamma_{3,2}^{NaivD}$	$\gamma_{3,1}^I$	$\gamma_{3,2}^I$
Bias	0.0004	0.0194	-0.0309	-0.1171	-0.0103	-0.0203	-0.0061	0.0032
SE	0.1342	0.0752	0.1281	0.0538	0.1316	0.0667	0.1279	0.0611
SD	0.1346	0.0787	0.1283	0.0547	0.1320	0.0701	0.1258	0.0621
CP	0.9541	0.9395	0.9415	0.4196	0.9499	0.9102	0.9582	0.9332

and distances of the two distinct normal distributions being mixed were chosen in a way such that the mean of the generated random coefficients β_{Xi} and β_{Wi} were the same as those in Scenario 1. The skewness of the generated random coefficients β_{Xi} were 0.33, -0.04, and -0.15; and the skewness of β_{Wi} were 0.20, -0.15, and -0.04. The variances of β_{Xi} were 0.86, 0.12, and 0.12; and the variances of β_{Wi} were 0.75, 0.23, and 0.13. The censoring rate was around 50%. The estimation results are presented in Table 2 and Figure 3. It is shown that the naive estimators are biased with poor coverage probabilities. The proposed estimators outperform the naive estimators. The coverage probabilities of $\hat{\eta}_1(t)$ around its minimum point was a bit lower than the nominal level due to the greater curvature. But it can be shown by the results of setting (b) which are presented in Appendix S1 that the coverage probabilities improve with the decrease of the variance of the measurement errors. The simulation result of setting (c) shows that the proposed estimators still perform well when the random coefficients were generated from other asymmetric distributions with greater skewness.

5 | APPLICATION

In this section, we apply the proposed model and estimation method to the ARIC study described in the Introduction. Our aim is to investigate the association of the biomarker trajectory and other established risk factors with recurrent MI, stroke, and total mortality. We restricted our attention to White male participants living in Washington County and Suburban Minneapolis, who had no history of stroke or MI. The events of interest were definite MI, probable MI, definite stroke, probable stroke, and all death events before 2011. The follow-up data of biomarker measurements and censoring information before 2011 were used. The final sample included 2764 subjects with 445 MI events, 210 stroke events, and 29.2% mortality rate. The covariates considered were longitudinal HDLC (mmol/L), longitudinal SBP (in mmHg) and the following baseline

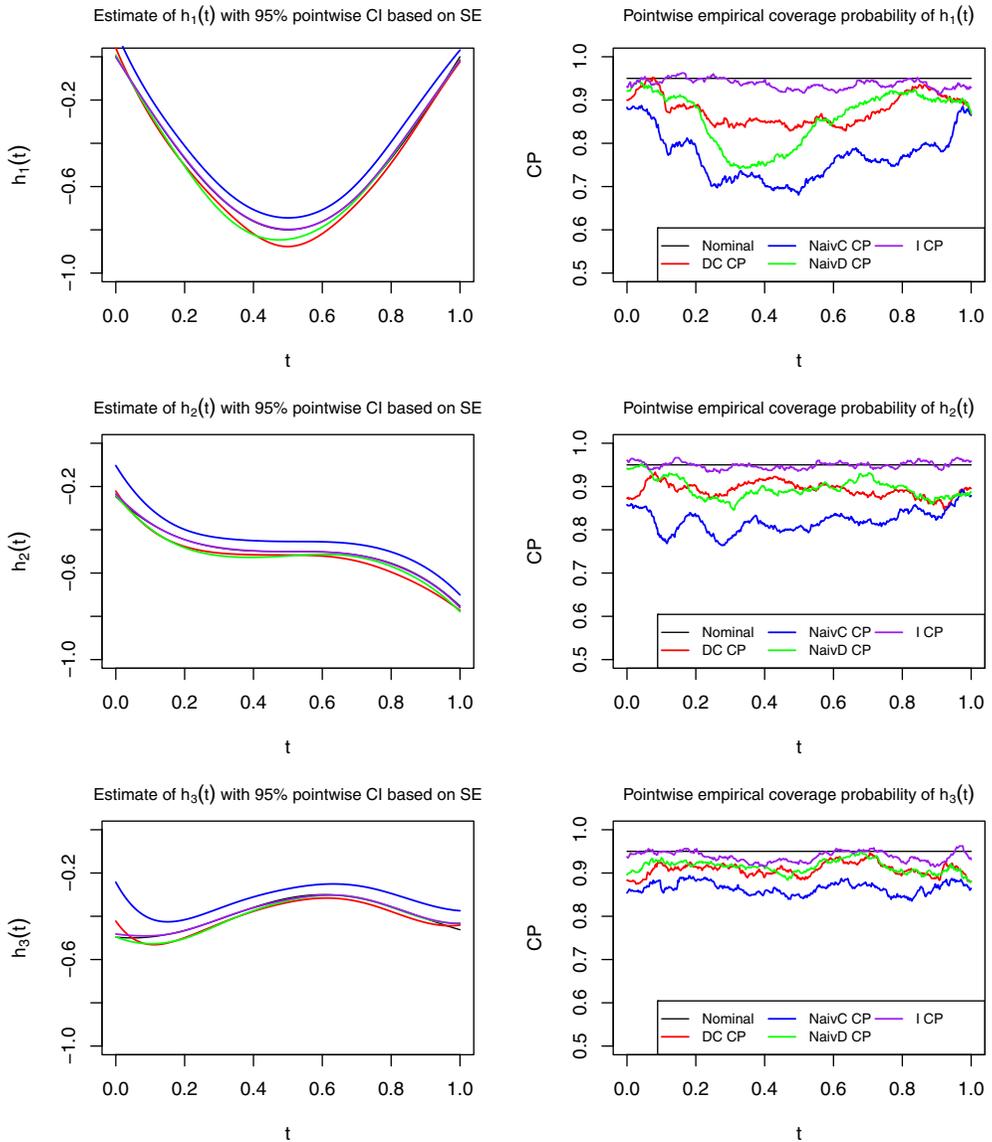


FIGURE 3 Estimation of the time-varying coefficients in Simulation 2 setting (a) [Colour figure can be viewed at wileyonlinelibrary.com]

covariates: age (in years); indicators for diabetes (1 for fasting glucose ≥ 126 mg/dl and 0 otherwise), indicators for hypertension medication (1 for hypertension lowering medication within past 2 weeks at baseline examination, and 0 otherwise) and smoking (1 for current smoker at baseline examination and 0 otherwise). Among them, age was divided by 10 and SBP measurements were divided by 50. All event times and longitudinal observation times were divided by 365. The structure of time-varying coefficients and constant parameters in the multi-state models were chosen based on some preliminary analysis treating the effects of all covariates nonparametrically. It was found that the coefficients of SBP, HDLC and hypertension did not change much over time and in addition, the effect of SBP and HDLC were not significant over most of the time.

Therefore we assumed that these three covariates have constant effects. Using the same notations of the sets of possible transition paths S_1 , S_2 , and S_3 in Section 4, we considered the following models

$$\begin{aligned} \lambda_{hg}^i(t) &= \lambda_{0,1}(t) \exp(\eta_{1,1}(t)\text{age}_i + \eta_{1,2}(t)\text{diab}_i + \eta_{1,3}(t)\text{smoke}_i \\ &\quad + \gamma_{1,1}\text{SBP}_i(t) + \gamma_{1,2}\text{HDLC}_i(t) + \gamma_{1,3}\text{hyp}_i), \quad h, g \in S_1, \\ \lambda_{hg}^i(t) &= \lambda_{0,2}(t) \exp(\eta_{2,1}(t)\text{age}_i + \eta_{2,2}(t)\text{diab}_i + \eta_{2,3}(t)\text{smoke}_i \\ &\quad + \gamma_{2,1}\text{SBP}_i(t) + \gamma_{2,2}\text{HDLC}_i(t) + \gamma_{2,3}\text{hyp}_i), \quad h, g \in S_2, \\ \lambda_{hg}^i(t) &= \lambda_{0,3}(t) \exp(\eta_{3,1}(t)\text{age}_i + \eta_{3,2}(t)\text{diab}_i + \eta_{3,3}(t)\text{smoke}_i \\ &\quad + \gamma_{3,1}\text{SBP}_i(t) + \gamma_{3,2}\text{HDLC}_i(t) + \gamma_{3,3}\text{hyp}_i), \quad h, g \in S_3, \end{aligned} \tag{11}$$

where M is the maximum number of nonterminal events observed among all subjects. $M = 6$ in our sample. The first model is for recurrent MI events, the second model is for recurrent stroke events and the third one is for fatal events. We assumed that the intensity functions of the same type of events share the same baseline intensities and regression coefficients. The models of SBP and HDLC were chosen as follows according to the longitudinal data trajectories shown in Appendix S1 and the biological knowledge discussed in Section 2. $SBP_{ij} = \beta_{S10} + \beta_{S11}t_{ij} + \beta_{S12}I\left(\left(\sum_{h,g} N_{hg}^i(t_{ij-})I(g \in \{1, \dots, M\})\right) > 0\right) + \varepsilon_{Sij}$, $HDLC_{ij} = \beta_{H10} + \beta_{H11}t_{ij} + \beta_{H12}I\left(\left(\sum_{h,g} N_{hg}^i(t_{ij-})I(g \in \{M + 1, \dots, 2M\})\right) > 0\right) + \varepsilon_{Hij}$. The first model indicates that the intercept of the SBP trajectory is associated with prior MI event history and the second model implies that the intercept of HDLC trajectory is associated with prior stroke event history. For comparison, both the estimation results from the full model (assuming event feedback in the longitudinal covariate) and the reduced model (assuming no event feedback) are reported. The estimated effects of SBP, HDLC and hypertension are presented in Table 3. The estimated effects of age, diabetes, and smoking are shown in Figures 4–6.

From Table 3, we can see that the current levels of SBP were positively and significantly associated with the intensity of stroke events. This suggests that an increase of 50 mmHg of SBP level was related to a $\exp(1.2239) = 3.4$ higher risk of stroke occurrences. This is consistent with existing results in the medical research of the ARIC data. Shahar et al. (2003) found that baseline SBP was positively associated with incident stroke events and Petruski-Ivleva et al. (2016) found that an increase in SBP independent of baseline is associated with increased risk of stroke. In our model

TABLE 3 Estimated parameters and SEs for the Atherosclerosis Risk in Communities data

		Full model			Reduced model		
		SBP	HDLC	hyp	SBP	HDLC	hyp
MI	Est	-0.1739	-0.4812	0.3510	0.2425	-2.6418	0.1413
	SE	1.1427	0.8025	0.1931	0.2047	0.5404	0.0856
Stroke	Est	1.2239	-0.5548	0.0888	0.7831	-1.2282	0.2352
	SE	0.3970	0.7219	0.1382	0.2043	0.5369	0.0840
Death	Est	-0.5909	-0.3459	0.5089	0.1148	-0.5786	0.2104
	SE	1.1547	0.8713	0.1670	0.1332	0.3740	0.0877

Abbreviations: HDLC, high-density lipoprotein cholesterol; SBP, systolic blood pressure.

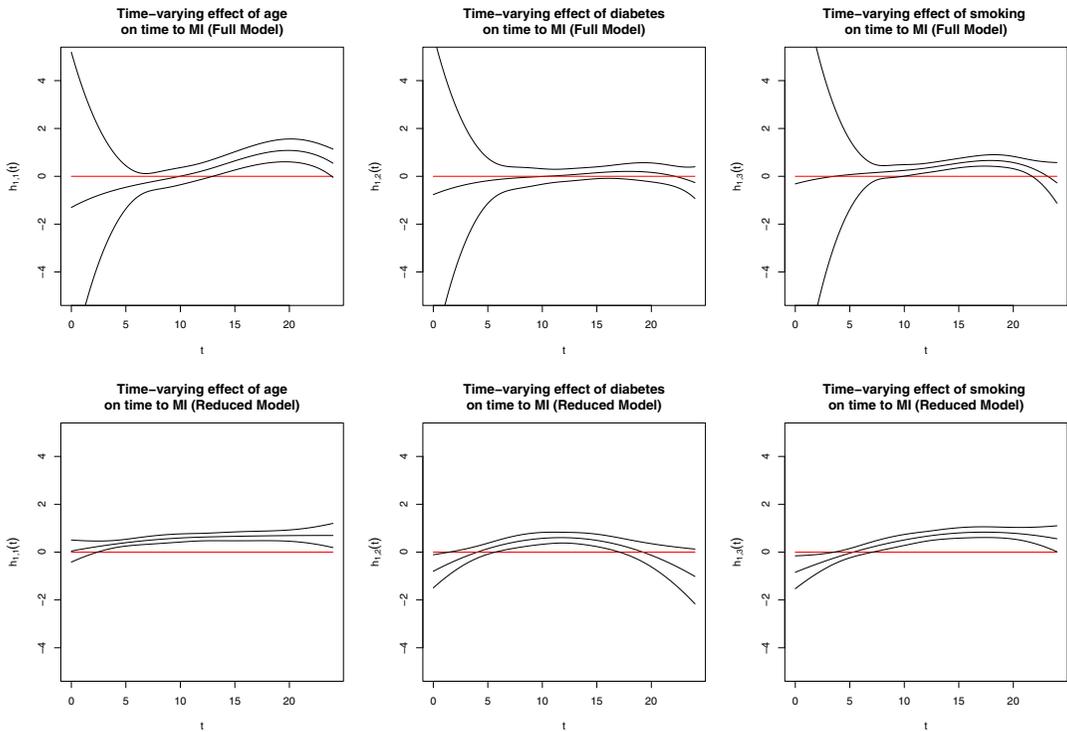


FIGURE 4 Estimation of time-varying effects of age (10 years), diabetes and smoking on time to recurrent myocardial infarction for the Atherosclerosis Risk in Communities data [Colour figure can be viewed at wileyonlinelibrary.com]

the current level of blood pressure, $SBP_i(t)$, is essentially the sum of the baseline SBP β_{Si0} and the change of SBP level along time $\beta_{Si1}t + \beta_{Si2}I\left(\left(\sum_{h,g} N_{hg}^i(t-I)(g \in \{1, \dots, M\})\right) > 0\right)$. Since both the baseline SBP level and the increase in SPB along time are positively associated with the risk of stroke occurrences, it is expected that the current levels of SBP, which summarize both effects, are positively associated with the stroke event intensity. As shown in Table 3, the current levels of SBP was found to be negatively associated with the risk of MI event and death, but these effects are not significant. Although existing research found that the baseline SBP was positively associated with the risk of MI and death (Mok et al., 2018), some studies found that the slope of SBP was negatively associated with CVD events (Barrett et al., 2019). This may help in understanding why the associations between the current levels of SBP and the risks of MI and death are not significant as the effects of the baseline SBP level and the effects of the slope of SBP have opposing signs. HDLC was negatively related to the risk of all types of events, but neither of the effects are significant. Taking hypertension medication was significantly and positively associated with death, indicating that subjects with hypertension medication use are more likely to experience a fatal event. It can be seen from Figures 4 and 6 that the baseline age and smoking status were positively associated with the risks of MI events and death. These effects become stronger and significant as time moves on. Figure 5 shows that the baseline age, diabetes and smoking status were all positively associated with the risks of stroke events over most of the time.

Comparing the results obtained from the full model and the reduced model, it can be found that the strength and the significance of some coefficients are quite different. We ran a preliminary

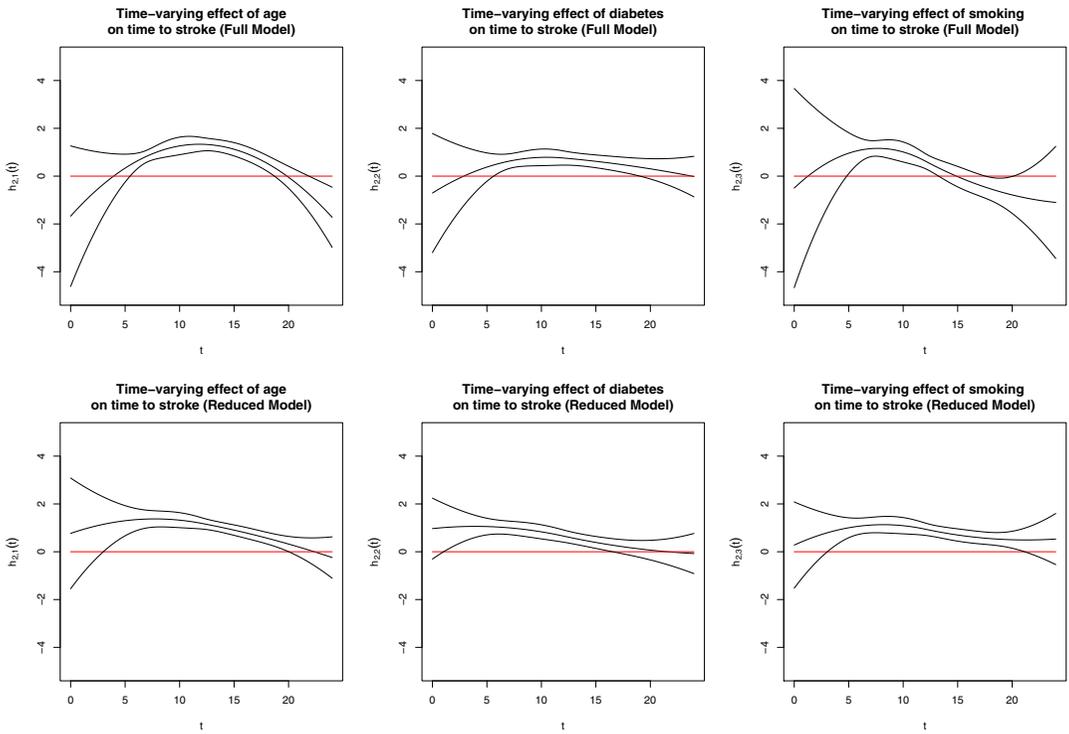


FIGURE 5 Estimation of time-varying effects of age (10 years), diabetes and smoking on time to recurrent stroke for the Atherosclerosis Risk in Communities data [Colour figure can be viewed at wileyonlinelibrary.com]

longitudinal data analysis of the SBP and HDLC data using linear mixed effects models and the results show that the fixed effects of the CVD event history are significant for both biomarkers. Therefore it is necessary to include the event history term in the time-dependent covariate models and ignoring the past event feedbacks yields biased estimation results as shown in the simulation studies. Therefore the statistical evidence provided by the reduced model may be invalid.

6 | DISCUSSION

This paper proposed a novel modeling framework for the analysis of multitype recurrent event and terminal event data. Our methodology has great flexibilities and are computationally efficient. First of all, the formulations of the intensity functions are allowed to depend on the event order or/and event types. Various forms of event feedbacks can be considered. Second, no distribution assumptions are required for the random effects in the covariate process. Third, standard Cox regression models are extended to include both time-dependent coefficients and constant coefficients. The use of regression splines provides an computationally efficient estimate of the time-dependent coefficients, comparing to the local polynomials approach where estimation should be implemented at a series of dense grid points. Fourth, an asymptotically unbiased estimating equation is developed and the variance estimators of the coefficients are derived. The variance estimators can be used to construct hypothesis tests and confidence intervals without going to bootstrap methods.

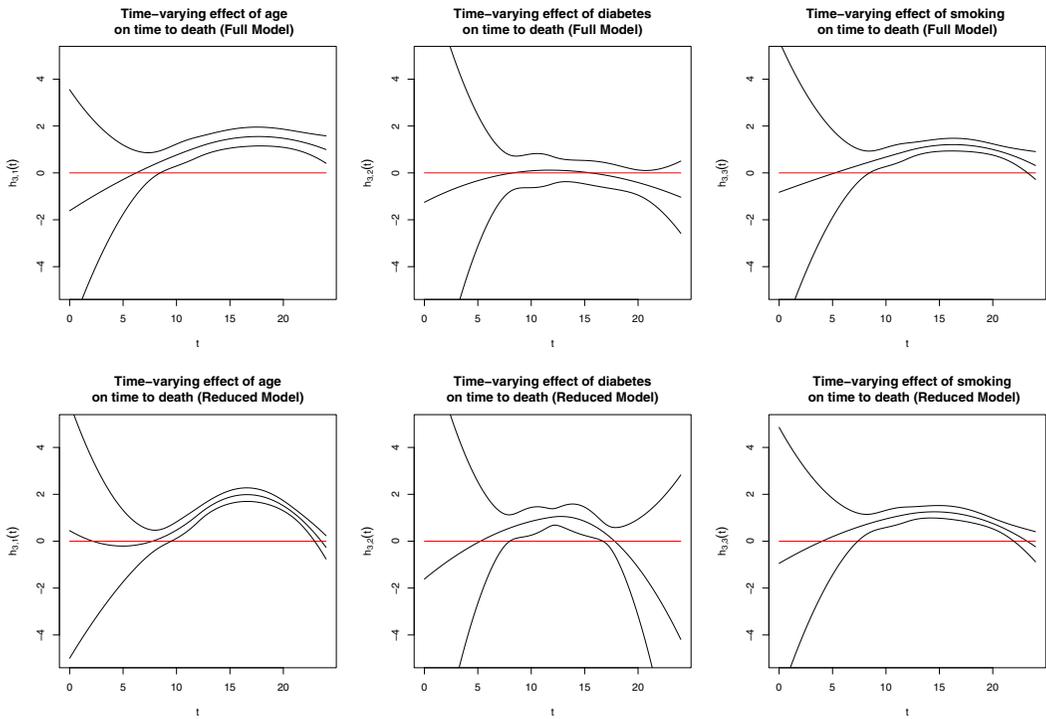


FIGURE 6 Estimation of time-varying effects of age (10 years), diabetes and smoking on time to death for the Atherosclerosis Risk in Communities data [Colour figure can be viewed at wileyonlinelibrary.com]

In the real analysis, we assumed that the prior CVD history provide feedback to the future occurrences of CVD events through the biomarkers SBP and HDLC. Though there is medical evidence for the association between prior CVD events and the trajectories of SBP and HDLC, a formal hypothesis test for testing the existence of past event feedback is desirable. In addition, we assumed a Markov model for the multistate events. However in some applications the form of past event feedback may result in assumption violations, for example when the event history is a function of state duration time. In this case, the multi-state model with state duration times as event feedbacks is a non-Markov model. Therefore model checking techniques to evaluate the Markov assumption is essential. And an extension of the proposed modeling framework to the semi-Markov or non-Markov settings will be left as future work.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the two anonymous reviewers for their insightful comments which helped improving the paper. This research was supported by the Dean's Award Scholarship, the University of Manchester.

ORCID

Jianxin Pan  <https://orcid.org/0000-0002-7460-6350>

REFERENCES

- Aalen, O. O., Fosen, J., Weedon-Fekjær, H., Borgan, Ø., & Husebye, E. (2004). Dynamic analysis of multivariate failure time data. *Biometrics*, 60, 764–773.

- Amarenco, P., Kim, J. S., Labreuche, J., Charles, H., Abtan, J., Béjot, Y., Cabrejo, L., Cha, J.-K., Ducrocq, G., Giroud, M., & Guidoux, C. (2020). A comparison of two LDL cholesterol targets after ischemic stroke. *New England Journal of Medicine*, *382*, 9–19.
- Amarenco, P., & Labreuche, J. (2009). Lipid management in the prevention of stroke: Review and updated meta-analysis of statins for stroke prevention. *The Lancet Neurology*, *8*, 453–463.
- Barrett, J. K., Huille, R., Parker, R., Yano, Y., & Griswold, M. (2019). Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC study. *Statistics in Medicine*, *38*, 1855–1868.
- Brown, P. M., & Ezekowitz, J. A. (2019). Frailty modelling for multitype recurrent events in clinical trials. *Statistical Modelling*, *1471082X17746538*, *19*, 140–156.
- Cai, Z., & Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*, *30*, 93–111.
- Chen, B. E., Cook, R. J., Lawless, J. F., & Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine*, *24*, 671–691.
- Chen, X., Wang, Q., Cai, J., & Shankar, V. (2012). Semiparametric additive marginal regression models for multiple type recurrent events. *Lifetime Data Analysis*, *18*, 504–527.
- Cook, R. J., Yi, G. Y., Lee, K.-A., & Gladman, D. D. (2004). A conditional Markov model for clustered progressive multistate processes under incomplete observation. *Biometrics*, *60*, 436–443.
- Cummins, D. J., Filloon, T. G., & Nychka, D. (2001). Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage. *Journal of the American Statistical Association*, *96*, 233–246.
- Dai, H., & Pan, J. (2018). Joint modelling of survival and longitudinal data with informative observation times. *Scandinavian Journal of Statistics*, *45*, 571–589.
- De Boor, C. (2001). *A practical guide to splines Applied Mathematical Series* (revised ed.). Springer-Verlag.
- Elisaf, M. S., Kalaitzidis, R. G., Goudevenos, J. A., Katsaraki, A. E., Sideris, D. A., & Siamopoulos, K. C. (1999). Blood pressure profile in patients with microvascular angina. *Coronary Artery Disease*, *10*, 257–259.
- Fisher, L. D., & Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, *20*, 145–157.
- Fosen, J., Borgan, Ø., Weedon-Fekjær, H., & Aalen, O. O. (2006). Dynamic analysis of recurrent event data using the additive hazard model. *Biometrical Journal*, *48*, 381–398.
- Gjessing, H. K., Røysland, K., Pena, E. A., & Aalen, O. O. (2010). Recurrent events and the exploding Cox model. *Lifetime Data Analysis*, *16*, 525–546.
- Huang, Y.-H., Hwang, W.-H., & Chen, F.-Y. (2016). Improving efficiency using the Rao–Blackwell theorem in corrected and conditional score estimation methods for joint models. *Biometrics*, *72*, 1136–1144.
- Ieva, F., Jackson, C. H., & Sharples, L. D. (2017). Multi-state modelling of repeated hospitalisation and death in patients with heart failure: The use of large administrative databases in clinical epidemiology. *Statistical Methods in Medical Research*, *26*, 1350–1372.
- Khan, N., Zhao, Y., Rabkin, S., Izadnegahdar, M., Park, J., Guan, M., Lee, M., Grubisic, M., Peng, D., Daniele, P., & Chan, S. (2017). How low is too low? analysis of systolic and diastolic blood pressure in patients with and without prior cardiovascular disease in the sprint trial. *Canadian Journal of Cardiology*, *33*, S199–S200.
- Khan, N. A., Rabkin, S. W., Zhao, Y., McAlister, F. A., Park, J. E., Guan, M., Chan, S., & Humphries, K. H. (2018). Effect of lowering diastolic pressure in patients with and without cardiovascular disease: Analysis of the sprint (systolic blood pressure intervention trial). *Hypertension*, *71*, 840–847.
- Kim, S., Zeng, D., Chambless, L., & Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in Biosciences*, *4*, 262–281.
- Li, G., & Lu, X. (2015). A bayesian approach for instrumental variable analysis with censored time-to-event outcome. *Statistics in Medicine*, *34*, 664–684.
- Mok, Y., Sang, Y., Ballew, S. H., Rebholz, C. M., Rosamond, W. D., Heiss, G., Folsom, A. R., Coresh, J., & Matsushita, K. (2018). American heart association's life's simple 7 at middle age and prognosis after myocardial infarction in later life. *Journal of the American Heart Association*, *7*, e007658.
- Nan, B., Lin, X., Lisabeth, L. D., & Harlow, S. D. (2005). A varying-coefficient Cox model for the effect of age at a marker event on age at menopause. *Biometrics*, *61*, 576–583.
- Nathoo, F., & Dean, C. (2008). Spatial multistate transitional models for longitudinal event data. *Biometrics*, *64*, 271–279.

- Paige, E., Barrett, J., Pennells, L., Sweeting, M., Willeit, P., Di Angelantonio, E., Gudnason, V., Nordestgaard, B. G., Psaty, B. M., Goldbourt, U., & Best, L. G. (2017). Use of repeated blood pressure and cholesterol measurements to improve cardiovascular disease risk prediction: An individual-participant-data meta-analysis. *American Journal of Epidemiology*, *186*, 899–907.
- Peña, E. A. (2006). Dynamic modelling and statistical analysis of event times. *Statistical Science*, *21*, 1.
- Petruski-Ivleva, N., Viera, A. J., Shimbo, D., Muntner, P., Avery, C. L., Schneider, A. L., Couper, D., & Kucharska-Newton, A. (2016). Longitudinal patterns of change in systolic blood pressure and incidence of cardiovascular disease: The atherosclerosis risk in communities study. *Hypertension*, *67*, 1150–1156.
- Shahar, E., Chambless, L. E., Rosamond, W. D., Boland, L. L., Ballantyne, C. M., McGovern, P. G., & Sharrett, A. R. (2003). Plasma lipid profile and incident ischemic stroke: The atherosclerosis risk in communities (ARIC) study. *Stroke*, *34*, 623–631.
- Song, X., & Wang, C. (2008). Semi-parametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics*, *64*, 557–566.
- Song, X., & Wang, L. (2017). Partially time-varying coefficient proportional hazards models with error-prone time-dependent covariates—an application to the aids clinical trial group 175 data. *The Annals of Applied Statistics*, *11*, 274–296.
- Steele, F. (2011). Multilevel discrete-time event history models with applications to the analysis of recurrent employment transitions. *Australian & New Zealand Journal of Statistics*, *53*, 1–20.
- The ARIC investigators. (1989). The atherosclerosis risk in communities (ARIC) study: Design and objectives. *American Journal of Epidemiology*, *129*, 687–702.
- Tsiatis, A. A., & Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, *88*, 447–458.
- Wattanakit, K., Folsom, A. R., Chambless, L. E., & Nieto, F. J. (2005). Risk factors for cardiovascular event recurrence in the atherosclerosis risk in communities (ARIC) study. *American Heart Journal*, *149*, 606–612.
- Woo, J., Lam, C., Kay, R., Wong, H., Teoh, R., & Nicholls, M. (1990). Acute and long-term changes in serum lipids after acute stroke. *Stroke*, *21*, 1407–1411.
- Yatsuya, H., Folsom, A. R., Alonso, A., Gottesman, R. F., & Rose, K. M. (2011). Postural changes in blood pressure and incidence of ischemic stroke subtypes: The ARIC study. *Hypertension*, *57*, 167–173.
- Yiu, S., Farewell, V. T., & Tom, B. D. (2018). Clustered multistate models with observation level random effects, mover–stayer effects and dynamic covariates: Modelling transition intensities and sojourn times in a study of psoriatic arthritis. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *67*, 481–500.
- Yu, H., Cheng, Y.-J., & Wang, C.-Y. (2018). Methods for multivariate recurrent event data with measurement error and informative censoring. *Biometrics*, *74*, 966–976.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ma, C., & Pan, J. (2021). Multistate analysis of multitype recurrent event and failure time data with event feedbacks in biomarkers. *Scandinavian Journal of Statistics*, 1–22. <https://doi.org/10.1111/sjos.12545>