MODELING PAST EVENT FEEDBACK THROUGH BIOMARKER DYNAMICS IN THE MULTI-STATE EVENT ANALYSIS FOR CARDIOVASCULAR DISEASE DATA

By Chuoxin Ma [*], Hongsheng Dai [†] and Jianxin Pan [*]

Department of Mathematics, The University of Manchester [*] and Department of Mathematical Sciences, University of Essex[†]

In cardiovascular studies, we often observe ordered multiple events along disease progression, which are essentially a series of recurrent events and terminal events with competing risk structure. One of the main interest is to explore the event specific association with the dynamics of longitudinal biomarkers. New statistical challenge arises when the biomarkers carry information from the past event history, providing feedbacks for the occurrences of future events, and particularly when these biomarkers are only intermittently observed with measurement errors. In this paper, we propose a novel modelling framework where the recurrent events and terminal events are modelled as multi-state process and the longitudinal covariates that account for event feedbacks are described by random effects models. Considering the nature of long-term observation in cardiac studies, flexible models with semiparametric coefficients are adopted. To improve computation efficiency, we develop an one-step estimator of the regression coefficients and derive their asymptotic variances for the computation of the confidence intervals, based on the proposed asymptotically unbiased estimating equation. Simulation studies show that the naive estimators which either ignore the past event feedbacks or the measurement errors are biased. Our method achieves better coverage probability, compared to the naive methods. The model is motivated and applied to a dataset from the Atherosclerosis Risk in Communities Study.

1. Introduction. Identifying risk factors associated with the course of cardiovascular disease (CVD) is of great medical interest. A main feature of the practice presents difficulties in the estimation of the association between risk factors and CVD events. That is, a constellation of events of different types are observed from the same subject and the occurrence of the previous events may affect the risk of the subsequent ones through the associated biomarkers. Specifically, a subject may experience for example recurrent myocardial infarction(MI) and may be followed by cardiovascu-

1

lar death. Events of different types occurred to the same subject can not be simply assumed to be independent. As pointed out by Kim et al. (2012) and Rogers et al. (2016), the risk of recurrent heart failures and myocardial infarction are associated with that of the fatal CVD events. The successive events are actually a nested series of competing risk events. After occurrence of MI, as long as the subject is still under observation, he/she can possibly encounter another MI or death in the future. The order of the recurrent events and terminal event is informative as it reflects disease progression. It is expected that risk factors associated with different types of events would be different. The strength of association can also vary between recurrences. Furthermore, it is very likely that CVD occurrences in the past can cause a change in the profile of the associated biomarkers and the resulting changes of the trajectory will further affect the proneness to new events in the future. When such effect accumulates with time, the feedback from past event history gets increasingly stronger. Some clinical trial experience (Elisaf et al., 1999; de la Sierra et al., 2010; Zhao et al., 2017) shows that the blood pressure, which is an established biomarker for CVD, is influenced by prior heart disease history. For illustration, Figure 1 depicts an example of the heart disease event flow and the influence of past event history.
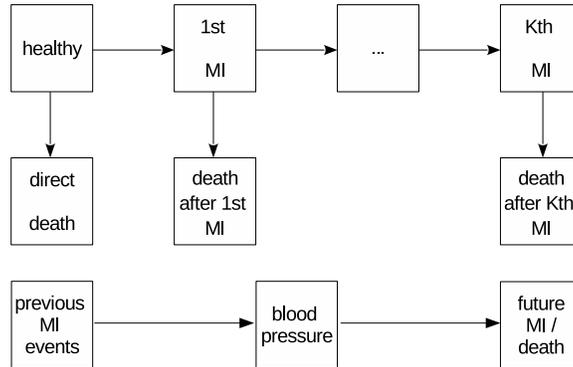
FIG 1. *Example for multiple events and past event feedback in CVD.*

The data we analyse in this paper are from the Atherosclerosis Risk in Communities (ARIC) Study (Wattanakit et al., 2005), which is an ongoing CVD surveillance study sponsored by the National Heart, Lung, and Blood Institute. One of the objectives is to investigate associations of the levels of, and changes in, risk factors with CVD occurrences. It is conducted in four US communities: Forsyth County, North Carolina; Jackson, Mississippi; Suburban Minneapolis, Minnesota and Washington County, Maryland. The study

investigates various types of clinical events (e.g., hospitalized myocardial infarction, stroke, heart failure and fatal coronary heart disease, etc.). Events are enumerated and validated through cohort/community surveillance and annual follow-up. In addition, information on cardiovascular risk factors are also collected in seven clinic examinations. The first four cohort examinations were conducted three-year apart in 1987-1989, 1990-1992, 1993-1995, 1996-1998, the fifth and sixth examinations were completed in 2011-2013 and 2016-2018. And the last one is currently in progress. Existing research on the ARIC data mostly focused on the incident heart disease events or cardiovascular death. Few of these literature considered the subsequent events which are possibly recurrent and of multi-type. However, as addressed by Ip et al. (2015) and Rogers et al. (2016), it is the series of recurrent events and fatal events that reflects the true CVD burden faced by an individual. Analysis of the overall disease burden invites a new model that can address the order and the timing of the cardiac events, as well as the correlation structure between them. Another methodological issue posed by the ARIC data is that the biomarkers are only measured in seven repeated examinations and are possibly error-contaminated. Even if the bias due to the imprecise and intermittent time-dependent covariate is corrected using existing methods, for example the corrected score approach (Wang, 2006; Song and Wang, 2008) or the conditional score approach (Tsiatis and Davidian, 2001), the past event feedback through the biomarkers is still not properly addressed. It is also worth noting that the ARIC study has been carrying out for more than 30 years. Models with time-independent regression coefficients may be too restrictive and unrealistic due to the nature of long-term observation. And sometimes the functional form of the relationships between event history and the biomarkers are of interest.

The multi-state model is a well-recognized tool for analysing the ordered multiple event data. The evolution of the CVD process can be viewed as a stochastic process with multiple states and the transition to a new state is made when a certain type of event is occurred, such as the occurrence of myocardial infarction or cardiovascular death. In this paper, we focus on Markov regression model for multi-state data, in which the dependence between future and past state transitions is fully captured by the covariate information at current state. Some authors considered conditional Markov model with unobserved random effects to explain the intra-subject correlation between different transitions intensities. Cook et al. (2004) introduced multivariate random effects to describe the dependence among events in clustered progressive multi-state processes. Bedair et al. (2016) considered a multivariate frailty model to characterize the event rate of multi-type

recurrent event data, taking into account the dependence among different event types. The random effects may help in understanding the dependence between state transitions. However, they may not be adequate for characterizing the dependence structure between the state transition times (Aalen et al., 2004). As shown in the heart disease example, the dynamic feedbacks in the longitudinal covariate also induce association between successive events. Using dynamic covariates instead of frailty terms allows mechanistic interpretation and makes the model easier to fit with a counting process model and suitable intensity processes (Aalen et al., 2004; Aalen, Borgan and Gjessing, 2008).

Several authors have proposed to incorporate the past event feedback in the form of a separate covariate in the intensity model like the number of prior events or some simple functions thereof. Existing literature includes, but not limits to Miloslavsky et al. (2004), Peña (2006), Borgan et al. (2007), Cook and Lawless (2007) and Gjessing et al. (2010). However, these methods did not consider the underlying process in the motivating data, where the impact of past events is mediated through biomarkers. Note that if for example, the blood pressure levels are known at all event time points and are exactly measured, the past event feedback shall be well reflected by the trajectory of the blood pressure. And a marginal analysis without specifying the probability structure of the event feedback mechanism may still be valid. However, given the blood pressure is only measured in seven clinic examinations, the estimators of the associations between event occurrences and the levels of biomarkers will be biased if the specific feedback mechanism is not incorporated into the analysis. Dai and Pan (2018) recently presented a joint model of survival and longitudinal data where the influence of informative observation times on the longitudinal biomarker is formulated as a function of event history. Although featuring the influence of past event history on the biomarker trajectory, their proposed method cannot be tailored to meet the needs of our motivating data. Treating time to intermediate events as informative time points is more specifically suited to analysis with a main focus on the longitudinal process, while our main research interest is in the time-to-event process. Therefore it is essential to develop a modelling framework which can describe the prior event feedback as dynamics of the multi-state process.

In this paper, we propose a class of multi-state models with past event feedback in longitudinal covariates and derive an efficient estimation method based on the asymptotically unbiased estimating equations and the one-step algorithm. This new methodology can fully characterize the features of CVD progression and obtain a valid estimate of the association between risk fac-

tors and the overall CVD burden. More specifically, the risks of the ordered recurrent events and terminal events are modelled as state transition intensity functions with true underlying time-dependent covariates and semiparametric coefficients. The true time-dependent covariate process follow a random effects model that allows past event impact and the longitudinal observations are the true values plus measurement errors. No distribution assumption is required for the random effects. We construct asymptotically unbiased estimating equations of the semiparametric coefficients and derive the one-step estimators which are computationally efficient. In addition, the asymptotic properties of the proposed estimators are derived. Using these results, hypothesis testing or confidence interval can be easily constructed without going to Monte Carlo resampling or bootstrap approach. The rest of this article is organized as follows. Section 2 describes the models and the estimation procedures. Section 3 presents the simulation study and real data analysis is given in Section 4. The theoretical properties of the proposed estimators and the corresponding proofs are provided in the Supplementary Material.

2. Models and Estimation Procedure.

2.1. *The Data and Models.* We first present the serial recurrent events and terminal event in a multi-state formulation. To facilitate the description of the event process and the derivation of the estimation method, we use counting process notations. The models of the event intensities and the associated biomarker trajectories are then presented and also compared to the existing models in literature.

Suppose that there are n independent individuals and the period of observation time is $[0, L]$. The event process is re-expressed in a way such that an individual will transit to the next intermediate state when a recurrent event occurs, or enter an absorbing state when a terminal event occurs. A diagram that illustrates the numbering of states and possible transition paths is presented in the Supplementary Material (Section 1). Let K be the maximum number of recurrent events observed across all individuals. We number the intermediate states as $\{0, 1, ..., K\}$ and the absorbing states as $\{K + 1, K + 2, ..., 2K + 1\}$. For example, being absorbed into state $K + 2$ indicates that an individual had previously experienced 1 transition to the intermediate state. This is equivalent to say that a recurrent event has been occurred to this individual before death. Let $N_{k-1,k}^i(t)$ denote the counting process of individual i recording the transition between intermediate states $k - 1$ and k before time t, for $i = 1, .., n$ and $k = 1, ..., K$. Denote $N_{k-(K+1),k}^i(t)$ as the counting process for the transition from an intermedi-

ate state $k - (K + 1)$ to a terminal state k, for $k = K + 1, ..., 2K + 1$. Note that $dN^i_{k-1,k}(t) = 1$ indicates that subject i experiences the kth recurrent event at time t and $dN^i_{k-(K+1),k}(t) = 1$ indicates that a terminal event is occurred at time t after $k - (K + 1)$ recurrent events. Let $Y^i_j(t)$ be an indicator taking value 1 if subject i is at risk for the transition out of state j at time t for $j = 0, ..., K$. Let $\mathcal{H}_i(t)$ be the history of the multi-state process and the covariate process.

Denote $\mathbf{X}_i(t)$ as a covariate vector with $\mathbf{X}_i(t) = (X_{i1}(t), ..., X_{ip_x}(t))^{\mathrm{T}}$ and $\mathbf{W}_i(t)$ as a covariate vector with $\mathbf{W}_i(t) = (W_{i1}(t), ..., W_{ip_w}(t))^{\mathrm{T}}$. These two sets of covariates are different in that $\mathbf{X}_i(t)$ have time-varying effects on the event intensities while $\mathbf{W}_i(t)$ have constant effects. The components in both $\mathbf{X}_i(t)$ and $\mathbf{W}_i(t)$ may be baseline covariate or time-dependent covariate. Some of them may contain past event history information. For ease of presentation, we incorporate the same covariates in all transition intensities. Extension to a more general case where intensities of different state transitions depend on different covariates can be easily derived by changing notations. Assuming multiplicative Markov models, the intensity process of the transition to an intermediate state can be written as

$$\lambda_{i,k}(t) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{pr}\{dN^i_{k-1,k}(t + \Delta t) = 1 \mid Y^i_{k-1}(t) = 1, \mathcal{H}_i(t)\}}{\Delta t}$$

$$(2.1) \qquad = Y^i_{k-1}(t)\lambda^k_0(t) \exp\left(\boldsymbol{\eta}^{\mathrm{T}}_k(t)\mathbf{X}_i(t) + \boldsymbol{\gamma}^{\mathrm{T}}_k\mathbf{W}_i(t)\right), k = 1, ..., K.$$

The intensity process of the transition to a terminal state can be written as

$$\lambda_{i,k}(t) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{pr}\{dN^i_{k-(K+1),k}(t + \Delta t) = 1 \mid Y^i_{k-(K+1)}(t) = 1, \mathcal{H}_i(t)\}}{\Delta t}$$

$$(2.2)$$

$$= Y^i_{k-(K+1)}(t)\lambda^k_0(t) \exp\left(\boldsymbol{\eta}^{\mathrm{T}}_k(t)\mathbf{X}_i(t) + \boldsymbol{\gamma}^{\mathrm{T}}_k\mathbf{W}_i(t)\right), k = K + 1, ..., 2K + 1,$$

where in both models, $\lambda^k_0(t)$ are unspecified baseline intensity functions, $\mathbf{X}_i(t)$ are covariates with time-varying coefficients $\eta_k(t)$ and $\mathbf{W}_i(t)$ are covariates with time-fixed coefficients γ_k. The above model implies that the connection between the past and the future is given by means of the intensity process, which is a function of the current levels of the risk factors and the event history summarized at the current time. Take CVD data as an example, the probability of the occurrence of another MI or cardiovascular death in the future depends on the levels of the risk factors as well as prior CVD history. Although written as functions of times, the components of the covariate vectors $\mathbf{X}_i(t)$ and $\mathbf{W}_i(t)$ can be chosen as time-dependent covariate

or baseline covariate. An example of the time-dependent covariate is the systolic blood pressure (SBP) level that changes over time. The current level of SBP is very likely to be associated with the instantaneous risk of developing a CVD event. Such an association with the CVD risk may be time-varying, which means that for example the impact of the SBP level measured 10 years ago could be different from the impact of the same level of SBP measured today. This is particularly important for our motivating ARIC data. Even if the levels of SBP measured at two different time points are the same, their effects on the CVD risks could be different, considering the long time span of the study and the possible change of physical condition of the participants during the study. Therefore a time-dependent association described by $\eta_k(t)$ in Models (2.1) and (2.2) fits better to this real scenario. Similar applications in CVD research can be found in Zhou et al. (2019), where the effect of longitudinal physical fitness on CVD mortality is assumed to vary with age.

The different regression coefficients for different transitions means that the magnitude of the dependence of CVD risk on the levels of risk factors may vary across recurrence. In some real applications, the number of observed events may be too small for estimation at a later stage (when the number of recurrence is greater than some values). As suggested by Yang et al. (2017) and Cook and Lawless (2007), we can either treat the subsequent events for some large values of k as censored or combine them into one strata and assume that they share the same intensity functions. As for the proportion of observations being censored or combined into one strata, we suggest trying different values and compute the corresponding prediction errors using the formulas in Section 5.3 of the Supplementary Material. The one that minimizes the prediction errors may indicate a best fit of the data.

Let $N_i^E(t-) = \sum_{k=1}^{K} N_{k-1,k}^i(t-)$ denotes the number of recurrent events that subject i has experienced over $[0, t)$. For the covariates $\mathbf{X}_i(t)$ and $\mathbf{W}_i(t)$, we assume that their true trajectories are generated from the following stochastic process

$$X_{ir}(t) = f_{Xr}\left(t, N_i^E(t-)\right)^{\mathrm{T}} \boldsymbol{\beta}_{Xir}, \ r = 1, ..., p_x,$$

(2.3) $$W_{is}(t) = f_{Ws}\left(t, N_i^E(t-)\right)^{\mathrm{T}} \boldsymbol{\beta}_{Wis}, \ s = 1, ..., p_w,$$

where $\boldsymbol{\beta}_{Xir}$ is a c_{Xr}-dimensional random effects, $\boldsymbol{\beta}_{Wis}$ is a c_{Ws}-dimensional random effects and f_{Xr} , f_{Ws} are known generic functions. We assume that $f_{Xr}(\cdot)$ and $f_{Ws}(\cdot)$ satisfy the conditions such that the transition intensity process shall not explode. Details of the conditions can be found in Gjessing et al. (2010). Model (2.3) suggests that the event feedback is manifested through $N_i^E(t-)$ and its impact on the biomarker trajectory varies

among individuals. One example of the past event information is the total number of previous events $\min\{N_i^E(t-), K_0\}$, truncated at a chosen number K_0. A small number of K_0 can prevent the value of the intensity process going to infinity. For example K_0 was chosen to be 5 in an asthma exacerbations study (page 196, Cook and Lawless (2007)). Other forms of past event information include $I\left(N_i^E(t-) > 0\right)$ which implies whether there is prior disease history and $N_i^E(t-)/t$ which is the average rate of past events.

Model (2.3) is flexible because no distribution assumptions are imposed on the random effects. In addition, it is capable of handling various types of covariates. For example, by setting $X_{ir}(t) = \beta_{Xir0} + \beta_{Xir1} t + \cdots + \beta_{Xira} t^a + \beta_{Xir(a+1)} N_i^E(t-)$ we have a time-dependent covariate of which the trajectory is a linear trend of time and the accumulated number of past events. And choosing $X_{ir}(t) = \beta_{Xir0}$ gives us a baseline covariate. Therefore f_{Xr} and f_{Ws} provide various combinations of the time trajectory and the event feedbacks. In practice, the choice of their functional forms is usually not self-evident or may be far from obvious, but may be suggested by biological understanding or biological hypothesis (Fisher and Lin, 1999). When the real application lacks prior knowledge of the generic functions, one may adopt flexible polynomial or fractional polynomial functions of time for a nonparametric approximation. For example by setting $X_{ir}(t) = \beta_{X_i0} + \beta_{X_i1}t_{ij}^2 + \beta_{X_i2}t_{ij}^2 ln(t_{ij}) + \beta_{X_i2}I(N_i^E(t_{ij}-) > 0) + \epsilon_{X_{ij}}$, we have a longitudinal model that is a $FP(2,2)$ fractional polynomial functions of time plus the past event feedbacks. More discussion about this issue can be found in Section 5.

Note that when $K = 0$ (i.e. there are no recurrent events), and when the term $N_i^E(t-)$ vanishes from Model (2.3), the modelling framework degenerates to a standard survival model with time-dependent covariates as discussed in Song and Wang (2008) and Song et al. (2017).

2.2. *The Estimating Equations and Computation Procedure.* In practice, the inference procedure is complicated by the fact that the true values of the covariates $\mathbf{X}_i(t)$ and $\mathbf{W}_i(t)$ are not observed at every event time point. Also the time-varying effects of the covariates cannot be directly estimated as they are essentially infinite dimensional parameters. For the first issue, a popular method is to perform a two-stage analysis in which the values of the time-dependent covariates at all event time points are estimated and incorporated in the intensity functions. Then the bias due to the use of the estimated values are corrected using some estimating equation approach. There has been considerable methodological development including the con-

ditional score approach and the corrected score approach mentioned in Section 1. We develop our estimation procedure based on an extension to the corrected score approach. For the second issue, there has been enormous development on survival analysis with time-varying coefficient (Cai and Sun, 2003; Song and Wang, 2008; Song et al., 2017). The kernel smoothing approach is widely used due to its simplicity and easy interpretation. However for the estimation of the semiparametric parameters, the nonparametric coefficients are usually estimated locally in the first step given the values of the parametric coefficients. And then the estimated nonparametric coefficients are plugged in the model and the parametric coefficients are estimated globally in the next step. These two steps are then iterated until convergence. Such full iterations between the nonparametric step and parametric step are computationally intensive and the expressions of the estimators are usually complicated. Therefore, we derive an one-step estimator of the semiparametric coefficients which are computationally efficient and concise.

Suppose for each component of the covariates, we observe a vector of longitudinal data $\mathbf{X}_{ir} = (X_{i1r}, ..., X_{im_{X_{ir}}r})^{\mathrm{T}}$ instead of the true values of $X_{ir}(t)$ at each time point t for $r = 1, ..., p_x$. Similarly, $\mathbf{W}_{is} = (W_{i1s}, ..., W_{im_{W_{is}}s})^{\mathrm{T}}$ is recorded instead of $W_{is}(t)$ for $s = 1, ..., p_w$. Assuming that the measurement errors are additive, the relationship between the observations and the true covariate process can be described as follows

$$X_{ijr} = X_{ir}(t_{ij}) + \varepsilon_{Xijr}, \, i = 1, ..., n; \, j = 1, ..., m_{Xir}; \, r = 1, ..., p_x,$$
$$(2.4) \quad W_{ijs} = W_{is}(t_{ij}) + \varepsilon_{Wijs}, \, i = 1, ..., n; \, j = 1, ..., m_{Wis}; \, s = 1, ..., p_w,$$

where ε_{Xijr} and ε_{Wijs} are mutually independent. ε_{Xijr} are independent across r with $\varepsilon_{Xijr} \overset{iid}{\sim} N(0, \sigma_{Xr}^2)$ and ε_{Wijs} are independent across s with $\varepsilon_{Wijs} \overset{iid}{\sim} N(0, \sigma_{Ws}^2)$. We assume that the observation time points t_{ij} and the measurement errors are non-informative about the multi-state process. We also assume that the measurement errors are independent of the true covariate process, the state transition times and the censoring time. In our real analysis, the covariates of interests are not all observed with errors. Model (2.4) is also able to handle covariates that are exactly observed. For example, additionally assuming $m_{Xir} = 1$ and $\varepsilon_{Xijr} = 0$ leads to a precisely measured baseline covariate.

Based on model (2.4), the least squares estimates $\hat{\mathbf{X}}_i(t)$, $\hat{\mathbf{W}}_i(t)$ and their conditional variance-covariance matrices $\mathbf{\Sigma}_{Xi}(t)$, $\mathbf{\Sigma}_{Wi}(t)$ can be easily calculated and are presented in the Supplementary Material (Section 2).

Let $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_{2K+1}\}$ and $\eta(u) = \{\eta_1(u), ..., \eta_{2K+1}(u)\}$. Assuming that the values of the time-dependent covariates are all known in Models (2.1) and

(2.2), the log-partial likelihood function $l\left(\boldsymbol{\gamma},\eta(u)\right)$ can be written as in (2.5). Taking the first derivative of the log-partial likelihood leads to the score function. For example, the score functions of the recurrent event parameters $\boldsymbol{\gamma}_k$ given $\eta_k(u)$ can be easily derived as in (2.6). Solving the score functions gives us a valid estimate of the parameters. However when the values of the time-dependent covariates are unknown, simply replacing the true values of the time-dependent covariates by their least squares estimates will lead to a biased estimating equation. For example, the expectation of the numerator in (2.6) is shown in (2.7) where the bias terms involving $\boldsymbol{\Sigma}_{Xj}(u)$ and $\boldsymbol{\Sigma}_{Wj}(u)$ are introduced.

(2.5)

$$
\begin{aligned}
& l\left(\boldsymbol{\gamma},\eta(u)\right) \\
&= \sum_{k=1}^{K}\sum_{i=1}^{n}\Bigg[\int_{0}^{L}\left(\eta_k^{\mathrm{T}}(u)\mathbf{X}_i(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_i(u)\right) \\
&\quad -\log\left(\sum_{j=1}^{n}Y_{k-1}^{j}(u)\exp\left(\eta_k^{\mathrm{T}}(u)\mathbf{X}_j(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_j(u)\right)\right)dN_{k-1,k}^{i}(u)\Bigg] \\
&\quad +\sum_{k=K+1}^{2K+1}\sum_{i=1}^{n}\Bigg[\int_{0}^{L}\left(\eta_k^{\mathrm{T}}(u)\mathbf{X}_i(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_i(u)\right) \\
&\quad -\log\left(\sum_{j=1}^{n}Y_{k-(K+1)}^{j}(u)\exp\left(\eta_k^{\mathrm{T}}(u)\mathbf{X}_j(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_j(u)\right)\right)dN_{k-(K+1),k}^{i}(u)\Bigg].
\end{aligned}
$$

(2.6)
$$
\begin{aligned}
& U(\boldsymbol{\gamma}_k) \\
&= \sum_{i=1}^{n}\int_{0}^{L}\left\{\mathbf{W}_i(u)-\frac{\sum_{j=1}^{n}Y_{k-1}^{j}(u)\mathbf{W}_j(u)\exp\left(\eta_k^{\mathrm{T}}(u)\mathbf{X}_j(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_j(u)\right)}{\sum_{j=1}^{n}Y_{k-1}^{j}(u)\exp\left(\eta_k^{\mathrm{T}}(u)\mathbf{X}_j(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_j(u)\right)}\right\}dN_{k-1,k}^{i}(u), \\
& k=1,...,K.
\end{aligned}
$$

(2.7) $$
\begin{aligned}
& \mathbb{E}\left\{\hat{\mathbf{W}}_j(u)\exp\left(\eta_k^{\mathrm{T}}(u)\hat{\mathbf{X}}_j(u)+\boldsymbol{\gamma}_k^{\mathrm{T}}\hat{\mathbf{W}}_j(u)\right)\big|\mathbf{X}_j(u),\mathbf{W}_j(u)\right\} \\
&= (\mathbf{W}_j(u)+\boldsymbol{\Sigma}_{Wj}(u)\boldsymbol{\gamma}_k)\exp\Big(\eta_k^{\mathrm{T}}(u)\mathbf{X}_j(u)-1/2\eta_k^{\mathrm{T}}(u)\boldsymbol{\Sigma}_{Xj}(u)\eta_k(u) \\
&\quad +\boldsymbol{\gamma}_k^{\mathrm{T}}\mathbf{W}_j(u)-1/2\boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\Sigma}_{Wj}(u)\boldsymbol{\gamma}_k\Big).
\end{aligned}
$$

In addition to bias correction, we also need to approximate the time-varying coefficients $\eta_k(t)$ with local polynomials. For $k = 1, ..., 2K + 1$, let $\eta_k(u) \approx \eta_k(t) + \eta_k'(t)(u - t)$ with $\eta'(t) = \partial\eta(t)/\partial t$. Denote $\mathbf{b}_k = \left(\eta_k^{\mathrm{T}}(t), \, \eta_k'^{\mathrm{T}}(t)\right)^{\mathrm{T}}$. Let $\boldsymbol{\gamma}_k(t)$ denote the local estimators of the time-independent coefficients in the nonparametric step. Note that they are not the final estimates. To keep our notations simple, we omit the argument t in $\boldsymbol{\gamma}_k(t)$ when introducing the estimating equations in the nonparametric step. Let $\boldsymbol{\alpha}_k = \left(\mathbf{b}_k^{\mathrm{T}}, \boldsymbol{\gamma}_k^{\mathrm{T}}\right)^{\mathrm{T}}$. Define $\mathbf{A} \otimes \mathbf{B}$ as the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$ throughout the paper. Let blkdiag$(\mathbf{A}, \mathbf{B})$ denote a block diagonal matrix with the main diagonal blocks $\mathbf{A}$ and $\mathbf{B}$. Let h be the bandwidth, $\mathbf{H} = \mathrm{diag}\left(\mathbf{I}_{p_x}, h\mathbf{I}_{p_x}\right)$, $u_t = (1, u-t)^{\mathrm{T}}$, $\mathbf{X}_i(u,t) = u_t \otimes \mathbf{X}_i(u)$, $\boldsymbol{\Sigma}_{Xi}(u,t) = u_t u_t^{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{Xi}(u)$, $\mathbf{Z}_i(u,t) = \left(\mathbf{X}_i^{\mathrm{T}}(u,t), \mathbf{W}_i^{\mathrm{T}}(u)\right)^{\mathrm{T}}$, $\boldsymbol{\Sigma}_{Zi}(u,t) = \mathrm{blkdiag}(\boldsymbol{\Sigma}_{Xi}(u,t), \boldsymbol{\Sigma}_{Wi}(u))$ and $K_h(u - t) = K(u-t)/h$, where $K(\cdot)$ is the kernel function. Let $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^{\mathrm{T}}$ for a general vector a. To preserve predictability which is essential for the derivation of the theoretical results, we only use the information before time t to estimate $\mathbf{X}_i(t)$ and $\mathbf{W}_i(t)$. Let $m_{Xir}(t)$ be the number of observations before time t for the rth component in $\mathbf{X}_i(t)$ and $m_{Wis}(t)$ be the number of observations before time t for the sth component in $\mathbf{W}_i(t)$. Let $I_i(t) = I(m_{Xir}(t) \geq c_{Xr}, m_{Wis}(t) \geq c_{Ws} : r = 1, ..., p_x; \, s = 1, ..., p_w)$ be an indicator function which takes value 1 if the number of longitudinal observations from individual i before time t is sufficient for the estimation of $\boldsymbol{\beta}_{Xir}$ and $\boldsymbol{\beta}_{Wis}$. Denote $Y_i(t) = I(C_i \geq t)$, where C_i is the censoring time of the ith individual and $Y_i(t) = 1$ if a subject is still under observation at time t. Together with the state transition at risk process $Y_j^i(t)$ in Section 2, we can define the overall at risk process $\bar{Y}_j^i(t) = I_i(t)Y_i(t)Y_j^i(t)$ for $j = 0, ..., K$.

The whole estimation procedure includes two steps. In the first step we estimate both the time-varying coefficients and the constant parameters in a local estimating equation. The estimators of the constant parameters are not efficient because the estimating equation only use a local subset of the data. Therefore in the next step, we construct a global estimating equation of the constant parameters with the time-varying coefficients substituted by their local estimates. The estimating equation is global in the sense that it utilises the whole data set. Unlike the profile method, the estimates of the time-varying coefficients being substituted in the second step are treated as constants instead of functions of the constant parameters. With the assumption about the bandwidth and using the order of convergence of the local estimator, it can be shown that the global estimator of the constant parameters achieves the optimal convergence rates. Details of the algorithm and its application in other models can be found in Carroll et al. (1997);

Cao and Yao (2012); Huang and Yao (2012).

In what follows, we use the subscript NP to represent the nonparametric estimate and use the subscript P to refer to the parametric estimate. Functions with subscript k ranging from 1 to K indicate quantities involved in the estimation of recurrent event parameters and those with subscripts k ranging from $K + 1$ to $2K + 1$ correspond to the terminal event parameters. To remove the bias according to the result in (2.7), we define the following functions of which the expectations equal to the numerator and denominator in a standard score function of the log-partial likelihood with true covariate values. For $l = 0, 1, 2$, let

$$
\begin{aligned}
S_{NP,k}^{(l)}(\boldsymbol{\alpha}_k, u, t) =& n^{-1} \sum_{j=1}^{n} \bar{Y}_{k-1}^{j}(u) \left(\hat{\mathbf{Z}}_j(u, t) - \boldsymbol{\Sigma}_{Zj}(u, t)\boldsymbol{\alpha}_k \right)^{\otimes l} \\
& \times \exp\left(\boldsymbol{\alpha}_k^{\mathrm{T}} \hat{\mathbf{Z}}_j(u, t) - \frac{1}{2}\boldsymbol{\alpha}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{Zj}(u, t)\boldsymbol{\alpha}_k \right), k = 1, ..., K, \\
S_{NP,k}^{(l)}(\boldsymbol{\alpha}_k, u, t) =& n^{-1} \sum_{j=1}^{n} \bar{Y}_{k-(K+1)}^{j}(u) \left(\hat{\mathbf{Z}}_j(u, t) - \boldsymbol{\Sigma}_{Zj}(u, t)\boldsymbol{\alpha}_k \right)^{\otimes l} \\
& \times \exp\left(\boldsymbol{\alpha}_k^{\mathrm{T}} \hat{\mathbf{Z}}_j(u, t) - \frac{1}{2}\boldsymbol{\alpha}_k^{\mathrm{T}} \boldsymbol{\Sigma}_{Zj}(u, t)\boldsymbol{\alpha}_k \right), k = K + 1, ..., 2K + 1.
\end{aligned}
$$

Given the variance of the measurement errors, the local estimating equations of $\boldsymbol{\alpha}_k$ can be written as

$$
\begin{aligned}
U_{NP,k}(\boldsymbol{\alpha}_k) =& (n\mathbf{H})^{-1} \sum_{i=1}^{n} \int_0^L K_h(u - t)\bar{Y}_{k-1}^{i}(u) \left\{ \hat{\mathbf{Z}}_i(u, t) - \frac{S_{NP,k}^{(1)}(\boldsymbol{\alpha}_k, u, t)}{S_{NP,k}^{(0)}(\boldsymbol{\alpha}_k, u, t)} \right\} \\
& dN_{k-1,k}^{i}(u), \ k = 1, ..., K, \\
U_{NP,k}(\boldsymbol{\alpha}_k) =& (n\mathbf{H})^{-1} \sum_{i=1}^{n} \int_0^L K_h(u - t)\bar{Y}_{k-(K+1)}^{i}(u) \left\{ \hat{\mathbf{Z}}_i(u, t) - \frac{S_{NP,k}^{(1)}(\boldsymbol{\alpha}_k, u, t)}{S_{NP,k}^{(0)}(\boldsymbol{\alpha}_k, u, t)} \right\}
\end{aligned}
$$

(2.8) $dN_{k-(K+1),k}^{i}(u), \ k = K + 1, ..., 2K + 1.$

In practical applications, we can substitute σ_{Xr}^2 and σ_{Ws}^2 with their consistent estimates, e.g. the moment estimator based on all longitudinal data and the theoretical properties of our proposed estimating equations remain valid. Replacing the unknown functions $\eta_k(u)$ with their local estimates $\hat{\eta}_k(u) \equiv \hat{\mathbf{b}}_k$, the constant parameters $\boldsymbol{\gamma}_k$ can be estimated in the following

parametric step. For $l = 0, 1, 2$, let

$$
S_{W,P,k}^{(l)}(\boldsymbol{\gamma}_k, \eta_k, u) = n^{-1} \sum_{j=1}^{n} \bar{Y}_{k-1}^{j}(u) \left(\hat{\mathbf{W}}_j(u) - \boldsymbol{\Sigma}_{Wj}(u)\boldsymbol{\gamma}_k \right)^{\otimes l} \exp\left(\eta_k^{\mathrm{T}}(u)\hat{\mathbf{X}}_j(u) \right.
$$

$$
\left. - \frac{1}{2}\eta_k^{\mathrm{T}}(u)\boldsymbol{\Sigma}_{Xj}(u)\eta_k(u) + \boldsymbol{\gamma}_k^{\mathrm{T}}\hat{\mathbf{W}}_j(u) - \frac{1}{2}\boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\Sigma}_{Wj}(u)\boldsymbol{\gamma}_k \right),
$$

$$
k = 1, ..., K,
$$

$$
S_{W,P,k}^{(l)}(\boldsymbol{\gamma}_k, \eta_k, u) = n^{-1} \sum_{j=1}^{n} \bar{Y}_{k-(K+1)}^{j}(u) \left(\hat{\mathbf{W}}_j(u) - \boldsymbol{\Sigma}_{Wj}(u)\boldsymbol{\gamma}_k \right)^{\otimes l} \exp\left(\eta_k^{\mathrm{T}}(u)\hat{\mathbf{X}}_j(u) \right.
$$

$$
\left. - \frac{1}{2}\eta_k^{\mathrm{T}}(u)\boldsymbol{\Sigma}_{Xj}(u)\eta_k(u) + \boldsymbol{\gamma}_k^{\mathrm{T}}\hat{\mathbf{W}}_j(u) - \frac{1}{2}\boldsymbol{\gamma}_k^{\mathrm{T}}\boldsymbol{\Sigma}_{Wj}(u)\boldsymbol{\gamma}_k \right),
$$

$$
k = K + 1, ..., 2K + 1.
$$

The global estimating equations of $\boldsymbol{\gamma}_k$ can be written as

$$
U_{P,k}(\boldsymbol{\gamma}_k, \hat{\eta}_k) = n^{-1} \sum_{i=1}^{n} \int_0^L \bar{Y}_{k-1}^{i}(u) \left\{ \hat{\mathbf{W}}_i(u) - \frac{S_{W,P,k}^{(1)}(\boldsymbol{\gamma}_k, \hat{\eta}_k, u)}{S_{W,P,k}^{(0)}(\boldsymbol{\gamma}_k, \hat{\eta}_k, u)} \right\} dN_{k-1,k}^{i}(u),
$$

$$
k = 1, ..., K,
$$

$$
U_{P,k}(\boldsymbol{\gamma}_k, \hat{\eta}_k) = n^{-1} \sum_{i=1}^{n} \int_0^L \bar{Y}_{k-(K+1)}^{i}(u) \left\{ \hat{\mathbf{W}}_i(u) - \frac{S_{W,P,k}^{(1)}(\boldsymbol{\gamma}_k, \hat{\eta}_k, u)}{S_{W,P,k}^{(0)}(\boldsymbol{\gamma}_k, \hat{\eta}_k, u)} \right\} dN_{k-(K+1),k}^{i}(u),
$$

(2.9)

$$
k = K + 1, ..., 2K + 1.
$$

The global estimators attain the optimal convergence rates as shown in the technical proof in the Supplementary Material (Section 6).

3. Simulation. We conducted extensive simulations to assess the performance of our proposed model and estimation method. Four scenarios were considered. Scenario I considered an example with moderate censoring rate and a reasonable number of longitudinal observations measured at irregular intervals. Different distributions of the random coefficients in the longitudinal model were considered. The influence of sample size was also evaluated. Scenario II considered a case where there were at most 5 longitudinal observations for each time-dependent covariate. The first four measurements were made at regular time intervals and the last measurement

was taking at a much longer interval. This setting mimicked the cohort surveillance examination cycles in the ARIC study. Scenario III aimed to illustrate the performance of our proposed estimator when there was actually no dynamic feedback in the system that generated the event data and longitudinal data. Scenario IV investigated the performance of our proposed estimator when the generic functions of the longitudinal data process were incorrectly specified. The multi-state models that generated the event times in all scenarios were $\lambda_{i,k}(t) = \lambda_0^E(t) \exp\left(\eta_E(t)X_i(t) + \gamma_{E1}W_{i1} + \gamma_{E2}W_{i2}(t)\right)$ for $k = 1, ..., K$ and $\lambda_{i,k}(t) = \lambda_0^D(t) \exp\left(\eta_D(t)X_i(t) + \gamma_{D1}W_{i1} + \gamma_{D2}W_{i2}(t)\right)$ for $k = K + 1, ..., 2K + 1$. Here we assume that the baseline intensity functions and the regression coefficients are the same across transitions to the intermediate states (indexed by E), and the baseline intensities and regression parameters are also shared across transitions to the terminal states (indexed by D). This is to mimic the models in the real analysis. The estimating equations can be similarly derived with some straightforward algebra and are presented in the Supplementary Material (Section 3). Multi-state event data were simulated using nested series of competing risks experiments and details for generating such data can be found in Beyersmann et al. (2011). The baseline intensity functions were chosen as $\lambda_0^E(t) = 2.5$, $\lambda_0^D(t) = 0.1 + t^{1/2}$. The maximum follow-up time was 1. Censoring times were generated from a uniform distribution $U[0, 3]$ and truncated at 1, leading to a censoring rate around 45%. The true values of the coefficients were $\eta_E(t) = -sin(\pi t), \eta_D(t) = \log(0.6t + 0.4)$, $\gamma_{E1} = 0.6$, $\gamma_{E2} = 0.4$, $\gamma_{D1} = 0.3$, $\gamma_{D2} = 0.5$. The baseline covariate W_{i1} was sampled from a Bernoulli distribution with probability 0.5. The observations of the time-dependent covariates were generated from the longitudinal models $X_{ij} = \beta_{Xi0} + \beta_{Xi1}t_{ij}^{1/2} + \beta_{Xi2}\log\left(N_i^E(t_{ij}-) + 1\right) + \varepsilon_{Xij}$ and $W_{ij2} = \beta_{Wi0} + \beta_{Wi1}t_{ij} + \beta_{Wi2}\min\left(N_i^E(t_{ij}-), 5\right) + \varepsilon_{Wij}$. We took log-transformation and truncation of the past event feedbacks to avoid explosions as discussed in Section 2. In Scenario III where there was no past event feedback in the true longitudinal process, we set $\beta_{Wi2} = \beta_{Xi2} = 0$ for all i. The variances of the measurement errors ε_{Wij} and ε_{Xij} were set as 0.1.

In each scenario, the sample size was chosen as 500 and simulation results were based on 500 Monte Carlo replications. The Epanechnikov kernel function $K(t) = 0.75(1 - t^2)I(|t| \leq 1)$ was used throughout the simulations. The unknown functions and coefficients in all scenarios were estimated in four ways: (i) Our proposed estimator (DC) where D is short for dynamic feedbacks and C for bias correction; (ii) the naive estimator that ignores the dynamic feedbacks (NaivD); (iii) the naive estimator us-

ing least squares estimate of the random coefficients without correcting bias (NaivC) and (iv) the ideal estimator (I) that assumes the random coefficients $\boldsymbol{\beta}_{Xi}$ and $\boldsymbol{\beta}_{Wi}$ to be known at each time point. All of the aforementioned estimators were computed with the one-step algorithm. The NaivD estimates were calculated assuming longitudinal models without event feedbacks $X_{ij} = \beta_{Xi0} + \beta_{Xi1} t_{ij}^{1/2} + \varepsilon_{Xij}$ and $W_{ij2} = \beta_{Wi0} + \beta_{Wi1} t_{ij} + \varepsilon_{Wij}$. 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 standard deviation (SD), the average of the estimated standard error (SE), and the 95% empirical coverage probability (CP). The performance of the local estimator was evaluated at 50 equally spaced grid points, using bandwidth $h = 0.05$ and $h = 0.1$. We chose such small values of the bandwidths because the theoretical results require the nonparametric estimates to be undersmoothed. Details can be found in the Supplementary Material (Condition B, Section 6.2). The average of the estimates, the 95 % pointwise confidence bands and the pointwise coverage probabilities were calculated at each grid point. For better presentation the span of the vertical axis in all plots of the pointwise coverage probabilities of $\eta_D(t)$ is restricted to $[0.7, 1]$ because the coverage rates by different estimators vary only over this small range. In addition, we mark the bias and coverage probabilities of the proposed DC estimator and NaivD estimator in bold in all tables to highlight the difference of the performance. To save space, the two simulation studies investigating the effect of sample size and random effect distribution in Scenario I, and all the simulation results in Scenario III and Scenario IV are given in the Supplementary Material (Section 4).

3.1. *Scenario I, Setting 1.* The longitudinal measurement times were $t_{ij} = (0,\ 0.05, 0.09, 0.15, 0.2, 0.3, 0.4, 0.5, 0.9)$ and observations were made until the occurrence of terminal event or censoring. We first considered $\boldsymbol{\beta}_{Xi}$ generated from a normal distribution with mean $(1, -0.2, 0.3)$, variance $(0.5, 0.01, 0.05)$ and correlations $corr(\beta_{Xi0}, \beta_{Xi1}) = 0.2$, $corr(\beta_{Xi0}, \beta_{Xi2}) = 0.8$, $corr(\beta_{Xi1}, \beta_{Xi2}) = 0.6$. $\boldsymbol{\beta}_{Wi}$ were sampled from a normal distribution with mean $(1, -0.5, 0.2)$, variance $(0.4, 0.01, 0.05)$ and correlations $corr(\beta_{Wi0}, \beta_{Wi1}) = 0.2$, $corr(\beta_{Wi0}, \beta_{Wi2}) = 0.8$, $corr(\beta_{Wi1}, \beta_{Wi2}) = 0.6$. About 75% subjects had at least one recurrent event, among which the average number of recurrent events was 2.5 per subject. Table 1 gives the estimates of the time-independent parameters.

From Table 1, we can see that our proposed estimation procedure performs well. The DC estimators of the time-independent parameters $\gamma_{E,1}$, $\gamma_{E,2}$,

$\gamma_{D,1}$ and $\gamma_{D,2}$ are unbiased and the 95% empirical coverage probabilities are close to the nominal value. Estimates based on different bandwidth values are similar. For $\gamma_{E,2}$ and $\gamma_{D,2}$ which are the coefficients of the longitudinal covariates, the naive estimators are substantially biased and their coverage probabilities are below the nominal value. The naive estimators seem to yield unbiased results for $\gamma_{E,1}$ and $\gamma_{D,1}$, which are the coefficients of the baseline covariates. This is expected because the past event feedback impacts the time-dependent covariates more than the baseline covariates and also the baseline covariates are assumed to be exactly observed. Thus using naive methods have limited impact on the estimation of $\gamma_{E,1}$ and $\gamma_{D,1}$.

Figures 2 - 3 present the true and fitted curves of the time-varying coefficients and it can be seen that the fitted curves based on our proposed method show a close agreement with the true functions. The proposed DC estimators outperform the naive estimators in terms of the 95% pointwise coverage probabilities.

TABLE 1
Simulation results in the case of normal random effects with 45% censoring rate and sample size n=500.

	$\gamma_{E,1}^{DC}$	$\gamma_{E,2}^{DC}$	$\gamma_{E,1}^{NaivC}$	$\gamma_{E,2}^{NaivC}$	$\gamma_{E,1}^{NaivD}$	$\gamma_{E,2}^{NaivD}$	$\gamma_{E,1}^{I}$	$\gamma_{E,2}^{I}$
				h=0.05				
Bias	0.0077	**0.0065**	0.0057	-0.0618	0.0038	**-0.0212**	0.0061	-0.0021
SD	0.0788	0.0380	0.0785	0.0333	0.0746	0.0332	0.0674	0.0307
SE	0.0764	0.0369	0.0762	0.0318	0.0724	0.0324	0.0686	0.0310
CP	0.9420	**0.9300**	0.9440	0.5180	0.9360	**0.8760**	0.9540	0.9640
				h=0.1				
Bias	0.0073	**0.0062**	0.0056	-0.0617	0.0036	**-0.0214**	0.0060	-0.0021
SD	0.0786	0.0379	0.0785	0.0333	0.0745	0.0331	0.0674	0.0307
SE	0.0762	0.0367	0.0762	0.0318	0.0723	0.0323	0.0686	0.0310
CP	0.9420	**0.9320**	0.9420	0.5160	0.9360	**0.8740**	0.9520	0.9640
	$\gamma_{D,1}^{DC}$	$\gamma_{D,2}^{DC}$	$\gamma_{D,1}^{NaivC}$	$\gamma_{D,2}^{NaivC}$	$\gamma_{D,1}^{NaivD}$	$\gamma_{D,2}^{NaivD}$	$\gamma_{D,1}^{I}$	$\gamma_{D,2}^{I}$
				h=0.05				
Bias	0.0005	**0.0119**	0.0030	-0.0600	-0.0055	**-0.0395**	0.0010	0.0023
SD	0.1258	0.0585	0.1226	0.0486	0.1240	0.0566	0.1223	0.0528
SE	0.1266	0.0578	0.1226	0.0479	0.1251	0.0530	0.1227	0.0518
CP	0.9560	**0.9360**	0.9560	0.7580	0.9540	**0.8500**	0.9520	0.9480
				h=0.1				
Bias	0.0010	**0.0120**	0.0038	-0.0595	-0.0050	**-0.0390**	0.0017	0.0027
SD	0.1255	0.0584	0.1225	0.0486	0.1238	0.0566	0.1222	0.0527
SE	0.1263	0.0576	0.1225	0.0479	0.1248	0.0529	0.1226	0.0518
CP	0.9560	**0.9340**	0.9560	0.7640	0.9540	**0.8540**	0.9520	0.9480

FIG 2. *Estimation of $\eta_E(t)$ in the case of normal random effects with 45% censoring rate and sample size n=500.*

Fɪɢ 3. *Estimation of $\eta_D(t)$ in the case of normal random effects with 45% censoring rate and sample size n=500.*

3.2. *Scenario II.* In this scenario, we evaluate the performance of our proposed estimating procedure when the longitudinal observations of the time-dependent covariate are sparse, which is the case of the ARIC study. The longitudinal measurements were taken at $t = (0, 0.125, 0.25, 0.375, 0.875)$. The range of grid points in this scenario was set as $[0.25, 1]$ considering the fact that all subjects in the simulated data set had less than 3 longitudinal observations before $t = 0.25$. Specifically, the smoothing windows of any grid point $t < 0.25$ contained very few data that was feasible for estimation and hence the nonparametric estimate before $t = 0.25$ might not be reliable. The selected range of the grid points did not influence the estimation of the constant parameters because the local estimates at event times earlier than 0.25 were not used in the global estimating equation anyway. The rest of the simulation setup was the same as in Scenario I.

As shown in Table 2, we can see that the proposed DC estimators outperform the other two naive ones in terms of the 95% coverage probabilities. Figures 4 - 5 also show that the proposed method achieves satisfactory performance regarding the pointwise 95% coverage probability. Although the absolute bias of the estimators of the time-independent parameters and time-varying effects using the proposed method are slightly larger than those using

the naive method which ignores the past event feedbacks. This should be evaluated together with the scale of the standard errors of both estimators. It can be seen from Table 2 that the standard errors of the proposed DC estimators are larger than the NaiveD estimators. From Figures 4 - 5, we can also find that the confidence bands of the proposed DC estimators are wider than those of the NaiveD estimators. Therefore the performance measures should be considered jointly and we can see that the proposed estimators actually achieve better coverage probabilities than the naive estimators.

The additional simulation studies in the Supplementary Material evaluate the impact of the random effect distribution, sample size, existence of event feedback and the misspecification of the generic functions. It is found that our estimating procedure provided similar results under a different random effect distribution setting. The performance of the proposed method also improved with a larger sample size. When there is no event feedback in the true model, our approach gives results comparable to those based on methods using a perfect model for the data. In addition, when the functional form of the longitudinal process is misspecified, the performance of the proposed estimators is still satisfactory and outperform the naive estimators. Taken together, our method is reliable and provides better results than other methods in many different scenarios.

FIG 4. *Estimation of $\eta_E(t)$ in the case of sparse longitudinal observations.*

TABLE 2
Simulation results in the case of sparse longitudinal observations.

	$\gamma_{E,1}^{DC}$	$\gamma_{E,2}^{DC}$	$\gamma_{E,1}^{NaivC}$	$\gamma_{E,2}^{NaivC}$	$\gamma_{E,1}^{NaivD}$	$\gamma_{E,2}^{NaivD}$	$\gamma_{E,1}^{I}$	$\gamma_{E,2}^{I}$
				h=0.05				
Bias	0.0070	**0.0323**	-0.0369	-0.1001	-0.0099	**-0.0239**	-0.0028	-0.0007
SD	0.1185	0.0633	0.1017	0.0416	0.0912	0.0409	0.0700	0.0324
SE	0.1196	0.0618	0.1046	0.0407	0.0887	0.0383	0.0683	0.0308
CP	0.9512	**0.9146**	0.9248	0.3089	0.9411	**0.8821**	0.9431	0.9451
				h=0.1				
Bias	0.0031	**0.0297**	-0.0372	-0.1001	-0.0109	**-0.0241**	-0.0033	-0.0006
SD	0.1148	0.0610	0.1016	0.0416	0.0904	0.0404	0.0697	0.0320
SE	0.1150	0.0599	0.1048	0.0407	0.0883	0.0380	0.0683	0.0308
CP	0.9451	**0.9207**	0.9228	0.3089	0.9390	**0.8841**	0.9431	0.9512
	$\gamma_{D,1}^{DC}$	$\gamma_{D,2}^{DC}$	$\gamma_{D,1}^{NaivC}$	$\gamma_{D,2}^{NaivC}$	$\gamma_{D,1}^{NaivD}$	$\gamma_{D,2}^{NaivD}$	$\gamma_{D,1}^{I}$	$\gamma_{D,2}^{I}$
				h=0.05				
Bias	-0.0023	**0.0264**	-0.0071	-0.1223	-0.0088	**-0.0503**	0.0008	-0.0026
SD	0.1453	0.0748	0.1325	0.0493	0.1328	0.0567	0.1268	0.0509
SE	0.1479	0.0754	0.1362	0.0505	0.1317	0.0556	0.1228	0.0516
CP	0.9573	**0.9451**	0.9573	0.3150	0.9512	**0.8130**	0.9431	0.9533
				h=0.1				
Bias	-0.0022	**0.0256**	-0.0063	-0.1219	-0.0086	**-0.0498**	0.0015	-0.0019
SD	0.1443	0.0740	0.1322	0.0492	0.1322	0.0565	0.1265	0.0507
SE	0.1471	0.0748	0.1361	0.0505	0.1313	0.0553	0.1227	0.0516
CP	0.9553	**0.9492**	0.9593	0.3191	0.9512	**0.8232**	0.9431	0.9573

FIG 5. *Estimation of $\eta_D(t)$ in the case of sparse longitudinal observations.*

4. Real Analysis.   In this section, we applied our proposed model and estimation procedure to the data set from the ARIC study. We investigate the effect of SBP level and baseline risk factor on time to recurrent MI and death in White male participants living in Washington County and Suburban Minneapolis, who had no prior coronary heart disease history. We chose this subgroup of participants because the risk factors for CVD differ considerably by race and gender (Kim et al., 2012) and also we excluded those participants with previous coronary heart disease events because the form of past event feedback adopted in this real data analysis was an indicator function of whether there were previous MI events. We restricted our attention to the event times and follow-up data until 2011 because there were few events after this time point. Participants with missing values in baseline covariates were also excluded. This sample included 2932 subjects with 876 fatal events and 482 recurrent events. The number of participants who had at least one recurrent event is 379. Among these participants, the maximum number of recurrent events observed per subject is 6 and the average number is around 1.27. Around 30% participants experienced the fatal event.

We fitted Models (2.1-2.2) to the time-to-event data, and additionally assume that all transitions related to the recurrent events are modelled with the same baseline intensity functions and parameters, and all terminal events share the same baseline intensities and regression coefficients. This is because the average number of MI occurrences was small among those subjects who had recurrent events. Therefore all the recurrent events are combined into one strata, and so are the terminal events. Nevertheless, the progression feature of the disease can still be reflected in the event feedbacks. The time scale we used is the time since the participant entered the study. We considered a longitudinal covariate SBP (in 50mmHg) and the following baseline covariates: age (in years); baseline level of high-density lipoprotein cholesterol (HDLC, in mmol/L); indicators for diabetes (1 for fasting glucose $\geq$ 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. To determine the time-varying/independent structure in the multi-state models, we ran a preliminary analysis in which the coefficients of all covariates were treated as time-dependent and were estimated non-parametrically. The results suggested that the coefficients of SBP, HDL-cholesterol and hypertension medication did not change much over time. In addition, the effect of SBP on time to recurrent MI and the

effect of HDL-cholesterol on time-to-death were not significant over most of the follow-up time. Hence the coefficients of these three covariates were chosen to be time-independent and the others were chosen to be time-varying effects. As for the longitudinal covariate, we fitted the following model to the SBP measurements

$$(4.1) \qquad W_{ij} = \beta_{Xi0} + \beta_{Xi1}t_{ij} + \beta_{Xi2}I\left(N_i^E(t_{ij}-) > 0\right) + \varepsilon_{Xij}.$$

This model implies that the intercept of the SBP trajectory changes after the first MI occurrence. The form of the longitudinal model was chosen according to the spaghetti plot of SBP shown in the Supplementary Material (Section 5). It can be seen that the levels of SBP seem to change with the occurrences of MI. In addition, there are substantial heterogeneities in the patterns of the individual trajectories, suggesting the use of random effects to characterize the variation across subjects.

We used adaptive bandwidth to estimate the time-varying coefficients because the event times are non-uniformly distributed. The idea is that we use a fixed quantile of total number of event times to estimate the local parameters at every grid point and the size of the local window varies with the density of the observed event time points. Details of bandwidth selection can be found in the Supplementary Material (Section 5). The time-varying coefficients were estimated at 50 equally spaced grid points. For comparison, we estimate the coefficients under two models. The estimation results using our proposed models assuming past event feedback in the longitudinal covariate are reported under the full model and those from the models assuming no event feedback are reported under the reduced model.

The estimated constant effects of SBP, HDL-cholesterol (HDL-C) and hypertension medication (HypMed) are displayed in Table 3. It can be seen that the sign of the estimated effects are the same comparing the results based on the full model and the reduced model. Being taken hypertension lowering medication was associated with an increase in the hazard of death and this effect was significant at the 95% confidence level. On the other hand, HDL-cholesterol had protective effect against recurrent MI, which means that an increase in the level of HDL-cholesterol would reduce the risk of MI recurrences. Current level of SBP was not significantly associated with the risk of death and recurrent MI. The estimated associations between the current level of SBP and the risk of recurrent MI and death are negative, which is counterintuitive to what we would expect: as higher level of blood pressure is usually associated with greater risk of cardiovascular events. Under our modelling framework, the current level of SBP is essentially a sum of the baseline SBP (β_{Xi0} in (4.1)), the change of SBP levels

since baseline examination ($\beta_{Xi1}t_{ij}$ in (4.1)) and the MI occurrence history ($\beta_{Xi2}I\left(N_i^E(t_{ij}-)>0\right)$ in (4.1)). The random slope β_{Xi1} represents the rate of change of SBP. Therefore the negative association between the current level of SBP and the risk of CVD events may be explained by a negative effect of the baseline SBP on the CVD risk, or a negative association between the rate of change of SBP and the CVD risk, or a combination of the both. A previous statistical analysis of the ARIC data (Barrett et al., 2019) has similar findings. Using two-stage method and joint modelling approach on the full ARIC data set, they found that the rate of change of SBP is negatively associated with the risk of incident CVD event. That is to say, a positive slope of the SBP level leads to a lower risk of CVD event.

Figures 6 - 7 present the estimates of the time-varying effects of age, diabetes and smoking. All of the three risk factors were significant for both types of events over most of the time span. Age and smoking were positively associated with the risk of MI occurrences and these effects increased along time. It can be seen from Figure 6 that the estimated effects of age and smoking at early time points under the full model are different from those obtained from the reduced models. This discrepancy might be explained by the different event time data involved in the estimating equations under full model and reduced model, especially for equations evaluated at the first few grid points. As shown in Figure 7, age, diabetes and smoking were all positively associated with time-to-death. The estimated coefficients of these factors obtained from the full model and reduced model are similar.

Comparing the results obtained from the full model and the reduced model, it is found that the estimated effects of the longitudinal SBP levels are quite different. The standard errors of our proposed method are larger. This is related to the model complexity as our proposed model takes into account the past event feedbacks and the reduced model assumes a simple linear trajectory of time for the longitudinal data. The estimation results of other baseline variables such as HDLC, hypertension medication, age, diabetes and smoking from the full model and the reduced model are more similar. This is expected because the past event feedback impacts the longitudinal covariates more than the baseline covariates. Though there is medical evidence supporting the existence of past event feedback in SBP (Section 1) and the preliminary longitudinal data analysis also shows that the fixed effects of the event feedback are significant (Supplementary Material Section 5), model selection methods that compare different longitudinal models would be useful to provide further statistical evidence. This remains an open research question.

TABLE 3
Estimated parameters and standard errors for the ARIC data.

	Full model			Reduced model		
	SBP	HDL-C	HypMed	SBP	HDL-C	HypMed
Recurrent MI						
Est	-0.5248	-1.1622	0.2607	-0.0859	-1.1254	0.1723
SE	1.7465	0.2377	0.3604	0.2130	0.2197	0.1499
Death						
Est	-0.5453	0.1172	0.3829	-0.1090	0.1202	0.3004
SE	2.8472	0.1191	0.1162	0.1207	0.1140	0.0807
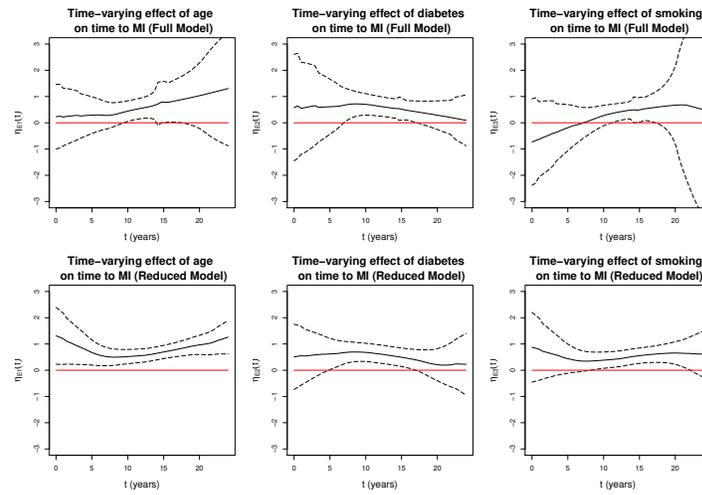
FIG 6. *Estimation of time-varying effects of age (10 years), diabetes and smoking on time to recurrent MI for the ARIC data.*
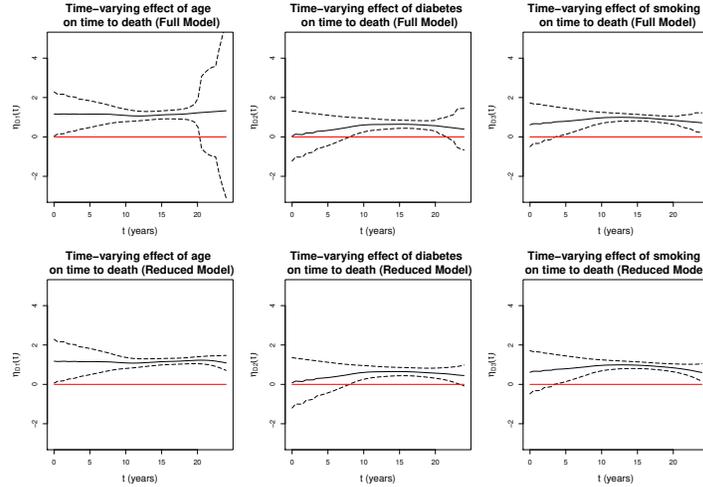
FIG 7. *Estimation of time-varying effects of age (10 years), diabetes and smoking on time to death for the ARIC data.*

5. Discussion. This paper presented a class of multi-state models to estimate the time-varying/independent associations between covariates and time-to-events in the presence of past event feedback. If covariates are measured exactly at each event time point, there shall be no extra effort to address the past event impact as they are directly reflected in the precise values of the measurements. However in practice, some covariates are measured at intermittent time points and are subject to measurement errors, which makes it difficult to take into account the past event impact. To tackle this problem, we proposed to model the trajectory of the time-dependent covariate with past event history.

The proposed estimation method relies on the assumption that the generic functions of the trajectories of the time-dependent covariates are known. While in practice some applications may lack the theories or background knowledge to determine the true forms of model. One option is to consider using flexible polynomial or fractional polynomial functions of time as a nonparametric approximation. We have shown via a simulation study that a linear mixed effects model of time and past event feedbacks provides a reasonable approximation to the true nonlinear trajectories and the performance of the proposed estimators are still satisfactory. The results are presented in Section 4.4 in the Supplementary Material. The theoretical properties of such nonparametric approximation remains unknown and would be an interesting topic.

The proposed models also assume that the set of covariates with time-

varying effects can be predetermined. However in practical applications, it is likely that data of a large number of variables are collected but which covariates have null, constant or time-varying effect remains unknown. Though preliminary analysis treating all covariates nonparametrically with graphical check can help identify some patterns of the coefficients, a more formal method could be considered. For example, the model structure selection approach in Yan and Huang (2012) or Xiao, Lu and Zhang (2016) might be employed to develop a simultaneous time-varying/independent effects selection and estimation method.

Our estimation procedure only uses longitudinal measurements before a given event time point. When the repeated measurements are sparse or the observation times are very irregular, the performance of the proposed estimators could be influenced as demonstrated in simulations. Therefore, a more efficient method that utilizes all the longitudinal information from the same subject is required. Huang, Hwang and Chen (2016) has proposed to use Monte Carlo error augmentation procedure to utilize the entire longitudinal information and provided the theoretical properties of the estimators. Cao et al. (2015) developed a kernel weighted score function to tackle the sparsity problem of the longitudinal covariates in the proportional hazards model. These methods provide insights for solutions when the the longitudinal data are sparse. Methods that work with multi-state event and dynamic longitudinal covariates will be considered as future topics.

References.

AALEN, O., BORGAN, O. and GJESSING, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

AALEN, O. O., FOSEN, J., WEEDON-FEKJÆR, H., BORGAN, Ø. and HUSEBYE, E. (2004). Dynamic analysis of multivariate failure time data. *Biometrics* **60** 764–773.

BARRETT, J. K., HUILLE, R., PARKER, R., YANO, Y. and 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.

BEDAIR, K., HONG, Y., LI, J. and AL-KHALIDI, H. R. (2016). Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial. *Computational Statistics & Data Analysis* **101** 161–173.

BEYERSMANN, J. et al. (2011). *Competing risks and multistate models with R*. Springer Science & Business Media.

BORGAN, Ø., FIACCONE, R. L., HENDERSON, R. and BARRETO, M. L. (2007). Dynamic analysis of recurrent event data with missing observations, with application to infant diarrhoea in Brazil. *Scandinavian Journal of Statistics* **34** 53–69.

CAI, Z. and SUN, Y. (2003). Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models. *Scandinavian Journal of Statistics* **30** 93–111.

CAO, J. and YAO, W. (2012). Semiparametric mixture of binomial regression with a degenerate component. *Statistica Sinica* 27–46.

CAO, H., CHURPEK, M. M., ZENG, D. and FINE, J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. *Journal of the American Statistical Association* **110** 1187–1196.

CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* **92** 477–489.

COOK, R. J. and LAWLESS, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.

COOK, R. J., YI, G. Y., LEE, K.-A. and GLADMAN, D. D. (2004). A conditional Markov model for clustered progressive multistate processes under incomplete observation. *Biometrics* **60** 436–443.

DAI, H. and PAN, J. (2018). Joint modelling of survival and longitudinal data with informative observation times. *Scandinavian Journal of Statistics* **45** 571–589.

DE LA SIERRA, A., SEGURA, J., GOROSTIDI, M., BANEGAS, J. R., DE LA CRUZ, J. J. and RUILOPE, L. M. (2010). Diurnal blood pressure variation, risk categories and antihypertensive treatment. *Hypertension Research* **33** 767.

ELISAF, M. S., KALAITZIDIS, R. G., GOUDEVENOS, J. A., KATSARAKI, A. E., SIDERIS, D. A. and SIAMOPOULOS, K. C. (1999). Blood pressure profile in patients with microvascular angina. *Coronary artery disease* **10** 257–259.

FISHER, L. D. and LIN, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual review of public health* **20** 145–157.

GJESSING, H. K., RØYSLAND, K., PENA, E. A. and AALEN, O. O. (2010). Recurrent events and the exploding Cox model. *Lifetime data analysis* **16** 525–546.

HUANG, Y.-H., HWANG, W.-H. and 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.

HUANG, M. and YAO, W. (2012). Mixture of regression models with varying mixing proportions: a semiparametric approach. *Journal of the American Statistical Association* **107** 711–724.

IP, E. H., EFENDI, A., MOLENBERGHS, G. and BERTONI, A. G. (2015). Comparison of risks of cardiovascular events in the elderly using standard survival analysis and multiple-events and recurrent-events methods. *BMC medical research methodology* **15** 15.

KIM, S., ZENG, D., CHAMBLESS, L. and LI, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in biosciences* **4** 262–281.

MILOSLAVSKY, M., KELEŞ, S., LAAN, M. J. and BUTLER, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 239–257.

PEÑA, E. A. (2006). Dynamic modelling and statistical analysis of event times. *Statistical science: a review journal of the Institute of Mathematical Statistics* **21** 1.

ROGERS, J. K., YAROSHINSKY, A., POCOCK, S. J., STOKAR, D. and POGODA, J. (2016). Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in medicine* **35** 2195–2205.

SONG, X. and WANG, C. (2008). Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time-Varying Coefficients. *Biometrics* **64** 557–566.

SONG, X., WANG, L. et al. (2017). Partially time-varying coefficient proportional hazards models with error-prone time-dependent covariatesan application to the AIDS Clinical Trial Group 175 data. *The Annals of Applied Statistics* **11** 274–296.

TSIATIS, A. A. and DAVIDIAN, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88** 447–

458.

Wang, C.-Y. (2006). Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica* 235–253.

Wattanakit, K., Folsom, A. R., Chambless, L. E. and Nieto, F. J. (2005). Risk factors for cardiovascular event recurrence in the Atherosclerosis Risk in Communities (ARIC) study. *American heart journal* **149** 606–612.

Xiao, W., Lu, W. and Zhang, H. H. (2016). Joint structure selection and estimation in the time-varying coefficient Cox model. *Statistica Sinica* **26** 547.

Yan, J. and Huang, J. (2012). Model selection for Cox models with time-varying coefficients. *Biometrics* **68** 419–428.

Yang, W., Jepson, C., Xie, D., Roy, J. A., Shou, H., Hsu, J. Y., Anderson, A. H., Landis, J. R., He, J., Feldman, H. I. et al. (2017). Statistical methods for recurrent event analysis in cohort studies of CKD. *Clinical Journal of the American Society of Nephrology* **12** 2066–2073.

Zhao, Y., Park, J., Izadnegahdar, M., Lee, M., Khan, N., Rabkin, S., Guan, M., Grubisic, M., Peng, D., Daniele, P. et al. (2017). FACTORS ASSOCIATED WITH VERY LOW DIASTOLIC BLOOD PRESSURE IN THE SPRINT TRIAL. *Canadian Journal of Cardiology* **33** S199.

Zhou, J., Zhang, J., Mclain, A. C., Lu, W., Sui, X. and Hardin, J. W. (2019). A varying-coefficient generalized odds rate model with time-varying exposure: An application to fitness and cardiovascular disease mortality. *Biometrics* **75** 853–863.

Address of the First and Third authors

Department of Mathematics, The University of Manchester, Oxford Road, Manchester, M13 9PL, U.K.
E-mail: chuoxin.ma@manchester.ac.uk
        Jianxin.Pan@manchester.ac.uk

Address of the Second author

Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, U.K.
E-mail: hdaia@essex.ac.uk