The Annals of Applied Statistics
2024, Vol. 18, No. 3, 2444–2461
https://doi.org/10.1214/24-AOAS1889

JOINT MODELING OF MULTISTATE AND NONPARAMETRIC MULTIVARIATE LONGITUDINAL DATA

BY LU YOU[1,a], FALASTIN SALAMI[2,c], CARINA TÖRN[2,d], ÅKE LERNMARK[2,e] AND ROY TAMURA[1,b]

[1]*Health Informatics Institute, University of South Florida,* [a]*lu.you@epi.usf.edu,* [b]*roy.tamura@epi.usf.edu*
[2]*Department of Clinical Sciences, Lund University,* [c]*falastin.salami@med.lu.se,* [d]*carina.torn@med.lu.se,* [e]*ake.lernmark@med.lu.se*

It is oftentimes the case in studies of disease progression that subjects can move into one of several disease states of interest. Multistate models are an indispensable tool to analyze data from such studies. The Environmental Determinants of Diabetes in the Young (TEDDY) is an observational study of at-risk children from birth to onset of type-1 diabetes (T1D) up through the age of 15. A joint model for simultaneous inference of multistate and multivariate nonparametric longitudinal data is proposed to analyze data and answer the research questions brought up in the study. The proposed method allows us to make statistical inferences, test hypotheses, and make predictions about future state occupation in the TEDDY study. The performance of the proposed method is evaluated by simulation studies. The proposed method is applied to the motivating example to demonstrate the capabilities of the method.

1. Introduction. To understand the natural progression of diseases, it is helpful to define some stages, disease subtypes, and intermediate events to gain insight into the process of disease development. In many clinical studies, researchers will specify several disease states of interest and collect data on the timing of state transitions. Multistate models are an indispensable tool to analyze such data.

In this paper we present a method for jointly modeling multistate and longitudinal data. In addition to the data on state transitions, longitudinal biomarkers will also be collected during the follow-up period to investigate the relationship between state transitions and the longitudinal pattern of biomarkers. The proposed method intends to enrich existing multistate models to incorporate longitudinal data. This research is primarily motivated by The Environmental Determinants of Diabetes in the Young (TEDDY) study that follows children at risk for type-1 diabetes (T1D) up through the age of 15. T1D is a chronic autoimmune disease caused by the destruction of insulin-producing beta cells in the pancreas (Atkinson, Eisenbarth and Michels (2014)). The presence of several T1D-related autoantibodies is one of the most important markers of beta-cell autoimmunity in T1D (Krischer et al. (2015)). Though research has shown that children with more than one type of autoantibody present much higher risks of T1D and different types of autoantibodies carry differential risks, there are still limited discussions on how different combinations of autoantibodies may interact and change the risks of T1D. In the TEDDY study, children with genetic predispositions of T1D are monitored for the presence of autoantibodies. We obtained a dataset from this study with repeated measures of body mass index (BMI) and hemoglobin A1c (HbA1c) tests as well as the time when each participant is tested positive for certain autoantibodies. This dataset provides an opportunity to investigate the risks of T1D for different autoantibody status and the chronological order of autoantibody development. In the analysis we will treat each autoantibody status as a state and use a multistate model to quantify the transition probabilities between different states.

Received June 2023; revised February 2024.
Key words and phrases. Joint modeling, multistate model, spline regression model, type-1 diabetes.

The event history data and longitudinal data will be modeled jointly in one model to make inferences about the relationship between longitudinal and event history data.

Initial efforts in joint modeling include those of Schluchter (1992) and Gruttola and Tu (1994) along with the influential works by Wulfsohn and Tsiatis (1997) and Tsiatis and Davidian (2004). A comprehensive review of the field is found in Albert (2019). Joint models have been increasingly used to assemble and analyze the information from different outcome variables and make simultaneous inference by combining several different models. In the literature there have been some proposals of joint models for longitudinal and multistate data, but most existing methods cannot be directly applied here due to the complexity of the current dataset. The earliest works deal with a special case where the event history data can be described by a competing risks model where each individual has only one observed transition (c.f., Huang, Li and Elashoff (2010), Li et al. (2012), Williamson et al. (2008)). Alafchi et al. (2021) proposed a three-state model with a univariate longitudinal marker that is described by quadratic growth curves. Ferrer et al. (2016) considered a more complex multistate model with a univariate parametric longitudinal model. Yiu and Tom (2017) considered a multistate model with a special structure that follows the time-homogeneous Markov property. Other alternative methods model the sequence of state transitions as multivariate survival data and use a shared frailty term to model all survival endpoints of interest (e.g., Chi and Ibrahim (2006), Li, Lesperance and Wu (2022), Liu, Huang and O'Quigley (2008)). These methods can not be applied to our current problem due to the following reasons. First, our dataset includes multiple longitudinal markers whose trajectories are not adequately modeled by a linear or a parametric model of analytical functions. The parametric assumptions made in these methods make them difficult to adequately describe the longitudinal pattern in our dataset. Second, our study defines seven states by participants' autoantibody and disease status with a nontrivial structure. Some of these methods that presume a specific structure on the multistate models cannot be directly applied on that account. Finally, most multivariate survival models assume that the multivariate survival times are independent, given the common frailty. They would not be particularly helpful when the chronological order of the event sequences is of main interest in the analysis. For these reasons we will present a new method to analyze this dataset.

The remainder of the paper will be organized as follows. In Section 2 we introduce the proposed model as well as the method for model estimation and statistical inference. In Section 3 we present simulation studies to evaluate and validate the proposed method. In Section 4 we define various states in the aforementioned TEDDY dataset, analyze the TEDDY dataset as of mid-2021, and give interpretations of the results. Section 5 concludes the paper by discussing several possible extensions and areas of future research.

2. Methodology.

2.1. *Data and notation.* Suppose that in a study where m subjects are being followed, we have access to both longitudinal data and event history data. To establish a relationship between these two types of data, we align them using a common time origin, such as study enrollment or birth. We note that all subsequent times mentioned in the analysis will be directly referenced to this time origin unless otherwise specified. The longitudinal dataset consists of n_y longitudinal markers. The jth observation of the kth longitudinal variable for the ith subject is obtained at time t_{ikj}, which will be denoted by y_{ikj} ($1 \le i \le m$ and $1 \le j \le n_{ik}$). There are n_s discrete states under consideration. The relationship between these states can be described by a directed graph $(\mathbb{V}, \mathbb{E})$, where the set of vertices $\mathbb{V}$ represent all the n_s states, and the set of arrows $\mathbb{E}$ consists of all possible transitions between the states. In this paper we will number the states by $1, 2, \ldots, n_s$ (i.e., $\mathbb{V} = \{1, \ldots, n_s\}$) and $(s_1, s_2) \in \mathbb{E}$ if and only if the transition from s_1 to s_2 is possible. We will let $\mathbb{N}(s_1) = \{s_2 : (s_1, s_2) \in \mathbb{E}\}$

denote the states that can be transitioned to from state s_1. In this paper we will restrict our analysis to cases where there are no cycles in the directed graph $(\mathbb{V}, \mathbb{E})$. For $(s_1, s_2) \in \mathbb{E}$, we let $T^*_{i,s_1\to s_2}$ denote the underlying true transition times when the ith subject transitions from state s_1 to s_2, and $T^*_{i,s_1\to s_2} = \infty$ if the ith subject did not to make a transition from s_1 to s_2. ($T^*_{i,s_1\to s_2}$ are uniquely defined since there are no cycles in the directed graph.) Like most survival models, the transition times are subject to censoring and may not be observed in the study period. We let E_i and C_i denote the study entry time and censoring time of the ith subject. We let $\delta_{i,s_1\to s_2} = 1$ if the ith subject is observed to transition from s_1 to s_2 (i.e., $E_i < T^*_{i,s_1\to s_2} \le C_i$) and 0 otherwise. The observed transition time is denoted by $T_{i,s_1\to s_2}$, where $T_{i,s_1,\to s_2} = T^*_{i,s_1,\to s_2}$ if $\delta_{i,s_1\to s_2} = 1$ and $T_{i,s_1,\to s_2} = \infty$ otherwise. For a state s_1 and a time t, we let $y_{i,s_1}(t) = 1$ indicate if the ith subject is observed to be in state s_1 at time t and $y_{i,s_1}(t) = 0$ if the ith subject is in other states or t is beyond the interval $(E_i, C_i]$. Additional covariates can be included to model the longitudinal data and multistate data. We let $\mathbf{x}_i(t)$ be the n_x-dimensional vector of covariates for subject i in the longitudinal model and $\mathbf{z}_i(t)$ be the n_z-dimensional vector of covariates for subject i in the multistate model. This manuscript primarily focuses on scenarios where the covariates are nontime-varying, (i.e., $\mathbf{x}_i(t) = \mathbf{x}_i$ and $\mathbf{z}_i(t) = \mathbf{z}_i$).

2.2. *Joint models of multistate and longitudinal data.* The longitudinal data are observed with errors and the true underlying trajectory of the kth longitudinal marker of the ith subject will be denoted by $m_{ik}(t)$. The longitudinal model can be written as follows:

$$y_{ikj} = m_{ik}(t_{ikj}) + \epsilon_{ikj}, \tag{1}$$

where ϵ_{ikj} is the measurement error term that is assumed to follow normal distribution with mean 0 and variance σ_k^2. The trajectories $m_{ik}(t)$ will be modeled by the following nonparametric mixed effects model:

$$m_{ik}(t) = \sum_{l=1}^{n_{Bk}} b_{ikl} B_{kl}(t) + \boldsymbol{\xi}_k^\top \mathbf{x}_i(t) = \sum_{l=1}^{n_{Bk}} c_{kl} B_{kl}(t) + \sum_{l=1}^{n_{Bk}} a_{ikl} B_{kl}(t) + \boldsymbol{\xi}_k^\top \mathbf{x}_i(t), \tag{2}$$

where $\{B_{kl}(t) : l = 1, \ldots, n_{Bk}\}$ is a collection of spline basis functions for modeling the nonparametric trajectory of $m_{ik}(t)$, $\mathbf{b}_{ik} = (b_{ik1}, \ldots, b_{ikn_{Bk}})^\top$ is the vector of coefficients, and $\boldsymbol{\xi}_k$ is the coefficients for covariates $\mathbf{x}_i(t)$. $\mathbf{b}_{ik}$ can be split into the sum of the fixed effects coefficients $\mathbf{c}_k = (c_{k1}, \ldots, c_{kn_{Bk}})^\top$ and the random effects coefficients $\mathbf{a}_{ik} = (a_{ik1}, \ldots, a_{ikn_{Bk}})^\top$, representing population mean coefficients and subject-specific variation from the mean respectively. In this paper we consider using B-spline (de Boor (1978)) basis functions with knots at evenly spaced quantiles of observation times, which is a common method in previous research (e.g., Brown, Ibrahim and DeGruttola (2005)). While other types of splines may also be used, for a detailed discussion on the comparisons between spline choices, we refer readers to Wand (2000) and Perperoglou et al. (2019). Let $\mathbf{b}_i = (\mathbf{b}_{i1}^\top, \ldots, \mathbf{b}_{in_y}^\top)^\top$, $\mathbf{c} = (\mathbf{c}_1^\top, \ldots, \mathbf{c}_{n_y}^\top)^\top$, and $\mathbf{a}_i = (\mathbf{a}_{i1}^\top, \ldots, \mathbf{a}_{in_y}^\top)^\top$ be the n_b-dimensional vector obtained by stacking the subvectors $\mathbf{b}_{ik}$, $\mathbf{c}_k$, and $\mathbf{a}_{ik}$ ($n_b = \sum_{k=1}^{n_y} n_{Bk}$). We assume that $\mathbf{a}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}_b$, and thus $\mathbf{b}_i$ follows a normal distribution with mean $\mathbf{c}$ and variance-covariance matrix $\boldsymbol{\Sigma}_b$. In order to simplify the formula for parameter estimation in matrix form, we introduce the following notations: $\mathbf{y}_{ik} = (y_{ik1}, \ldots, y_{ikn_{ik}})^\top$, $\mathbf{B}_k(t) = (B_{k1}(t), \ldots, B_{kn_{Bk}}(t))^\top$, $\mathbf{B}_{ik} = [\mathbf{B}_k(t_{ik1}), \ldots, \mathbf{B}_k(t_{ikn_{ik}})]^\top$, $\mathbf{X}_{ik} = [\mathbf{x}_i(t_{ik1}), \ldots, \mathbf{x}_i(t_{ikn_{ik}})]^\top$, and $\boldsymbol{\epsilon}_{ik} = (\epsilon_{ik1}, \ldots, \epsilon_{ikn_{ik}})^\top$. Using these notations, we can express $\mathbf{y}_{ik}$ as the sum of three components $\mathbf{B}_{ik}\mathbf{b}_{ik} + \mathbf{X}_{ik}\boldsymbol{\xi}_k + \boldsymbol{\epsilon}_{ik}$. Under these notations we can see that the variance-covariance matrix $\boldsymbol{\Sigma}_b$ captures the correlations between the observed longitudinal data. The covariance between $\mathbf{y}_{ik_1}$ and $\mathbf{y}_{ik_2}$ can be expressed as

$\mathrm{Cov}(\mathbf{y}_{ik_1}, \mathbf{y}_{ik_2}) = \mathbf{B}_{ik_1}\mathbf{J}_{k_1}\boldsymbol{\Sigma}_b\mathbf{J}_{k_2}^\top\mathbf{B}_{ik_2}^\top$, where $\mathbf{J}_k = (\mathbf{0}_{n_{Bk}\times n_{B1}}, \ldots, \mathbf{I}_{n_{Bk}\times n_{Bk}}, \ldots, \mathbf{0}_{n_{Bk}\times n_{Bn_y}})$ is a row subview of the identity matrix $\mathbf{I}_{n_b\times n_b}$ corresponding to the dimensions of $\mathbf{b}_{ik}$.

The multistate model aims to quantify the probability of making certain state transitions. We use

$$\lambda_{i,s_1\to s_2}(t) = \lim_{dt\downarrow 0} P\{T^*_{i,s_1\to s_2} \in (t, t+dt] | y_{i,s_1}(t) = 1\}/dt$$

to denote the true cause-specific hazard of transitioning from state s_1 to s_2 for the ith subject, which is the instantaneous probability of entering s_2, given that the subject is in s_1 at time t. $\lambda_{i,s_1\to s_2}(t)$ is quantified by the following proportional hazards model:

$$\lambda_{i,s_1\to s_2}(t) = \lambda_{0,s_1}(t)\exp\{\alpha_{s_1\to s_2} + \boldsymbol{\beta}_{s_1\to s_2}^\top\mathbf{m}_i(t) + \boldsymbol{\gamma}_{s_1\to s_2}^\top\mathbf{z}_i(t)\}, \tag{3}$$

where $\alpha_{s_1\to s_2}$ is a constant parameter that models the relative risks of transitioning from s_1 to s_2, $\boldsymbol{\beta}_{s_1\to s_2}$ is a n_y-dimensional regression coefficient for $\mathbf{m}_i(t)$, and $\boldsymbol{\gamma}_{s_1\to s_2}$ is a n_z-dimensional regression coefficient for $\mathbf{z}_i(t)$. By this formulation the log hazard ratio of transitioning to s_2' vs. s_2 is

$$\log\{\lambda_{i,s_1\to s_2'}(t)/\lambda_{i,s_1\to s_2}(t)\} = \alpha_{s_1\to s_2'} - \alpha_{s_1\to s_2}$$

when the values of $\mathbf{m}_i(t)$ and $\mathbf{z}_i(t)$ are held constant. To simplify notations throughout this paper, we let $\boldsymbol{\theta}_{s_1\to s_2} = (\alpha_{s_1\to s_2}, \boldsymbol{\beta}_{s_1\to s_2}^\top, \boldsymbol{\gamma}_{s_1\to s_2}^\top)^\top$, and $\boldsymbol{\eta}_i(t) = (1, \mathbf{m}_i(t)^\top, \mathbf{z}_i(t)^\top)^\top$. Accordingly, $\lambda_{i,s_1\to s_2}(t)$ can be expressed as $\lambda_{i,s_1\to s_2}(t) = \lambda_{0,s_1}(t)\exp\{\boldsymbol{\theta}_{s_1\to s_2}^\top\boldsymbol{\eta}_i(t)\}$.

2.3. *The likelihood function.* Next, we will formulate the likelihood function based on the defined model. The likelihood function for joint models is comprised of three components, one for the longitudinal model, one for the multistate model, and a mixed effects model that links the longitudinal and multistate models. Let $\mathcal{Y}_i = \{(y_{ikj}, t_{ikj}) : k = 1, \ldots, n_y, j = 1, \ldots n_{ik}\}$ be the collection of observed longitudinal data of the ith subject and $\mathcal{E}_i = \{(\delta_{i,s_1\to s_2}, T_{i,s_1\to s_2}) : (s_1, s_2) \in \mathbb{E}\}$ be the collection of event history data for the ith subject. Based on the model, we can derive the log probability density for $\mathcal{Y}_i$, given $\mathbf{b}_i$,

$$\log f(\mathcal{Y}_i|\mathbf{b}_i; \boldsymbol{\Theta}_{\xi,\sigma}) = \sum_{k=1}^{n_y}\left[-\frac{1}{2\sigma_k^2}\sum_{j=1}^{n_{ik}}\big(y_{ikj} - \mathbf{b}_{ik}^\top\mathbf{B}_k(t_{ikj}) - \boldsymbol{\xi}_k^\top\mathbf{x}_i(t_{ikj})\big)^2 - \frac{n_{ik}}{2}\log\{2\pi\sigma_k^2\}\right],$$

where $\boldsymbol{\Theta}_{\xi,\sigma} = \{\boldsymbol{\xi}_k, \sigma_k^2 : k = 1, \ldots, n_y\}$. The log probability density for $\mathcal{E}_i$, given $\mathbf{b}_i$ is

$$\begin{aligned}
&\log f(\mathcal{E}_i|\mathbf{b}_i; \boldsymbol{\Theta}_{\alpha,\beta,\gamma})\\
&\quad= \sum_{(s_1,s_2)\in\mathbb{E}}\left[\delta_{i,s_1\to s_2}\log\{\lambda_{i,s_1\to s_2}(T_{i,s_1\to s_2})\} - \int_{E_i}^{C_i} y_{i,s_1}(t)\lambda_{i,s_1\to s_2}(t)\,dt\right]\\
&\quad= \sum_{(s_1,s_2)\in\mathbb{E}}\Bigg[\delta_{i,s_1\to s_2}\{\log\lambda_{0,s_1}(T_{i,s_1\to s_2}) + \boldsymbol{\theta}_{s_1\to s_2}^\top\boldsymbol{\eta}_i(T_{i,s_1\to s_2})\}\\
&\qquad- \sum_{i'=1}^{m}\sum_{s_2':(s_1,s_2')\in\mathbb{E}}\delta_{i',s_1\to s_2'}y_{i,s_1}(T_{s_1\to s_2',i'})\lambda_{0,s_1}(T_{s_1\to s_2',i'})\exp\{\boldsymbol{\theta}_{s_1\to s_2}^\top\boldsymbol{\eta}_i(T_{s_1\to s_2',i'})\}\Bigg],
\end{aligned}$$

where $\boldsymbol{\Theta}_{\alpha,\beta,\gamma} = \{\alpha_{s_1\to s_2}, \boldsymbol{\beta}_{s_1\to s_2}, \boldsymbol{\gamma}_{s_1\to s_2} : (s_1, s_2) \in \mathbb{E}\}$ is the collection of all parameters in the multistate model. Lastly, the log probability density for $\mathbf{b}_i$ is

$$\log f(\mathbf{b}_i|\boldsymbol{\Theta}_b) = -\frac{1}{2}(\mathbf{b}_i - \mathbf{c})^\top\boldsymbol{\Sigma}_b^{-1}(\mathbf{b}_i - \mathbf{c}) - \frac{1}{2}\log\det\{2\pi\,\boldsymbol{\Sigma}_b\},$$

where $\boldsymbol{\Theta}_b = \{\mathbf{c}, \boldsymbol{\Sigma}_b\}$ is the collection of parameters in the nonparametric mixed-effects model. The joint likelihood for all parameters is then given by

$$L(\boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b) = \prod_{i=1}^{m} \int f(\mathcal{Y}_i|\mathbf{b}_i; \boldsymbol{\Theta}_{\xi,\sigma}) f(\mathcal{E}_i|\mathbf{b}_i; \boldsymbol{\Theta}_{\alpha,\beta,\gamma}) f(\mathbf{b}_i|\boldsymbol{\Theta}_b)\, d\mathbf{b}_i.$$

In order to differentiate it from the complete and profile likelihood functions that we will introduce later, we will use the term observed likelihood to refer to the likelihood described above.

Estimates of the parameters are obtained by maximizing the above joint likelihood function. Here we should give an important remark that certain constraints on $\alpha_{s_1\to s_2}$ are needed in the above model to ensure that there is a unique parameter set that maximizes the joint likelihood function. Suppose that we obtained parameters that maximize joint likelihood function with $\hat{\lambda}^{(0)}_{i,s_1\to s_2}(t)$ and $\hat{\alpha}^{(0)}_{s_1\to s_2}$ being the estimates for $\lambda_{i,s_1\to s_2}(t)$ and $\alpha_{s_1\to s_2}$. Then we can obtain a family of estimates that also minimizes the above function by defining $\hat{\lambda}^{(\tau)}_{i,s_1\to s_2}(t) = \hat{\lambda}^{(0)}_{i,s_1\to s_2}(t)\exp\{\tau\}$ and $\hat{\alpha}^{(\tau)}_{i,s_1\to s_2} = \hat{\alpha}^{(0)}_{i,s_1\to s_2}\exp\{-\tau\}$ for any real-valued τ. The appropriate constraint should rule out the extra one degree of freedom in the parameterization for all transitions out of the same state s_1. Throughout this paper we will apply the constraint that $\alpha_{s_1\to s_2} = 0$ for $s_2 = \min\{s_2 : (s_1, s_2) \in \mathbb{E}\}$, which is similar to the corner-point constraint in the analysis of variance models. The corresponding transition (s_1, s_2) with $\alpha_{s_1\to s_2}$ imposed to 0 is the reference transition for all transitions out of s_1. Also similar to the analysis of variance models, $\alpha_{s_1\to s_2}$ can be interpreted as the log hazard ratio of transitioning to s_2 compared to the reference transition.

2.4. *Model estimation by EM algorithm.* The joint likelihood function involves an intractable integration on a multidimensional space, so the parameter estimates derived from the maximum likelihood inference do not have a closed-form solution. A numerical algorithm is needed to find the parameter estimates. In the next section, we will present an expectation-maximization (EM) algorithm (c.f., Dempster, Laird and Rubin (1977)) to solve this problem.

The derivation of the EM algorithm here closely follows the one presented in Wulfsohn and Tsiatis (1997). To simplify the notations a bit, we let $\tilde{\boldsymbol{\Theta}}_{\xi,\sigma}$, $\tilde{\boldsymbol{\Theta}}_{\alpha,\beta,\gamma}$, $\tilde{\boldsymbol{\Theta}}_b$ be the estimated parameters at the current iteration, $\tilde{\mathrm{E}}_i[\cdot] = \mathrm{E}[\cdot|\mathcal{Y}_i, \mathcal{E}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b]$ be the expectation of random variables, given the observed data of the ith subject, and $\tilde{\mathrm{E}}[\cdot] = \mathrm{E}[\cdot|\{\mathcal{Y}_i, \mathcal{E}_i\}_{i=1}^m; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b]$ be the expectation of random variables, given all observed data. The EM algorithm is composed as an iterative rundown of the E-step and the M-step. The E-step concerns evaluating the expectation of the complete log-likelihood based on the observed data and the current estimates of parameters. The M-step then attempts to maximize the expectation of the complete log-likelihood and update the estimates of parameters. As the maximization of the expectation of the complete log-likelihood may not have a closed-form solution for all parameters, for some parameters we will update the parameters by a Newton–Raphson step. To derive the algorithm, we first write out the complete log-likelihood function, which is the log joint probability density of the complete data $(\mathcal{Y}_i, \mathcal{E}_i, \mathbf{b}_i)$,

$$(4)\quad l(\boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b) = \sum_{i=1}^{m} \big[\log f(\mathcal{Y}_i|\mathbf{b}_i; \boldsymbol{\Theta}_{\xi,\sigma}) + \log f(\mathcal{E}_i|\mathbf{b}_i; \boldsymbol{\Theta}_{\alpha,\beta,\gamma}) + \log f(\mathbf{b}_i|\boldsymbol{\Theta}_b)\big].$$

By differentiating $\tilde{\mathrm{E}}[l(\boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b)]$ with respect to each parameter and solving the system of equations, we can derive the following formula for updating the parameters:

$$
\begin{aligned}
\tilde{\boldsymbol{\xi}}_k &\leftarrow \left(\sum_{i=1}^{m} \mathbf{X}_{ik}^{\mathsf{T}} \mathbf{X}_{ik}\right)^{-1} \left[\sum_{i=1}^{m} \mathbf{X}_{ik}^{\mathsf{T}} \big(\mathbf{y}_{ik} - \mathbf{B}_{ik} \tilde{\mathrm{E}}_i[\mathbf{b}_{ik}]\big)\right], \\
\tilde{\sigma}_k^2 &\leftarrow \sum_{i=1}^{m} \sum_{j=1}^{n_{ik}} \tilde{\mathrm{E}}_i\big[\big(y_{ikj} - \mathbf{b}_{ik}^{\mathsf{T}} \mathbf{B}(t_{ikj}) - \boldsymbol{\xi}_k^{\mathsf{T}} \mathbf{x}_i(t_{ikj})\big)^2\big] / \sum_{i=1}^{m} n_{ik}, \\
\tilde{\mathbf{c}} &\leftarrow \sum_{i=1}^{m} \tilde{\mathrm{E}}_i[\mathbf{b}_i]/m, \\
\tilde{\boldsymbol{\Sigma}}_b &\leftarrow \sum_{i=1}^{m} \tilde{\mathrm{E}}_i\big[(\mathbf{b}_i - \mathbf{c})(\mathbf{b}_i - \mathbf{c})^{\mathsf{T}}\big]/m, \\
\tilde{\boldsymbol{\theta}}_{s_1 \to s_2} &\leftarrow \boldsymbol{\theta}_{s_1 \to s_2} - \left[\frac{\partial^2 \tilde{\mathrm{E}}[l]}{\partial \boldsymbol{\theta}_{s_1 \to s_2}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1 \to s_2}}\right]^g \left[\frac{\partial \tilde{\mathrm{E}}[l]}{\partial \boldsymbol{\theta}_{s_1 \to s_2}}\right], \\
\tilde{\lambda}_{0,s_1}(T_{s_1 \to s_2, i}) &\leftarrow \frac{1}{\sum_{s_2' \in \mathbb{N}(s_1)} \sum_{i'=1}^{m} y_{i',s_1}(T_{i,s_1 \to s_2}) \tilde{\mathrm{E}}_i[\exp\{\boldsymbol{\theta}_{s_1 \to s_2'}^{\mathsf{T}} \boldsymbol{\eta}_{i'}(T_{i,s_1 \to s_2})\}]},
\end{aligned}
\tag{5}
$$

where we will evaluate all the parameters on the right-hand-side of the formula with $\boldsymbol{\Theta}_{\xi,\sigma} = \tilde{\boldsymbol{\Theta}}_y$, $\boldsymbol{\Theta}_{\alpha,\beta,\gamma} = \tilde{\boldsymbol{\Theta}}_{\alpha,\beta,\gamma}$, $\boldsymbol{\Theta}_b = \tilde{\boldsymbol{\Theta}}_b$ on the right-hand side of the assignment operators ($\leftarrow$), we adopt the convention that $\frac{\partial}{\partial \alpha_{s_1 \to s_2}} = 0$ if $\alpha_{s_1 \to s_2}$ is fixed to be 0, and $[\cdot]^g$ is the Moore–Penrose generalized inverse when matrices are rank-deficient. When updating the parameter $\boldsymbol{\theta}_{s_1 \to s_2}$, a profile likelihood technique can be used to facilitate convergence (c.f., Hsieh, Tseng and Wang (2006), Murphy and van der Vaart (2000)). By this technique the big numbers of parameters in $\lambda_{0,s_1}(t)$ can be treated as nuisance parameters, which allows us to focus on the estimation of $\boldsymbol{\theta}_{s_1 \to s_2}$. Notice that

$$
\lambda_{0,s_1}(T_{s_1 \to s_2, i}) = \frac{1}{\sum_{s_2' \in \mathbb{N}(s_1)} \sum_{i'=1}^{m} y_{i',s_1}(T_{i,s_1 \to s_2}) \tilde{\mathrm{E}}_i[\exp\{\boldsymbol{\theta}_{s_1 \to s_2'}^{\mathsf{T}} \boldsymbol{\eta}_{i'}(T_{i,s_1 \to s_2})\}]}
\tag{6}
$$

when the expected log complete likelihood is maximized. By treating $\lambda_{0,s_1}(T_{s_1 \to s_2, i})$ as an implicit function of $\boldsymbol{\theta}_{s_1 \to s_2}$ when differentiating $\tilde{\mathrm{E}}[l(\boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b)]$, we can refine the updating procedure (5) as follows:

$$
\tilde{\boldsymbol{\theta}}_{s_1 \to s_2} \leftarrow \boldsymbol{\theta}_{s_1 \to s_2} - \left[\frac{\partial^2 pl(\boldsymbol{\theta}_{s_1 \to s_2})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1 \to s_2}}\right]^g \left[\frac{\partial pl(\boldsymbol{\theta}_{s_1 \to s_2})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}}\right],
$$

where $pl(\boldsymbol{\theta}_{s_1 \to s_2})$ is the profile likelihood function,

$$
\begin{aligned}
\frac{\partial pl(\boldsymbol{\theta}_{s_1 \to s_2})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}} &= \sum_{s_2' \in \mathbb{N}(s_1)} \sum_{i'=1}^{m} \delta_{i',s_1 \to s_2'} \left\{-\frac{1}{A_{s_1}(T_{i',s_1 \to s_2'})} \frac{\partial A_{s_1}(T_{i',s_1 \to s_2'})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}}\right\}, \\
\frac{\partial^2 pl(\boldsymbol{\theta}_{s_1 \to s_2})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1 \to s_2}} &= \sum_{s_2' \in \mathbb{N}(s_1)} \sum_{i'=1}^{m} \delta_{i',s_1 \to s_2'} \left\{-\frac{1}{A_{s_1}(T_{i',s_1 \to s_2'})} \frac{\partial^2 A_{s_1}(T_{i',s_1 \to s_2'})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1 \to s_2}}\right. \\
&\qquad \left. + \frac{1}{A_{s_1}(T_{i',s_1 \to s_2'})^2} \left[\frac{\partial A_{s_1}(T_{i',s_1 \to s_2'})}{\partial \boldsymbol{\theta}_{s_1 \to s_2}}\right]^{\otimes 2}\right\},
\end{aligned}
$$

and

$$A_{s_1}(t) = \sum_{s_2' \in \mathbb{N}(s_1)} \sum_{i'=1}^{m} y_{i',s_1}(t) \tilde{\mathrm{E}}_{i'}\big[\exp\{\boldsymbol{\theta}_{s_1 \to s_2'} \boldsymbol{\eta}_{i'}(t)\}\big].$$

A detailed derivation of the profile likelihood is presented in Section S1.1 of Supplementary Material (You et al. (2024)). Following Booth and Hobert (1999) and Ye, Li and Guan (2015), the convergence of parameter estimates is declared when the relative changes in the individual parameters are sufficiently small. Other considerations for facilitating convergence and the details about the convergence assessment are given in Section S5 of Supplementary Material (You et al. (2024)).

The estimation procedure described above involves several conditional expectations of the random variable $\mathbf{b}_i$ that do not have closed-form solutions. Suppose that we need to evaluate the conditional expectation of some function $\boldsymbol{\psi}$ of $\mathbf{b}_i$; we can apply Bayes' rule to rewrite the expression as a set of multidimensional integrals with respect to $\mathbf{b}_i$,

$$\begin{aligned}
\tilde{\mathrm{E}}_i\big[\boldsymbol{\psi}(\mathbf{b}_i)\big] &= \mathrm{E}\big[\boldsymbol{\psi}(\mathbf{b}_i) | \mathcal{Y}_i, \mathcal{E}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b\big] \\
&= \int \boldsymbol{\psi}(\mathbf{b}_i) f(\mathbf{b}_i | \mathcal{Y}_i, \mathcal{E}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b)\, d\mathbf{b}_i \\
&= \frac{\int \boldsymbol{\psi}(\mathbf{b}_i) f(\mathcal{E}_i | \mathbf{b}_i, \mathcal{Y}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b) f(\mathbf{b}_i | \mathcal{Y}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_b)\, d\mathbf{b}_i}{\int f(\mathcal{E}_i | \mathbf{b}_i, \mathcal{Y}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_{\alpha,\beta,\gamma}, \boldsymbol{\Theta}_b) f(\mathbf{b}_i | \mathcal{Y}_i; \boldsymbol{\Theta}_{\xi,\sigma}, \boldsymbol{\Theta}_b)\, d\mathbf{b}_i}.
\end{aligned}$$

The multidimensional integrations above can be evaluated using low-discrepancy quasi-Monte Carlo sequences. The quasi-Monte Carlo method is a numerical integration technique that aims to approximate high-dimensional integrals using low-discrepancy sequences of points in the integration domain (c.f., Niederreiter (1978), Sobol (1976)). Unlike the traditional Monte Carlo methods, which rely on randomly generated points, quasi-Monte Carlo methods use deterministic sequences that are designed to cover the multidimensional integration domain in a more regular and evenly spaced way. Both theoretical and numerical studies have shown that quasi-Monte Carlo methods offer improved accuracy compared to traditional Monte Carlo methods when the dimensionality of integration is high (e.g., dimensionality > 10) (c.f., Joy, Boyle and Tan (1996), Sloan and Woźniakowski (1998)). For details on the calculation, please refer to Section S1.2 of the Supplementary Material (You et al. (2024)).

2.5. *Selection of tuning parameters and statistical inference.* The number of B-spline basis functions $\{n_{Bk}\}_{k=1}^{n_y}$ for modeling the trajectory of longitudinal markers is a set of tuning parameters in the model that needs to be determined appropriately. A larger value of n_{Bk} allows for more flexibility in describing the trajectories but may suffer from the problem of overfitting. There have been many existing methods for properly selecting n_{Bk}. In this paper we will adopt the method by Yao (2007), which uses the Akaike information criterion (AIC) to determine n_{Bk}. Let $\hat{\boldsymbol{\Theta}}_\sigma$, $\hat{\boldsymbol{\Theta}}_{\alpha,\beta,\gamma}$ and $\hat{\boldsymbol{\Theta}}_\sigma$ be the estimates of $\boldsymbol{\Theta}_\sigma$, $\boldsymbol{\Theta}_{\alpha,\beta,\gamma}$ and $\boldsymbol{\Theta}_b$, respectively. AIC is then defined by

$$2N_{\text{parameter}} - 2\log L(\hat{\boldsymbol{\Theta}}_\sigma, \hat{\boldsymbol{\Theta}}_{\alpha,\beta,\gamma}, \hat{\boldsymbol{\Theta}}_b),$$

where the first term is the number of parameters in the model multiplied by 2, and the second term is the log-likelihood function evaluated at the parameter estimates multiplied by 2. The values of $\{n_{Bk}\}_{k=1}^{n_y}$ that minimize AIC are searched by exploring a grid of points.

Next, we discuss methods to make statistical inferences about the coefficients $\alpha_{s_1 \to s_2}$, $\boldsymbol{\beta}_{s_1 \to s_2}$ and $\gamma_{s_1 \to s_2}$. One commonly used method is to take the second derivatives of the

profile likelihood function. Let $\boldsymbol{\theta}_{s_1\to\bullet} = \{\boldsymbol{\theta}_{s_1\to s_2} : s_2 \in \mathbb{N}(s_1)\}$. The information matrix of $\boldsymbol{\theta}_{s_1\to\bullet}$, based on the profile likelihood function, is given by

$$\mathbf{I}_p(\boldsymbol{\theta}_{s_1\to\bullet}) = -\left[\frac{\partial^2 pl}{\partial \boldsymbol{\theta}_{s_1\to\bullet}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1\to\bullet}}\right],$$

where the formula is derived in Section S1.3 of the Supplementary Material (You et al. (2024)). To estimate the covariance matrix of $\hat{\boldsymbol{\theta}}_{s_1\to\bullet}$, we plug in the estimated parameters to the expression and take the generalized inverse of $\mathbf{I}_p(\boldsymbol{\theta}_{s_1\to\bullet})$. The covariance matrix can then be used to construct standard errors, p-values, and confidence intervals. However, Hsieh, Tseng and Wang (2006) pointed out that the above information matrix, based on the profile likelihood, is an approximation of the true Fisher's information matrix, as it does not account for the additional variability due to $\lambda_{s_1,0}(t)$. To address this issue, they outlined an alternative approach for deriving Fisher's information by differentiating the original log-likelihood function. In this study we follow this idea to investigate this method further. We consider the following projection to obtain the Fisher information for $\boldsymbol{\theta}_{s_1\to\bullet}$:

$$\mathbf{I}(\boldsymbol{\theta}_{s_1\to\bullet}) = -\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\theta}_{s_1\to\bullet}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1\to\bullet}}\right] - \left[\frac{\partial^2 \log L}{\partial \boldsymbol{\lambda}_{s_1}^{\mathsf{T}} \partial \boldsymbol{\theta}_{s_1\to\bullet}}\right]\left[\frac{\partial^2 \log L}{\partial \boldsymbol{\lambda}_{s_1}^{\mathsf{T}} \partial \boldsymbol{\lambda}_{s_1}}\right]^g \left[\frac{\partial^2 \log L}{\partial \boldsymbol{\theta}_{s_1\to\bullet}^{\mathsf{T}} \partial \boldsymbol{\lambda}_{s_1}}\right],$$

where $\boldsymbol{\lambda}_{s_1}$ is the collection of all the parameters to be estimated in $\lambda_{s_1,0}(t)$. Similarly, we can obtain the estimated covariance matrix of $\hat{\boldsymbol{\theta}}_{s_1\to\bullet}$ by plugging in the estimated parameters and take the generalized inverse of $\mathbf{I}(\boldsymbol{\theta}_{s_1\to\bullet})$. It has been shown in Hsieh, Tseng and Wang (2006) that standard errors based on $\mathbf{I}_p(\boldsymbol{\theta}_{s_1\to\bullet})$ are often more optimistic than those based on $\mathbf{I}(\boldsymbol{\theta}_{s_1\to\bullet})$. However, calculating $\mathbf{I}(\boldsymbol{\theta}_{s_1\to\bullet})$ involves inverting a large matrix, which can be computationally expensive and requires significant memory. As a comparison, the calculation of $\mathbf{I}_p(\boldsymbol{\theta}_{s_1\to\bullet})$ is relatively straightforward, and its diagonal elements can be obtained from the last iteration of the EM algorithm. In summary, the choice of using $\mathbf{I}(\boldsymbol{\theta}_{s_1\to\bullet})$ or $\mathbf{I}_p(\boldsymbol{\theta}_{s_1\to\bullet})$ depends on the trade-off between computational efficiency and accuracy. In the following sections, we will compare the performance of these two methods through a simulation study and a data application.

3. Simulation studies. In this section we will present the results of simulation studies to investigate the performance of the proposed model. In the simulation studies, we let $m = 500$, $n_y = 2$, $n_z = 2$, and $n_x = 1$ and considered two cases with different state-transition diagrams and longitudinal trajectories. Case (A) assumes a five-state model while Case (B) assumes a seven-state model, as illustrated in Figure 1. In Case (A) we let the longitudinal trajectories follow quadratic trajectories, similar to other parametric joint models (c.f., Xu and Zeger (2001), Yang, Yu and Gao (2016)). Longitudinal data are simulated from model (1) with $m_{ik}(t) = \xi_{ik0} + \xi_{ik1}t + \xi_{ik2}t^2$, where ξ_{ik0}, ξ_{ik1}, and ξ_{ik2} follow independent $N(0, 1)$, $N(1, 1)$, and $N(-1, 1)$ distribution, respectively. The covariates z_i follow independent Bernoulli(0.5) distribution, $\xi_1 = 0.2$, and $\xi_2 = -0.2$. The error terms ϵ_{ikj} follow independent $N(0, 2)$ distribution. The entry times $E_i = 0$ and censoring times C_i follow independent Weibull(2,2) distribution. The observation times t_{ikj} follow independent uniform distribution in $[(j-1)/100, j/100]$, and the observations (y_{ikj}, t_{ikj}) are subject to censoring by E_i and C_i (i.e., observed only when $E_i < t_{ikj} \le C_i$). The true values of regression coefficients $\alpha_{s_1\to s_2}$, $\boldsymbol{\beta}_{s_1\to s_2}$, and $\boldsymbol{\gamma}_{s_1\to s_2}$ are presented in the columns "True" in Table 1, and the true baseline hazards are $\lambda_{0,1}(t) = 1$ and $\lambda_{0,2}(t) = \lambda_{0,3}(t) = \sqrt{t}$. The transition times $T_{i,s_1\to s_2}$ are simulated based on the parameters and are also subject to censoring by E_i and C_i (i.e., observed only if $E_i < T_{i,s_1\to s_2} \le C_i$). In Case (B) we consider a more complicated

scenario where the longitudinal trajectory is nonparametric. The trajectories $m_{ik}(t)$ are now assumed to be

$$m_{i1}(t) = \xi_{i10} + \xi_{i11}\sin(\omega_{i1}\pi t + \phi_{i1}),$$
$$m_{i2}(t) = \xi_{i20} + \xi_{i21}\cos(\omega_{i2}\pi t + \phi_{i2}),$$

where $\xi_{i10}, \xi_{i11}, \xi_{i20}, \xi_{i21}$ follow independent $N(1, 0.1)$ distribution, ω_{i1}, ω_{i2} follow independent uniform distribution in $[1, 2]$, and ϕ_{i1}, ϕ_{i2} follow uniform distribution in $[0, 2\pi]$. The covariates z_i follow independent Bernoulli(0.5) distribution, $\xi_1 = 0.2$, and $\xi_2 = -0.2$. The error terms ϵ_{ikj} follow independent $N(0, 2)$ distribution. The entry times $E_i = 0$ and censoring times C_i follow independent Weibull(2,2) distribution. The observation times t_{ikj} follow independent uniform distribution in $[(j-1)/100, j/100]$. The true values of regression coefficients $\alpha_{s_1\to s_2}$, $\boldsymbol{\beta}_{s_1\to s_2}$, and $\boldsymbol{\gamma}_{s_1\to s_2}$ are given in the columns "True" in Table 1, and the true baseline hazards are $\lambda_{0,s_1}(t) = 1$ for all s_1. The transition times $T_{i,s_1\to s_2}$ are simulated based on the parameters. Similarly, both the observations (y_{ikj}, t_{ikj}) and transition times $T_{i,s_1\to s_2}$ are subject to censoring. All simulations are repeated 500 times.

3.1. *Joint models and the statistical inferences.* In Table 1 we evaluated the bias of estimated coefficients and the probability that the true values will fall in the 95% confidence intervals (CI). The bias of estimated coefficients are presented in the columns "Bias." Two versions of confidence intervals (CI) are compared: one derived from the information matrix of the profile likelihood and the other derived from the information matrix of the observed likelihood. The coverage probabilities of these confidence intervals are presented in the columns labeled "Coverage (Profile)" and "Coverage (Observed)," respectively. In most cases the biases of estimated coefficients are negligible, and the probabilities that the 95% CI cover the true value of parameters are all close to 0.95. This indicates that the proposed method for model inference performs effectively across different scenarios. Specifically, it appears that the approximation of the high-dimensional integrals are satisfactory for datasets of similar sizes. The CI based on the observed likelihood are slightly more conservative compared to those based on the profile likelihood. Nevertheless, we do not observe significant differences in the coverage probabilities, implying that the approximate inference based on the profile likelihood can also be a reliable approach. In Section S2 of Supplementary Material (You et al. (2024)), we repeated the simulation study with a different sample size $m = 250$. The results similarly showed that these 95% CI have a coverage rate close to 0.95. However, we observe more cases where the CI based on the profile likelihood are smaller than 0.95, and there are more cases where the CI based on the observed likelihood is closer to 0.95 compared to those based on the profile likelihood.

3.2. *Joint models vs. separate competing risk models.* Additionally, we compared the proposed method with the method that uses separate competing-risks regression models with time-varying covariates. In the comparison model, the following proportional hazards model is assumed as a substitute for model (3):

$$\lambda_{i,s_1\to s_2}(t) = \lambda_{0,s_1}(t)\exp\{\alpha_{s_1\to s_2} + \boldsymbol{\beta}_{s_1\to s_2}^{\mathsf{T}}\mathbf{y}_i(t) + \boldsymbol{\gamma}_{s_1\to s_2}^{\mathsf{T}}\mathbf{z}_i(t)\},$$

where both $\mathbf{y}_i(t)$ and $\mathbf{z}_i(t)$ are included as time-varying covariates that are generated by y_{ikj} and $\mathbf{z}_i(t)$. For transitions that start from the same state s_1, this model can be fitted separately using a software package for implementing competing risks models. In Table 2 we compared the estimates obtained from the proposed joint model and separate competing-risks models by biases and root-mean-squared errors (RMSE). We can see that, when the effect sizes are nonzero, using separate competing risks models will usually underestimate the effect, and the RMSE of the estimates obtained from separate models are usually larger than those obtained

TABLE 1

True values, biases of estimated parameters, and probabilities of true values within the 95% CI. The left and right panels give the results of Cases (A) and (B), respectively. Values in the parentheses are the corresponding standard errors

Parameter	True	Bias	Coverage (Profile)	Coverage (Observed)
$\beta_{1\to2,1}$	0.4	0.003 (0.003)	0.953 (0.009)	0.957 (0.009)
$\beta_{1\to2,2}$	0.2	−0.004 (0.003)	0.943 (0.010)	0.949 (0.010)
$\gamma_{1\to2,1}$	0.1	−0.003 (0.003)	0.953 (0.009)	0.953 (0.009)
$\gamma_{1\to2,2}$	0.3	0.005 (0.003)	0.939 (0.011)	0.941 (0.011)
$\alpha_{1\to2}$	0.0	—	—	—
$\beta_{1\to3,1}$	0.2	0.003 (0.003)	0.943 (0.010)	0.953 (0.009)
$\beta_{1\to3,2}$	0.4	0.002 (0.003)	0.927 (0.012)	0.935 (0.011)
$\gamma_{1\to3,1}$	0.3	0.005 (0.003)	0.941 (0.011)	0.941 (0.011)
$\gamma_{1\to3,2}$	0.1	0.001 (0.003)	0.935 (0.011)	0.937 (0.011)
$\alpha_{1\to3}$	0.4	0.007 (0.005)	0.955 (0.009)	0.949 (0.010)
$\beta_{2\to4,1}$	0.0	−0.003 (0.006)	0.929 (0.012)	0.933 (0.011)
$\beta_{2\to4,2}$	0.3	0.008 (0.006)	0.939 (0.011)	0.939 (0.011)
$\gamma_{2\to4,1}$	0.1	0.011 (0.007)	0.933 (0.011)	0.933 (0.011)
$\gamma_{2\to4,2}$	0.2	0.007 (0.007)	0.949 (0.010)	0.949 (0.010)
$\alpha_{2\to4}$	0.0	—	—	—
$\beta_{2\to5,1}$	0.3	0.019 (0.005)	0.931 (0.011)	0.937 (0.011)
$\beta_{2\to5,2}$	0.0	−0.002 (0.004)	0.955 (0.009)	0.962 (0.009)
$\gamma_{2\to5,1}$	0.2	−0.002 (0.005)	0.953 (0.009)	0.953 (0.009)
$\gamma_{2\to5,2}$	0.1	0.005 (0.005)	0.957 (0.009)	0.957 (0.009)
$\alpha_{2\to5}$	0.4	0.016 (0.010)	0.962 (0.009)	0.957 (0.009)
$\beta_{3\to4,1}$	−0.2	−0.001 (0.004)	0.927 (0.012)	0.929 (0.012)
$\beta_{3\to4,2}$	0.4	0.012 (0.004)	0.947 (0.010)	0.947 (0.010)
$\gamma_{3\to4,1}$	0.1	0.002 (0.005)	0.947 (0.010)	0.951 (0.010)
$\gamma_{3\to4,2}$	0.2	−0.001 (0.005)	0.945 (0.010)	0.947 (0.010)
$\alpha_{3\to4}$	0.0	—	—	—
$\beta_{3\to5,1}$	0.4	0.012 (0.005)	0.945 (0.010)	0.953 (0.009)
$\beta_{3\to5,2}$	−0.2	0.003 (0.005)	0.927 (0.012)	0.933 (0.011)
$\gamma_{3\to5,1}$	0.2	0.009 (0.006)	0.935 (0.011)	0.935 (0.011)
$\gamma_{3\to5,2}$	0.1	0.007 (0.006)	0.955 (0.009)	0.955 (0.009)
$\alpha_{3\to5}$	−0.4	−0.002 (0.009)	0.955 (0.009)	0.955 (0.009)

Parameter	True	Bias	Coverage (Profile)	Coverage (Observed)
$\beta_{1\to4,1}$	0.4	0.006 (0.007)	0.929 (0.012)	0.939 (0.011)
$\beta_{1\to4,2}$	0.2	−0.006 (0.007)	0.931 (0.011)	0.949 (0.010)
$\gamma_{1\to4,1}$	−0.2	0.007 (0.005)	0.962 (0.009)	0.962 (0.009)
$\gamma_{1\to4,2}$	0.4	0.017 (0.006)	0.947 (0.010)	0.947 (0.010)
$\alpha_{1\to4}$	0.0	—	—	—
$\beta_{1\to5,1}$	0.2	−0.000 (0.007)	0.933 (0.011)	0.947 (0.010)
$\beta_{1\to5,2}$	0.2	0.000 (0.007)	0.929 (0.012)	0.945 (0.010)
$\gamma_{1\to5,1}$	−0.2	0.000 (0.006)	0.943 (0.010)	0.945 (0.010)
$\gamma_{1\to5,2}$	0.4	0.015 (0.006)	0.949 (0.010)	0.949 (0.010)
$\alpha_{1\to5}$	0.3	0.005 (0.017)	0.939 (0.011)	0.955 (0.009)
$\beta_{2\to5,1}$	0.3	−0.004 (0.006)	0.935 (0.011)	0.945 (0.010)
$\beta_{2\to5,2}$	0.3	0.024 (0.007)	0.923 (0.012)	0.951 (0.010)
$\gamma_{2\to5,1}$	0.4	0.009 (0.005)	0.953 (0.009)	0.953 (0.009)
$\gamma_{2\to5,2}$	−0.2	−0.004 (0.005)	0.964 (0.008)	0.964 (0.008)
$\alpha_{2\to5}$	0.0	—	—	—
$\beta_{2\to6,1}$	0.3	0.014 (0.007)	0.953 (0.009)	0.962 (0.009)
$\beta_{2\to6,2}$	0.3	0.001 (0.008)	0.927 (0.012)	0.949 (0.010)
$\gamma_{2\to6,1}$	0.4	0.004 (0.006)	0.941 (0.011)	0.943 (0.010)
$\gamma_{2\to6,2}$	−0.2	−0.011 (0.006)	0.935 (0.011)	0.935 (0.011)
$\alpha_{2\to6}$	−0.3	−0.009 (0.017)	0.929 (0.012)	0.949 (0.010)
$\beta_{3\to4,1}$	0.4	0.011 (0.008)	0.933 (0.011)	0.949 (0.010)
$\beta_{3\to4,2}$	0.2	−0.003 (0.007)	0.937 (0.011)	0.951 (0.010)
$\gamma_{3\to4,1}$	0.2	−0.005 (0.006)	0.955 (0.009)	0.955 (0.009)
$\gamma_{3\to4,2}$	0.2	0.001 (0.006)	0.955 (0.009)	0.955 (0.009)
$\alpha_{3\to4}$	0.0	—	—	—
$\beta_{3\to6,1}$	0.2	−0.003 (0.007)	0.911 (0.013)	0.935 (0.011)
$\beta_{3\to6,2}$	0.4	0.008 (0.007)	0.927 (0.012)	0.939 (0.011)
$\gamma_{3\to6,1}$	0.2	0.005 (0.005)	0.945 (0.010)	0.947 (0.010)
$\gamma_{3\to6,2}$	0.2	−0.007 (0.005)	0.945 (0.010)	0.945 (0.010)
$\alpha_{3\to6}$	0.3	0.013 (0.017)	0.937 (0.011)	0.951 (0.010)
$\beta_{4\to7,1}$	0.2	0.012 (0.006)	0.945 (0.010)	0.955 (0.009)
$\beta_{4\to7,2}$	0.2	−0.000 (0.006)	0.951 (0.010)	0.962 (0.009)
$\gamma_{4\to7,1}$	0.4	0.011 (0.005)	0.949 (0.010)	0.949 (0.010)
$\gamma_{4\to7,2}$	0.0	0.006 (0.005)	0.943 (0.010)	0.945 (0.010)
$\alpha_{4\to7}$	0.0	—	—	—
$\beta_{5\to7,1}$	0.0	0.007 (0.005)	0.949 (0.010)	0.957 (0.009)
$\beta_{5\to7,2}$	0.4	−0.001 (0.006)	0.957 (0.009)	0.968 (0.008)
$\gamma_{5\to7,1}$	0.0	0.002 (0.005)	0.917 (0.012)	0.917 (0.012)
$\gamma_{5\to7,2}$	0.4	0.002 (0.005)	0.945 (0.010)	0.947 (0.010)
$\alpha_{5\to7}$	0.0	—	—	—
$\beta_{6\to7,1}$	0.4	0.004 (0.006)	0.903 (0.013)	0.921 (0.012)
$\beta_{6\to7,2}$	0.0	−0.006 (0.006)	0.957 (0.009)	0.957 (0.009)
$\gamma_{6\to7,1}$	0.3	−0.001 (0.005)	0.945 (0.010)	0.947 (0.010)
$\gamma_{6\to7,2}$	0.3	0.001 (0.005)	0.953 (0.009)	0.955 (0.009)
$\alpha_{6\to7}$	0.0	—	—	—

from the proposed joint model. It is disadvantageous to use separate competing-risks models when the main purpose of the analysis is to discover baseline and longitudinal markers that have major effects on the state transitions. The proposed method demonstrates good performance, as it combines the longitudinal model and the multistate model that incorporates all

TABLE 2

Comparing the bias and root-mean-squared errors (RMSE) of the parameters estimates using joint models and separate competing risks models. Values in the parentheses are the corresponding standard errors

			$\beta_{s_1\to s_2,1}$			$\beta_{s_1\to s_2,2}$			$\gamma_{s_1\to s_2,1}$			$\gamma_{s_1\to s_2,2}$			$\alpha_{s_1\to s_2}$		
s_1	s_2	Method	Truth	Bias	RMSE	Truth	Bias	RMSE	Truth	Bias	RMSE	Truth	Bias	RMSE	Truth	Bias	RMSE
Case (A)																	
1	2	Joint Models	0.4	0.003 (0.003)	0.072 (0.017)	0.2	−0.004 (0.003)	0.075 (0.017)	0.1	−0.003 (0.003)	0.074 (0.016)	0.3	0.005 (0.003)	0.077 (0.016)	–	–	–
1	2	Separate Models	0.4	−0.262 (0.002)	0.265 (0.004)	0.2	−0.137 (0.002)	0.143 (0.006)	0.1	−0.020 (0.003)	0.075 (0.016)	0.3	−0.012 (0.003)	0.076 (0.015)	–	–	–
1	3	Joint Models	0.2	0.003 (0.003)	0.064 (0.016)	0.4	0.002 (0.003)	0.064 (0.015)	0.3	0.005 (0.003)	0.066 (0.016)	0.1	0.001 (0.003)	0.067 (0.016)	0.4	0.007 (0.005)	0.111 (0.016)
1	3	Separate Models	0.2	−0.131 (0.002)	0.136 (0.006)	0.4	−0.262 (0.002)	0.265 (0.003)	0.3	−0.015 (0.003)	0.067 (0.015)	0.1	−0.012 (0.003)	0.067 (0.016)	0.4	−0.024 (0.005)	0.107 (0.015)
2	4	Joint Models	0.0	−0.003 (0.006)	0.134 (0.017)	0.3	0.008 (0.006)	0.132 (0.017)	0.1	0.011 (0.007)	0.162 (0.019)	0.2	0.007 (0.007)	0.159 (0.018)	–	–	–
2	4	Separate Models	0.0	−0.003 (0.003)	0.076 (0.017)	0.3	−0.175 (0.003)	0.191 (0.009)	0.1	−0.002 (0.007)	0.157 (0.018)	0.2	−0.003 (0.007)	0.156 (0.018)	–	–	–
2	5	Joint Models	0.3	0.019 (0.005)	0.108 (0.018)	0.0	−0.002 (0.004)	0.093 (0.015)	0.2	−0.002 (0.005)	0.116 (0.017)	0.1	0.005 (0.005)	0.119 (0.016)	0.4	0.016 (0.010)	0.222 (0.017)
2	5	Separate Models	0.3	−0.178 (0.003)	0.188 (0.007)	0.0	−0.005 (0.003)	0.061 (0.016)	0.2	−0.014 (0.005)	0.114 (0.018)	0.1	−0.006 (0.005)	0.116 (0.017)	0.4	0.098 (0.009)	0.218 (0.017)
3	4	Joint Models	−0.2	−0.001 (0.004)	0.091 (0.016)	0.4	0.012 (0.004)	0.096 (0.018)	0.1	0.002 (0.005)	0.105 (0.016)	0.2	−0.001 (0.005)	0.111 (0.018)	–	–	–
3	4	Separate Models	−0.2	0.119 (0.003)	0.133 (0.009)	0.4	−0.235 (0.003)	0.241 (0.005)	0.1	−0.005 (0.004)	0.099 (0.016)	0.2	−0.012 (0.005)	0.108 (0.017)	–	–	–
3	5	Joint Models	0.4	0.012 (0.005)	0.111 (0.016)	−0.2	0.003 (0.005)	0.110 (0.016)	0.2	0.009 (0.006)	0.132 (0.017)	0.1	0.007 (0.006)	0.127 (0.017)	−0.4	−0.002 (0.009)	0.210 (0.016)
3	5	Separate Models	0.4	−0.229 (0.003)	0.239 (0.006)	−0.2	0.117 (0.003)	0.136 (0.011)	0.2	0.001 (0.006)	0.127 (0.017)	0.1	−0.002 (0.006)	0.124 (0.017)	−0.4	0.015 (0.009)	0.190 (0.015)
Case (B)																	
1	4	Joint Models	0.4	0.006 (0.007)	0.160 (0.015)	0.2	−0.006 (0.007)	0.157 (0.015)	−0.2	0.007 (0.005)	0.116 (0.015)	0.4	0.017 (0.006)	0.129 (0.016)	–	–	–
1	4	Separate Models	0.4	−0.308 (0.009)	0.315 (0.014)	0.2	−0.171 (0.010)	0.186 (0.026)	−0.2	0.023 (0.020)	0.145 (0.091)	0.4	−0.003 (0.015)	0.112 (0.044)	–	–	–
1	5	Joint Models	0.2	−0.000 (0.007)	0.154 (0.018)	0.2	0.000 (0.007)	0.152 (0.019)	−0.2	0.000 (0.006)	0.122 (0.015)	0.4	0.015 (0.006)	0.125 (0.016)	0.3	0.005 (0.017)	0.374 (0.016)
1	5	Separate Models	0.2	−0.135 (0.008)	0.148 (0.026)	0.2	−0.150 (0.011)	0.169 (0.032)	−0.2	−0.010 (0.022)	0.165 (0.074)	0.4	0.011 (0.018)	0.131 (0.048)	0.3	−0.232 (0.025)	0.294 (0.037)
2	5	Joint Models	0.3	−0.004 (0.006)	0.144 (0.016)	0.3	0.024 (0.007)	0.148 (0.015)	0.4	0.009 (0.005)	0.113 (0.017)	−0.2	−0.004 (0.005)	0.111 (0.016)	–	–	–
2	5	Separate Models	0.3	−0.227 (0.009)	0.236 (0.020)	0.3	−0.225 (0.008)	0.233 (0.017)	0.4	−0.019 (0.016)	0.117 (0.048)	−0.2	0.010 (0.018)	0.130 (0.097)	–	–	–
2	6	Joint Models	0.3	0.014 (0.007)	0.166 (0.017)	0.3	0.001 (0.008)	0.173 (0.016)	0.4	0.004 (0.006)	0.141 (0.017)	−0.2	−0.011 (0.006)	0.135 (0.018)	−0.3	−0.009 (0.017)	0.386 (0.017)
2	6	Separate Models	0.3	−0.238 (0.010)	0.249 (0.020)	0.3	−0.249 (0.011)	0.262 (0.021)	0.4	−0.018 (0.017)	0.127 (0.041)	−0.2	0.003 (0.020)	0.149 (0.072)	−0.3	0.055 (0.032)	0.242 (0.040)
3	4	Joint Models	0.4	0.011 (0.008)	0.170 (0.018)	0.2	−0.003 (0.007)	0.163 (0.015)	0.2	−0.005 (0.006)	0.128 (0.016)	0.2	0.001 (0.006)	0.124 (0.016)	–	–	–
3	4	Separate Models	0.4	−0.312 (0.009)	0.320 (0.015)	0.2	−0.171 (0.010)	0.185 (0.025)	0.2	0.006 (0.017)	0.126 (0.041)	0.2	−0.003 (0.018)	0.135 (0.044)	–	–	–
3	6	Joint Models	0.2	−0.003 (0.007)	0.147 (0.016)	0.4	0.008 (0.007)	0.149 (0.015)	0.2	0.005 (0.005)	0.113 (0.018)	0.2	−0.007 (0.005)	0.113 (0.017)	0.3	0.013 (0.017)	0.383 (0.017)
3	6	Separate Models	0.2	−0.150 (0.009)	0.163 (0.024)	0.4	−0.303 (0.010)	0.311 (0.015)	0.2	0.017 (0.017)	0.126 (0.058)	0.2	−0.009 (0.013)	0.099 (0.042)	0.3	−0.088 (0.029)	0.229 (0.038)
4	7	Joint Models	0.2	0.012 (0.006)	0.136 (0.015)	0.2	−0.000 (0.006)	0.133 (0.015)	0.4	0.011 (0.005)	0.112 (0.018)	0.0	0.006 (0.005)	0.110 (0.016)	–	–	–
4	7	Separate Models	0.2	−0.152 (0.010)	0.168 (0.027)	0.2	−0.158 (0.009)	0.171 (0.023)	0.4	0.008 (0.015)	0.108 (0.042)	0.0	−0.005 (0.017)	0.124 (0.058)	–	–	–
5	7	Joint Models	0.0	0.007 (0.005)	0.121 (0.015)	0.4	−0.001 (0.006)	0.123 (0.016)	0.0	0.002 (0.005)	0.106 (0.017)	0.4	0.002 (0.005)	0.104 (0.016)	–	–	–
5	7	Separate Models	0.0	−0.006 (0.008)	0.062 (0.038)	0.4	−0.308 (0.009)	0.314 (0.014)	0.0	−0.007 (0.016)	0.115 (0.076)	0.4	−0.019 (0.012)	0.087 (0.043)	–	–	–
6	7	Joint Models	0.4	0.004 (0.006)	0.144 (0.017)	0.0	−0.006 (0.006)	0.123 (0.016)	0.3	−0.001 (0.005)	0.106 (0.018)	0.3	0.001 (0.005)	0.110 (0.016)	–	–	–
6	7	Separate Models	0.4	−0.313 (0.009)	0.320 (0.014)	0.0	0.003 (0.009)	0.067 (0.050)	0.3	0.004 (0.012)	0.091 (0.067)	0.3	−0.015 (0.010)	0.079 (0.050)	–	–	–

transitions of interest, which allows the models to acquire information from each other and reduce variability in parameter estimates.

4. TEDDY application. In this section we apply the proposed method to the TEDDY dataset introduced in Section 1. In this study, participants are scheduled to undergo autoantibody tests every three months following the detection of a single autoantibody until the onset of T1D or age 15. The dataset consists of a cohort of 691 participants who developed glutamic acid decarboxylase autoantibodies (GADA) or micro insulin autoantibodies (IAA) as the first autoantibodies. All participants are followed for three types of autoantibodies, GADA, IAA, and insulinoma-associated antigen-2 antibodies (IA-2A). The ages at which these autoantibodies were consistently confirmed are recorded, and the observation times are aligned by birth in the analysis. A fourth autoantibody, zinc transporter 8 autoantibody, was also examined in the TEDDY study, but sampling for this autoantibody was not done for all subjects at all time points, so we excluded this autoantibody from our analysis. The cohort excludes participants who developed IA-2A as the first autoantibody because IA-2A is usually considered an indicator of advanced autoimmunity and is rarely present as a single autoantibody (c.f., Yu, Zhao and Steck (2017)). In this analysis we defined seven states with six states representing all possible combinations of autoantibody status and one state representing the development of T1D. In Figure 2 all possible transitions between autoantibody states are represented by arrows. Participants in each autoantibody state can transition to the T1D state, and those transitions are reduced to a bold arrow to simplify the figure. Besides event history data, there are also demographic data and longitudinal data in this dataset. Participants' demographics include gender and a binary variable indicating records of T1D in first-degree relatives (FDR), which are modeled as covariates in the multistate model (3). The longitudinal markers include measurements of BMI and HbA1c. These measures demonstrate significant changes over time among all children in the study and are modeled by (1) and (2) with adjustment for gender. The time origin is chosen to be the time of birth. All observation times and event times are expressed as years since the date of birth, and the baseline hazards and the mean trajectories are both modeled as functions of the age.

We applied the proposed method to this dataset. The total number of parameters in the model is $\sum_{k=1}^{2} n_{Bk} + (\sum_{k=1}^{2} n_{Bk})^2 + 71$ (see formulas in Section S1.4 of Supplementary Material, You et al. (2024)). The AIC determined that the number of basis functions n_{B1} is 14 for BMI and n_{B2} is four for HbA1c, resulting a total of 415 parameters. The median CPU time required to fit the model for various combinations of n_{B1} and n_{B2} is 38.95 hours. From the results we can obtain estimates of the means and standard deviations of BMI and HbA1c trajectories from the estimates of $\mathbf{c}$ and $\mathbf{\Sigma}_b$. The estimated mean and standard deviation functions are presented in Figure 3, which describes the longitudinal trajectories of the two markers in this population. The estimates of the regression coefficients $\boldsymbol{\beta}_{s_1 \to s_2}$ and $\boldsymbol{\gamma}_{s_1 \to s_2}$ are presented in Table 3 along with their corresponding 95% CI. We can observe that the CIs based on the observed likelihood is slightly wider than those based on the profile likelihood, though there is no major discrepancy in the interpretation of the results between the two approaches. The result concurs with the existing studies (Salami et al. (2022)) showing that elevated levels of HbA1c are related to the progression to T1D. We also ran a joint model replacing the B-splines with restricted cubic splines and found the results similar to Table 3; see Section S6 of the Supplementary Material (You et al. (2024)) for details. In TEDDY data a typical pathway of disease progression involves the development of IAA during early childhood (age < 4), followed by the development of GADA, and ultimately progression to T1D. The analysis shows that having a FDR is a significant risk factor for developing GADA after IAA, while elevated levels of HbA1c is a significant risk factor for progressing to T1D following the development of GADA and IAA. To examine the transition rates from a certain state to the next state, we calculated the cause-specific hazard functions of transitioning

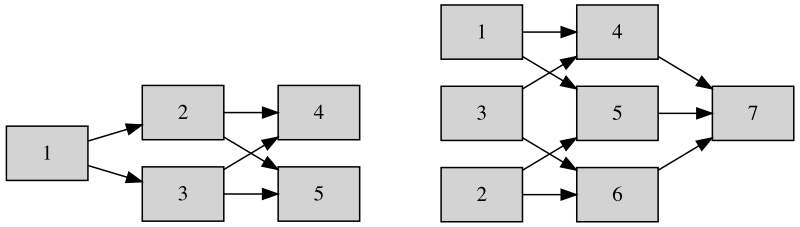

FIG. 1. *The diagram of possible transitions for simulations in Case* (*A*) (*left panel*) *and Case* (*B*) (*right panel*).

to the next state when covariates and longitudinal markers are held at the population mean levels and presented the results in Figure 4. From this figure it can be observed that the risk of developing autoantibodies and T1D peaks between the ages of one to three years and decline when children get older. When a participant is only persistently tested positive for IAA, GADA, or IAA and GADA, the risks of progressing to T1D are relatively small, compared to the risks of moving to another state with more autoantibodies. The development of IA-2A usually indicates an elevated risk of T1D. Participants who are tested positive for all the three types of autoantibodies have the highest risks to develop T1D around the age of two. The proposed multistate model allows us to understand this dataset in many different ways. It provides information about the chronological development of autoantibodies in children with T1D risks and offers insights into how the transitions are related to covariates and longitudinal markers. The proposed model can help us answer the research questions that cannot be directly addressed by survival and regression models with a single endpoint.

5. Discussion. Motivated by the TEDDY study, we developed a joint model for simultaneous inference of multistate and multivariate longitudinal data. This model uses a flexible longitudinal model to describe the nonparametric trajectories of longitudinal markers and directly estimates the effects of longitudinal markers on the state transitions. The proposed method provides a joint estimation of parameters in the longitudinal and multistate models. Extensive prior research has consistently showed the superiority of joint inference approaches over the two-step inference approaches where parameters in the time-to-event model are estimated subsequent to fitting the longitudinal processes (c.f., Sayers et al. (2017), Yang, Yu and Gao (2016)). The multistate model can also be conceptualized as a compilation of interconnected time-to-event models. The proposed method facilitates a comprehensive analysis that explores the complex evolution of state transition events and their associations with the dynamics of longitudinal markers.

The current study is still limited in several aspects, and a few possible extensions can be considered to improve the proposed method in future research. First, in the longitudinal model, we used B-splines to describe the nonparametric trajectories and tuned the number of basis functions by AIC. The performance of B-splines is sensitive to the placement of knots,

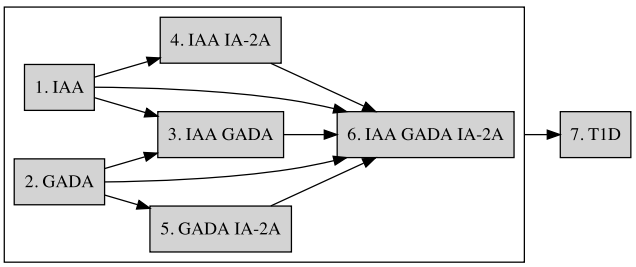

FIG. 2. *The diagram of possible state transitions in the real data application.*

TABLE 3
Estimated values of $\boldsymbol{\beta}_{s_1 \to s_2}$ and $\boldsymbol{\gamma}_{s_1 \to s_2}$ and their corresponding SE, p-values, and 95% CI in the data application

Variable	Estimate	95% CI (Profile)	95% CI (Observed)
IAA → IAA GADA			
BMI	0.073	(−0.026,0.172)	(−0.025,0.172)
HbA1c	−0.014	(−0.570,0.542)	(−0.677,0.649)
Sex	−0.064	(−0.468,0.339)	(−0.458,0.329)
FDR	0.470	(0.017,0.922)	(0.021,0.919)
IAA → IAA IA-2A			
BMI	−0.117	(−0.291,0.057)	(−0.294,0.059)
HbA1c	0.166	(−0.529,0.861)	(−0.680,1.011)
Sex	0.463	(−0.090,1.016)	(−0.090,1.016)
FDR	0.616	(0.033,1.199)	(0.033,1.200)
IAA → IAA GADA IA-2A			
BMI	−0.077	(−0.447,0.294)	(−0.454,0.301)
HbA1c	−0.201	(−1.739,1.338)	(−2.176,1.774)
Sex	−0.226	(−1.371,0.918)	(−1.374,0.921)
FDR	1.175	(0.013,2.338)	(0.012,2.339)
IAA → T1D			
BMI	0.097	(−0.093,0.288)	(−0.101,0.296)
HbA1c	2.663	(2.008,3.317)	(1.726,3.600)
Sex	0.281	(−0.732,1.295)	(−0.763,1.326)
FDR	0.852	(−0.149,1.853)	(−0.186,1.890)
GADA → IAA GADA			
BMI	0.034	(−0.052,0.120)	(−0.050,0.118)
HbA1c	−0.420	(−1.184,0.344)	(−1.270,0.430)
Sex	0.056	(−0.344,0.456)	(−0.316,0.428)
FDR	0.309	(−0.185,0.802)	(−0.177,0.795)
GADA → GADA IA-2A			
BMI	0.090	(−0.002,0.182)	(0.000,0.180)
HbA1c	−0.573	(−1.480,0.334)	(−1.601,0.455)
Sex	0.202	(−0.293,0.697)	(−0.293,0.698)
FDR	0.174	(−0.455,0.803)	(−0.455,0.803)
GADA → IAA GADA IA-2A			
BMI	−0.136	(−0.420,0.147)	(−0.423,0.150)
HbA1c	1.371	(−0.752,3.493)	(−0.946,3.687)
Sex	−0.337	(−1.487,0.813)	(−1.488,0.814)
FDR	1.382	(0.232,2.533)	(0.230,2.535)
GADA → T1D			
BMI	−0.031	(−0.238,0.176)	(−0.256,0.194)
HbA1c	6.605	(5.111,8.099)	(4.745,8.466)
Sex	−0.054	(−1.180,1.072)	(−1.321,1.213)
FDR	1.628	(0.462,2.794)	(0.283,2.973)
IAA GADA → IAA GADA IA-2A			
BMI	0.111	(0.020,0.202)	(0.019,0.204)
HbA1c	−0.015	(−0.626,0.596)	(−0.672,0.642)
Sex	0.428	(0.053,0.803)	(0.057,0.799)
FDR	0.147	(−0.271,0.564)	(−0.269,0.563)
IAA GADA → T1D			
BMI	0.086	(−0.099,0.270)	(−0.134,0.306)
HbA1c	5.752	(4.363,7.141)	(4.057,7.447)
Sex	−1.205	(−2.254,−0.155)	(−2.392,−0.018)
FDR	0.317	(−0.716,1.349)	(−0.809,1.442)
IAA IA-2A → IAA GADA IA-2A			
BMI	0.113	(−0.151,0.378)	(−0.158,0.385)
HbA1c	−0.223	(−1.589,1.143)	(−1.739,1.293)
Sex	−0.820	(−1.670,0.031)	(−1.672,0.033)
FDR	−0.819	(−2.108,0.470)	(−2.118,0.479)
IAA IA-2A → T1D			
BMI	0.104	(−0.073,0.281)	(−0.086,0.294)
HbA1c	1.700	(0.990,2.410)	(0.925,2.475)
Sex	−0.486	(−1.297,0.324)	(−1.306,0.333)
FDR	0.136	(−0.667,0.939)	(−0.682,0.955)
GADA IA-2A → IAA GADA IA-2A			
BMI	0.087	(−0.088,0.262)	(−0.089,0.263)
HbA1c	0.294	(−1.363,1.951)	(−1.457,2.045)
Sex	−0.908	(−1.808,−0.008)	(−1.792,−0.024)
FDR	0.173	(−1.013,1.359)	(−1.010,1.357)
GADA IA-2A → T1D			
BMI	0.030	(−0.151,0.211)	(−0.157,0.217)
HbA1c	4.672	(3.023,6.320)	(2.877,6.467)
Sex	−0.000	(−1.149,1.148)	(−1.171,1.171)
FDR	0.816	(−0.676,2.308)	(−0.705,2.337)
IAA GADA IA-2A → T1D			
BMI	0.076	(0.017,0.135)	(0.018,0.134)
HbA1c	2.600	(2.173,3.028)	(2.086,3.115)
Sex	−0.374	(−0.730,−0.017)	(−0.734,−0.013)
FDR	0.130	(−0.290,0.549)	(−0.300,0.559)

though we obtained satisfactory results in this example; in some other applications, it may not be flexible enough to fit the data closely. In the literature, many alternative longitudinal models have been considered. For example, Hsieh, Tseng and Wang (2006) employed a parametric model that is well-suited to the longitudinal pattern of the specific data. Yue and Al Kontar (2021) investigated Gaussian process models while Brumback and Rice (1998) investigated smoothing spline. Nonparametric approaches offer improved flexibility in capturing intricate longitudinal patterns without the constraint of predefined knot locations, but they can become computational intensive when there are multiple longitudinal markers. The local smoothing method is a nonparametric approach that reduces computational burden in multidimensional integration (c.f., You and Qiu (2021)). However, its effectiveness in estimating longitudinal trajectories can be limited by the availability of longitudinal observations in a neighborhood. It is thus worthwhile to discuss various factors that can influence the suitabil-

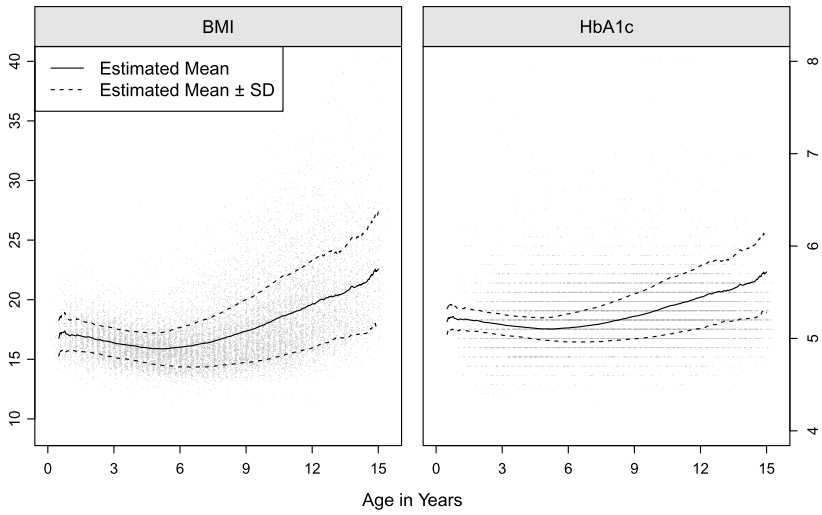

FIG. 3. *Estimated mean and mean ± SD trajectories for participants at risk of T1D in the TEDDY data application.*

ity of different longitudinal modeling approaches in the future research. Second, the multistate model considered in this paper is semiparametric and relies on the proportional hazards assumption. When the assumption is violated, this model can be made more flexible using several other techniques. For example, we may consider a time-varying effects model where $\alpha_{s_1 \to s_2}$, $\boldsymbol{\beta}_{s_1 \to s_2}$, and $\boldsymbol{\gamma}_{s_1 \to s_2}$ can vary with time or covariates. Third, the current analysis assumes that the timing of state transitions is directly observed through scheduled autoantibody tests conducted every three months. However, in other applications where the evaluation of

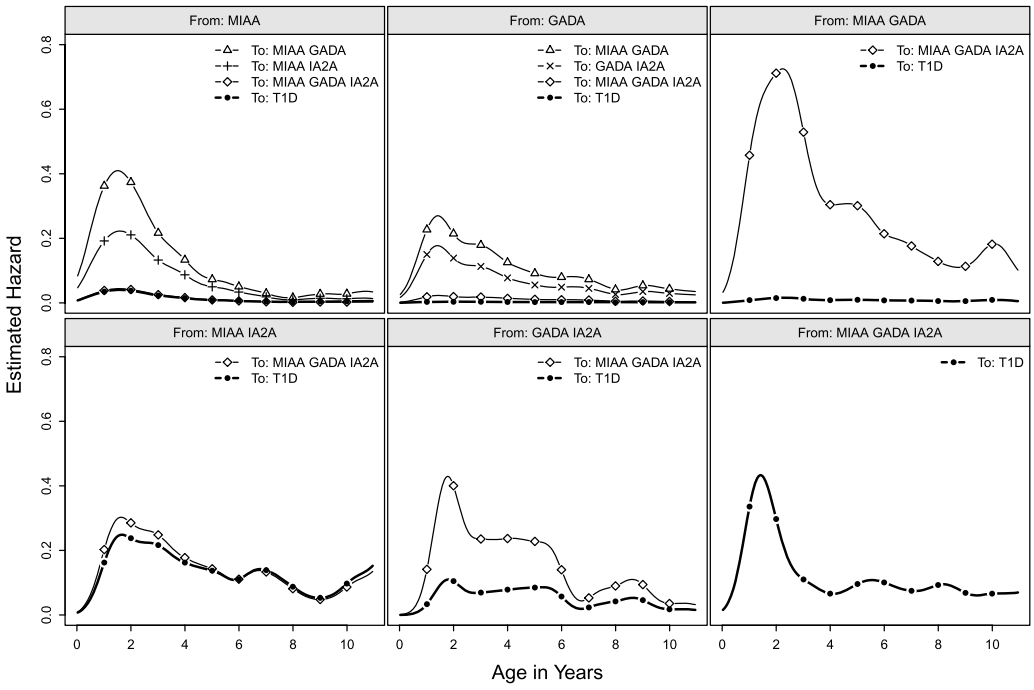

FIG. 4. *Estimated cause-specific hazard functions for different transitions when covariates and longitudinal markers are held at the population mean levels. The estimated hazard functions are smoothed using a bandwidth of half year.*

disease status occurs less frequently or missed visits are more likely, it is more accurate to treat transition times as interval-censored observations that occur between two consecutive evaluation visits. In order to extend the proposed method and accommodate interval-censored multistate data, we can investigate the likelihood-based method by Pak et al. (2017) and the EM-algorithm-based methods by Zhang et al. (2020) and Zeng and Lin (2021).

Our proposed model can be extended to a dynamic prediction model, similarly following the methods by Rizopoulos (2011). Given a participant's history observations, such a model will make predictions about the trajectories of longitudinal markers and estimate the probability of future state occupation and disease risks. This would be a practical tool to facilitate early prevention measures and can be used as a guide to provide personalized decisions. Also, the proposed method considers the case where the longitudinal markers are continuous and follow a normal distribution. Some existing research has considered the cases when the longitudinal markers are binary and count data and used generalized linear models to model the trajectories of longitudinal markers (c.f., Huang, Li and Guan (2014), Rizopoulos and Ghosh (2011)). Since binary and count longitudinal data are also common in many applications, future research could extend the current model to allow for these different types of longitudinal markers.

Data availability statement. Data analyzed for this study is available in the NIDDK Central Repository at https://www.niddkrepository.org/studies/teddy in accordance with the NIDDK's controlled-access authorization process.

Funding. Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Number R03DK135437. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The TEDDY study is funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, UC4 DK100238, UC4 DK106955, UC4 DK112243, UC4 DK117483, U01 DK124166, U01 DK128847, and Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Centers for Disease Control and Prevention (CDC), and JDRF. This work is supported, in part, by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR002535). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SUPPLEMENTARY MATERIAL

Supplement to "Joint modeling of multistate and nonparametric multivariate longitudinal data" (DOI: 10.1214/24-AOAS1889SUPPA; .pdf). The supplement provides additional details on model estimation, numerical computation, simulation studies, and TEDDY data application.

Code and software (DOI: 10.1214/24-AOAS1889SUPPB; .zip). The R package "jointmultistate" was created to implement the proposed method as a part of the supplementary materials. The package and code are also available for download from the GitHub repository located at https://github.com/luyouepiusf/jointmultistate.

REFERENCES

ALAFCHI, B., MAHJUB, H., TAPAK, L., ROSHANAEI, G. and AMIRZARGAR, M. A. (2021). Two-stage joint model for multivariate longitudinal and multistate processes, with application to renal transplantation data. *J. Probab. Stat.* Art. ID 6641602. MR4244007 https://doi.org/10.1155/2021/6641602

ALBERT, P. S. (2019). Shared random parameter models: A legacy of the biostatistics program at the National Heart, Lung and Blood Institute. *Stat. Med.* **38** 501–511. MR3902594 https://doi.org/10.1002/sim.8011

ATKINSON, M. A., EISENBARTH, G. S. and MICHELS, A. W. (2014). Type 1 diabetes. *Lancet* **383** 69–82.

BOOTH, J. G. and HOBERT, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 265–285.

BROWN, E. R., IBRAHIM, J. G. and DEGRUTTOLA, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61** 64–73. MR2129202 https://doi.org/10.1111/j.0006-341X.2005.030929.x

BRUMBACK, B. A. and RICE, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Amer. Statist. Assoc.* **93** 961–994. With comments and a rejoinder by the authors. MR1649194 https://doi.org/10.2307/2669837

CHI, Y.-Y. and IBRAHIM, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* **62** 432–445. MR2227491 https://doi.org/10.1111/j.1541-0420.2005.00448.x

DE BOOR, C. (1978). *A Practical Guide to Splines. Applied Mathematical Sciences* **27**. Springer, New York-Berlin. MR0507062

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. With discussion. MR0501537

FERRER, L., RONDEAU, V., DIGNAM, J., PICKLES, T., JACQMIN-GADDA, H. and PROUST-LIMA, C. (2016). Joint modelling of longitudinal and multi-state processes: Application to clinical progressions in prostate cancer. *Stat. Med.* **35** 3933–3948. MR3545618 https://doi.org/10.1002/sim.6972

GRUTTOLA, V. D. and TU, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50** 1003–1014.

HSIEH, F., TSENG, Y.-K. and WANG, J.-L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **62** 1037–1043. MR2297674 https://doi.org/10.1111/j.1541-0420.2006.00570.x

HUANG, H., LI, Y. and GUAN, Y. (2014). Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *J. Amer. Statist. Assoc.* **109** 1412–1424. MR3293600 https://doi.org/10.1080/01621459.2014.957286

HUANG, X., LI, G. and ELASHOFF, R. M. (2010). A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Stat. Interface* **3** 185–195. MR2659510 https://doi.org/10.4310/SII.2010.v3.n2.a6

JOY, C., BOYLE, P. P. and TAN, K. S. (1996). Quasi-Monte Carlo methods in numerical finance. *Manage. Sci.* **42** 926–938.

KRISCHER, J. P., LYNCH, K. F., SCHATZ, D. A., ILONEN, J., LERNMARK, Å., HAGOPIAN, W. A., REWERS, M. J., SHE, J.-X., SIMELL, O. G. et al. (2015). The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: The TEDDY study. *Diabetologia* **58** 980–987.

LI, G., LESPERANCE, M. and WU, Z. (2022). Joint modeling of multivariate survival data with an application to retirement. *Sociol. Methods Res.* **51** 1920–1946. MR4516875 https://doi.org/10.1177/0049124120914928

LI, N., ELASHOFF, R. M., LI, G. and TSENG, C.-H. (2012). Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Stat. Med.* **31** 1707–1721. MR2947519 https://doi.org/10.1002/sim.4507

LIU, L., HUANG, X. and O'QUIGLEY, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64** 950–958. MR2526647 https://doi.org/10.1111/j.1541-0420.2007.00954.x

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. With comments and a rejoinder by the authors. MR1803168 https://doi.org/10.2307/2669386

NIEDERREITER, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.* **84** 957–1041. MR0508447 https://doi.org/10.1090/S0002-9904-1978-14532-7

PAK, D., LI, C., TODEM, D. and SOHN, W. (2017). A multistate model for correlated interval-censored life history data in caries research. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 413–423. MR3611694 https://doi.org/10.1111/rssc.12186

PERPEROGLOU, A., SAUERBREI, W., ABRAHAMOWICZ, M. and SCHMID, M. (2019). A review of spline function procedures in R. *BMC Med. Res. Methodol.* **19** 1–16.

RIZOPOULOS, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67** 819–829. MR2829256 https://doi.org/10.1111/j.1541-0420.2010.01546.x

RIZOPOULOS, D. and GHOSH, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat. Med.* **30** 1366–1380. MR2828959 https://doi.org/10.1002/sim.4205

SALAMI, F., TAMURA, R., YOU, L., LERNMARK, Å., LARSSON, H. E., LUNDGREN, M., KRISCHER, J., ZIEGLER, A.-G., TOPPARI, J. et al. (2022). HbA1c as a time predictive biomarker for an additional islet autoantibody and type 1 diabetes in seroconverted TEDDY children. *Pediatric Diabetes* **23** 1586–1593.

SAYERS, A., HERON, J., SMITH, A. D. A. C., MACDONALD-WALLIS, C., GILTHORPE, M. S., STEELE, F. and TILLING, K. (2017). Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. *Stat. Methods Med. Res.* **26** 437–452. MR3592734 https://doi.org/10.1177/0962280214548822

SCHLUCHTER, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Stat. Med.* **11** 1861–1870. https://doi.org/10.1002/sim.4780111408

SLOAN, I. H. and WOŹNIAKOWSKI, H. (1998). When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals? *J. Complexity* **14** 1–33. MR1617765 https://doi.org/10.1006/jcom.1997.0463

SOBOL, I. M. (1976). Uniformly distributed sequences with an additional property of uniformity. *USSR Comput. Math. Math. Phys.* **16** 236–242.

TSIATIS, A. A. and DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statist. Sinica* **14** 809–834. MR2087974

WAND, M. P. (2000). A comparison of regression spline smoothing procedures. *Comput. Statist.* **15** 443–462. MR1818029 https://doi.org/10.1007/s001800000047

WILLIAMSON, P. R., KOLAMUNNAGE-DONA, R., PHILIPSON, P. and MARSON, A. G. (2008). Joint modelling of longitudinal and competing risks data. *Stat. Med.* **27** 6426–6438. MR2655125 https://doi.org/10.1002/sim.3451

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339. MR1450186 https://doi.org/10.2307/2533118

XU, J. and ZEGER, S. L. (2001). The evaluation of multiple surrogate endpoints. *Biometrics* **57** 81–87. MR1833292 https://doi.org/10.1111/j.0006-341X.2001.00081.x

YANG, L., YU, M. and GAO, S. (2016). Joint models for multiple longitudinal processes and time-to-event outcome. *J. Stat. Comput. Simul.* **86** 3682–3700. MR3547952 https://doi.org/10.1080/00949655.2016.1181760

YAO, F. (2007). Functional principal component analysis for longitudinal and survival data. *Statist. Sinica* **17** 965–983. MR2408647

YE, J., LI, Y. and GUAN, Y. (2015). Joint modeling of longitudinal drug using pattern and time to first relapse in cocaine dependence treatment data. *Ann. Appl. Stat.* **9** 1621–1642. MR3418738 https://doi.org/10.1214/15-AOAS852

YIU, S. and TOM, B. (2017). A joint modelling approach for multistate processes subject to resolution and under intermittent observations. *Stat. Med.* **36** 496–508. MR3592175 https://doi.org/10.1002/sim.7149

YOU, L. and QIU, P. (2021). Joint modeling of multivariate nonparametric longitudinal data and survival data: A local smoothing approach. *Stat. Med.* **40** 6689–6706. MR4352762 https://doi.org/10.1002/sim.9206

YOU, L., SALAMI, F., TÖRN, C., LERNMARK, Å. and TAMURA, R. (2024). Supplement to "Joint modeling of multistate and nonparametric multivariate longitudinal data." https://doi.org/10.1214/24-AOAS1889SUPPA, https://doi.org/10.1214/24-AOAS1889SUPPB

YU, L., ZHAO, Z. and STECK, A. K. (2017). T1D autoantibodies: Room for improvement? *Current Opinion in Endocrinology, Diabetes, and Obesity* **24** 285.

YUE, X. and AL KONTAR, R. (2021). Joint models for event prediction from time series and survival data. *Technometrics* **63** 477–486. MR4331448 https://doi.org/10.1080/00401706.2020.1832582

ZENG, D. and LIN, D. Y. (2021). Maximum likelihood estimation for semiparametric regression models with panel count data. *Biometrika* **108** 947–963. MR4341361 https://doi.org/10.1093/biomet/asaa091

ZHANG, H., KELVIN, E. A., CARPIO, A. and ALLEN HAUSER, W. (2020). A multistate joint model for interval-censored event-history data subject to within-unit clustering and informative missingness, with application to neurocysticercosis research. *Stat. Med.* **39** 3195–3206. MR4151928 https://doi.org/10.1002/sim.8663