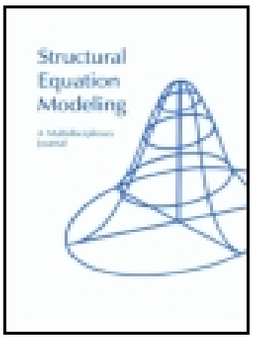

Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Causal Mediation Analysis for Multivariate Longitudinal Data and Survival Outcomes

Xiaoxiao Zhou & Xinyuan Song

To cite this article: Xiaoxiao Zhou & Xinyuan Song (2023): Causal Mediation Analysis for Multivariate Longitudinal Data and Survival Outcomes, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2022.2162406

To link to this article: https://doi.org/10.1080/10705511.2022.2162406

Published online: 24 Jan 2023.

Submit your article to this journal

Article views: 140

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hsem20

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2022.2162406

Routledge
Taylor & Francis Group

Check for updates

Causal Mediation Analysis for Multivariate Longitudinal Data and Survival Outcomes

Xiaoxiao Zhou and Xinyuan Song

Chinese University of Hong Kong

ABSTRACT

This study proposes a joint modeling approach to conduct causal mediation analysis that accommodates multivariate longitudinal data, dynamic latent mediator, and survival outcome. First, we introduce a confirmatory factor analysis model to characterize a time-varying latent mediator through multivariate longitudinal observable variables. Then, we establish a growth curve model to describe the linear trajectory of the dynamic latent mediator and simultaneously explore the relationship between the exposure and the mediating process. Finally, we link the mediating process to the survival outcome through a proportional hazards model. In addition, we use the mediation formula approach to assess the natural direct and indirect effects and prove the identifiability of the causal effects under sequential ignorability assumptions. A Bayesian approach incorporating the Markov chain Monte Carlo algorithm is developed to estimate the causal effects efficiently. Simulation studies are conducted to evaluate the empirical performance of the proposed method. An application to the Alzheimer's Disease Neuroimaging Initiative study further confirms the utility of the proposed method.

KEYWORDS

Causal mediation; dynamic latent variable; growth curve model; multivariate longitudinal data; survival outcome

1. Introduction

Examining a set of variables for mediation effects is a ubiquitous process in the social, behavioral, and health sciences owing to its extraordinary ability to identify and explain the mechanisms behind observations. When dealing with longitudinal data, growth curve modeling is a promising way to investigate individual differences in change over time and explore the predictors of these individual differences. It can be formulated in a multilevel modeling framework (Bryk & Raudenbush, 1992) or carried out in structural equation modeling (SEM) framework (Meredith & Tisak, 1990; Muthén & Curran, 1997). Cheong et al. (2003) proposed a method to evaluate a mediational process using latent growth curve (LGC) model by regarding the longitudinal mediator and outcome as two separate parallel processes. This method extended the mediation analysis to accommodate the longitudinal process. Afterward, Cheong (2011) investigated the accuracy of estimates and statistical power for testing meditation in latent growth curve modeling and provided guidelines for improving estimation accuracy. von Soest and Hagtvet (2011) proposed several longitudinal mediation models in the context of the LGC model and provided detailed explanations for constructing such models. Although other choices, such as the cross-lagged panel model, are also available to conduct longitudinal mediation analysis, Selig and Preacher (2009) acknowledged the advantages of the LGC model when one or more variables exhibit a meaningful trajectory of change. Nevertheless, the preceding analyses in this direction were mainly based on the normality assumption of the two processes and inapplicable to time-to-event outcomes.

The existing longitudinal mediation analyses incorporating the LGC model rely on a succession of linear structural equation models (SEMs) and assess the causal effects through a difference-in-coefficients approach or a product-of-coefficients method. One common criticism of such traditional methods is that they can provide only a black-box view of causality. Imai et al. (2010) proposed an alternative approach to overcome this drawback. Their proposed causal mediation analysis is based on the counterfactual framework without reference to any specific statistical model, thereby accommodating linear or nonlinear and parametric or nonparametric models as well as various data types of mediators and outcome variables. These appealing features facilitate the definition and interpretation of mediated effects on a time-to-event outcome. Abundant literature has endeavored to examine causal effects on the scales of hazard, hazard ratio, and survival probability under additive hazards (AH), proportional hazards (PH), and transformation models (Huang & Cai, 2016; Lange & Hansen, 2011; VanderWeele, 2011). Furthermore, Huang and Yang (2017) and Cho and Huang (2019) considered multiple mediators and examined complex path-specific effects in causal mediation models with survival outcomes. Sun et al. (2021) developed a joint mediation model involving a latent mediator to reduce the dimension of mediators and simultaneously relax the highly restrictive unconfoundedness assumptions.

CONTACT Xinyuan Song xysong@sta.cuhk.edu.hk Department of Statistics, Chinese University of Hong Kong, Hong Kong, Hong Kong.
© 2023 Taylor & Francis Group, LLC

Recent mediation studies have begun to explore the causal mechanism of the joint modeling of longitudinal and survival data, aiming to understand how much of the total effect is through the longitudinal mediator. Zheng and van der Laan (2017) considered general longitudinal settings with time-varying mediators, exposures, and survival outcomes. Upon establishing identifiability and the corresponding statistical estimands, they derived the efficient influence curves and demonstrated the robustness properties. Lin et al. (2017) altered the definition of effects to interventional ones such that the effects of interest can be well defined and identified. Distinguished from Lin et al. (2017), Didelez (2019) proposed the interventional approach based on an extended graphical method, an attractive alternative for any causal mediation setting. Later, Aalen et al. (2020) adopted this method and combined the g-formula with the AH model and a sequential linear model for the mediator process to obtain simple and interpretable expressions for direct and indirect effects. Under a counterfactual framework, Zheng and Liu (2022) quantified direct and indirect effects using a joint modeling approach. Zeng et al. (2022) further employed a functional principal component analysis approach to estimate the mediator process and derived a g-computation formula to express the causal estimands using the model coefficients.

On the other hand, one common phenomenon in medical and psychological research settings is that a single observed variable cannot fully characterize a latent trait. Instead, this latent trait must be measured together by multiple observed indicators from different perspectives. The multiple indicators are typically highly correlated, and simultaneously introducing them into a mediation analysis would render the causal diagram bewildering. While naively assuming the parallel pattern of the mediators may lead to information loss, the hypothesis of causally ordered relationships is also unplausible as it is hard to determine their orders. The factor analysis technique is a useful tool for addressing the problem. Unfortunately, none of the above methods have considered mediation analysis with survival outcomes and dynamic latent mediators.

This study proposes a joint modeling approach to conduct causal mediation analysis to accommodate multivariate longitudinal data observed at irregular time points and survival outcomes. This model comprises three components. The first component is a confirmatory factor analysis (CFA) model to characterize the latent mediator with several highly correlated observable surrogates. The second one is a growth curve model to describe the trajectory of the latent mediator and explore the relationship between the exposure and the mediating process. Finally, the mediating process is linked to the survival outcome through a PH model. Under this machinery, we can take full advantage of the longitudinal information and avoid the curse of dimensionality. Furthermore, we define causal direct and indirect effects on a counterfactual framework and show the identifiability of these effects under sequential ignorability assumptions. We develop a fully Bayesian approach with an efficient Markov chain Monte Carlo (MCMC) algorithm to simultaneously estimate the unknown parameters and causal effects. As far as we know, this study is the first to conduct causal mediation analysis with multivariate longitudinal and survival data.

The remainder of this paper is organized as follows. Section 2 introduces the proposed model and discusses the associated model identifiability issue. Section 3 derives the effects from the mediation formulation and certain identifiability assumptions, and Section 4 elucidates a Bayesian approach for conducting statistical inference. Sections 5 and 6 assess the empirical performance and utility of the proposed method through simulation studies and an application to the Alzheimer's disease dataset, respectively. Section 7 concludes the paper. Technical details are provided in the Appendix.

2. Model Description

2.1. CFA model

For each subject, let $\mathbf{Y}_{ij} = \mathbf{Y}_i(t_{ij}) = (y_{i1}(t_{ij}), ..., y_{ip}(t_{ij}))^T$ denote the $p \times 1$ vector of multivariate observations and $M_{ij} = M_i(t_{ij})$ denote the dynamic latent mediator for subject i $(i = 1, ..., n)$ at time t_{ij} $(j = 1, ..., J_i)$, where n is the total number of subjects and J_i is the total number of visits for subject i. A CFA model for characterizing the relationship between the observed variables and latent factors is defined as follows:

$$\mathbf{Y}_{ij} = \mathbf{\Lambda} M_{ij} + \boldsymbol{\epsilon}_{ij}, \tag{1}$$

where $\mathbf{\Lambda}$ is the $p \times q$ factor loading matrix; $\boldsymbol{\epsilon}_{ij}$ is a $p \times 1$ vector of measurement errors independent of M_{ij}, and $\boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \mathbf{\Psi})$ with $\mathbf{\Psi} = \text{diag}(\psi_1, ..., \psi_p)$.

2.2. LGC model

Let Z_i be the exposure variable, $\mathbf{X}_i = (x_{i1}, ..., x_{il})^T$ be the $l \times 1$ vector of baseline covariates, and $M_i(t)$ be the trajectory value of the latent mediator for subject i at time t. A growth curve model to depict the trajectory of $M_i(t)$ is defined as follows:

$$\begin{aligned} M_i(t) &= I_{M_i} + S_{M_i} t, \\ I_{M_i} &= \gamma_0 + \gamma_1 Z_i + \boldsymbol{\gamma}_2^T \mathbf{X}_i + \epsilon_{Ii}, \\ S_{M_i} &= \beta_0 + \beta_1 Z_i + \boldsymbol{\beta}_2^T \mathbf{X}_i + \epsilon_{Si}, \end{aligned} \tag{2}$$

where ϵ_{Ii} and ϵ_{Si} are random residuals and $(\epsilon_{Ii}, \epsilon_{Si})^T \sim \boldsymbol{N}(\mathbf{0}, \mathbf{\Sigma})$. Henceforth, the dynamic heterogeneous effect of Z_i on $M_i(t)$ is jointly described by the random intercept I_{M_i} and random slope S_{M_i} through the linear trajectory model Equation (2).

2.3. PH Model

Denote T_i^* and C_i as the event and censoring times, respectively, $\delta_i = I(T_i^* < C_i)$ as the failure indicator, and $T_i = \min(T_i^*, C_i)$ as the observed time. With the independent

censoring assumption, we consider a PH model to investigate the direct effect of the exposure Z_i and the indirect effect implemented through the time-dependent mediator $M_i(t)$ on the hazards of interest $\lambda_i(t)$:

$$\lambda_i(t|Z_i, M_i(t), \mathbf{X}_i) = \lambda_0(t)\exp(\alpha_1 Z_i + \alpha_2 M_i(t) + \boldsymbol{\alpha}_3^T \mathbf{X}_i), \quad (3)$$

where $\lambda_i(t)$ is the hazard function and $\lambda_0(t)$ is an unknown baseline hazard function.

2.4. Model Identifiability

The proposed model is unidentifiable without imposing appropriate identification constraints. For example, for an arbitrary constant b, $\mathbf{Y}_{ij} = \boldsymbol{\Lambda} M_{ij} + \boldsymbol{\epsilon}_{ij} = \boldsymbol{\Lambda} b b^{-1} M_{ij} + \boldsymbol{\epsilon}_{ij} = \boldsymbol{\Lambda} b b^{-1}(I_{M_i} + S_{M_i} t_{ij}) + \boldsymbol{\epsilon}_{ij} = \boldsymbol{\Lambda}^*(I^*_{M_i} + S^*_{M_i} t_{ij}) + \boldsymbol{\epsilon}_{ij}$, where $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda} b, I^*_{M_i} = b^{-1} I_{M_i}$, and $S^*_{M_i} = b^{-1} S_{M_i}$ with $\mathrm{cov}(I_{M_i}, S_{M_i}) = b^{-2}\boldsymbol{\Sigma}$. We follow the common practice (Song & Lee, 2012) to fix the first factor loading at 1.

3. Causal Effects Under Potential Outcome Framework

We quantify the direct and indirect effects of the exposure variable on the survival outcome of interest under the potential outcome framework (Imbens & Rubin, 2015). Let $M^z(t)$ denote the potential time-dependent mediator at time point t had the exposure Z been set to z, $\mathbf{M}^z = \{M^z(t), t \in [0, \tau]\}$ denote the potential mediator process, where τ is the largest follow-up time, and $T(z, \mathbf{m})$ be the potential time-to-event outcome that would be observed had Z and $\mathbf{M}$ been set to z and $\mathbf{m}$, respectively. Denote $f(T(z, \mathbf{M}^{z'}))$ as the function of time-to-event outcome T, which is a counterfactual notation defined by nested potential outcome. We abbreviate $f(T(z, \mathbf{M}^{z'}))$ to $f_{z,z'}$ for convenience. Apart from the commonly used assumptions in causal inference, such as consistency assumption and stable unit treatment value assumption (SUTVA) (Imbens & Rubin, 2015), we make two more assumptions to guarantee the identifiability of $f_{z,z'}$.

Assumption 1 (Ignorability). There are no unmeasured confounders for the effect of exposure on the potential mediator process and the potential survival time conditional on the observed covariates:

$$\{T(z, \mathbf{m}), \mathbf{M}^z\} \perp Z|\mathbf{X} \quad \text{for any } \mathbf{m} \text{ and } z \in \{0, 1\}. \quad (4)$$

Assumption 2 (Sequential ignorability). There are no unmeasured confounders for the effect of the mediator process $\mathbf{M}$ on the potential survival time conditional on the exposure and the observed covariates:

$$\mathbf{M} \perp T(z, \mathbf{m})|Z, \ \mathbf{X} \quad \text{for any } \mathbf{m}. \quad (5)$$

In the preceding assumptions, $A \perp B|C$ stands for the independence of A and B conditional on C. As mentioned by the literature (Preacher, 2015; Zeng et al., 2021), this ignorability assumption is strong and difficult to test. Therefore, its plausibility is worth further investigating.

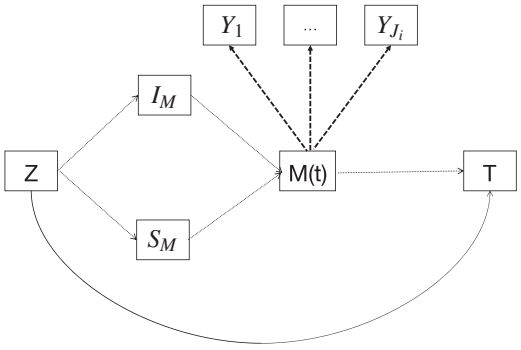

Figure 1. The path diagram of the proposed model. The direct effect is represented by solid line and the indirect effect is shown by dotted line.

Theorem 1. Given Assumptions 1 and 2, $f_{z,z'}$ is identifiable with the following mediation formula:

$$f_{z,z'} = \int f(T|z, \mathbf{w}) dF_{\mathbf{W}}(\mathbf{w}|z'), \quad (6)$$

where $\mathbf{W} = (I_M, S_M)$, $\mathbf{w} = (w_1, w_2)$ is the specific value of $\mathbf{W}$, and $F_{\mathbf{W}}(\mathbf{w}|z')$ represents the conditional distribution of $\mathbf{W}$ with the exposure z'.

Thus, the total and natural direct and indirect effects of exposure Z on the survival outcome can be calculated as follows:

$$\begin{aligned} TE(z, z') &= f_{z,z} - f_{z',z'} = \int f(T|z, \mathbf{w}) dF_{\mathbf{W}}(\mathbf{w}|z) \\ &\quad - \int f(T|z', \mathbf{w}) dF_{\mathbf{W}}(\mathbf{w}|z'), \\ DE(z, z') &= f_{z,z'} - f_{z',z'} = \int (f(T|z, \mathbf{w}) - f(T|z', \mathbf{w})) dF_{\mathbf{W}}(\mathbf{w}|z'), \\ IE(z, z') &= f_{z,z} - f_{z,z'} = \int f(T|z, \mathbf{w})(dF_{\mathbf{W}}(\mathbf{w}|z) - dF_{\mathbf{W}}(\mathbf{w}|z')). \end{aligned} \quad (7)$$

The above integrals can be approximated using Monte Carlo integration. There are various choices for f, such as the hazard function $\lambda(t)$ (Sun et al., 2021), the survival probability function (Zeng et al., 2022; Zheng & Liu, 2022;), and the restricted mean survival function (Zhou & Song, 2021). Under these circumstances, all the effects are functions of t. A path diagram of the proposed model is depicted in Figure 1.

We take the survival probability $S(t) = \exp(-\int_0^t \lambda(u)du)$ as an example to illustrate the computation. The counterfactual notation $S_{z,z'}(t)$ is a function of t and can be expressed as

$$\begin{aligned} S_{z,z'}(t) = \int \exp\Big(-\int_0^t \lambda_0(s) \exp(\alpha_1 z + \alpha_2(w_1 + w_2 s) \\ + \boldsymbol{\alpha}_3^T \mathbf{X}) ds\Big) dF_{\mathbf{W}}(\mathbf{w}|z'). \end{aligned} \quad (8)$$

For simplicity, we denote $DE(t) = S_{1,0}(t) - S_{0,0}(t)$, $IE(t) = S_{1,1}(t) - S_{1,0}(t)$ and $TE(t) = S_{1,1}(t) - S_{0,0}(t) = DE(t) + IE(t)$.

4. Bayesian Estimation

The derived $S_{z,z'}(t)$ is a function of regression parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \boldsymbol{\alpha}_3^T)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \boldsymbol{\beta}_2^T)^T$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \boldsymbol{\gamma}_2^T)^T$, $\lambda_0(t)$, and

$\boldsymbol{\Sigma}$. Denote $\boldsymbol{\theta}$ as the set of all unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$. In the Bayesian framework, the posterior distribution of $\boldsymbol{\theta}$, $\lambda_0(t)$, and the latent variables $\mathbf{w}$ can be obtained through posterior sampling. Then, the Monte Carlo estimates of $S_{z,z'}(t)$ can be obtained accordingly.

4.1. Prior Specification

For the unknown baseline hazard function $\lambda_0(t)$, we assume a piecewise exponential distribution (Ibrahim et al., 2001). Let $0 = u_0 < u_1 < \cdots < u_G$ be a finite partition of the time axis with $u_G > T_i$ for all $i = 1, \ldots, n$. In the gth interval, let $\lambda_0(t) = \lambda_g$ for $t \in (u_{g-1}, u_g]$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_G)^T$, and ν_{ig} denote the failure indicator in the gth interval $(u_{g-1}, u_g]$, such that $\nu_{ig} = 1$ if $T_i \in (u_{g-1}, u_g]$.

We follow the common practice (Ibrahim et al., 2001; Song & Lee, 2012) to assign the following conjugate prior distributions to the parameters in models Equations (1)–(3):

$$\begin{gathered}\Lambda_k|\psi_k \sim N(\Lambda_{k0}, \psi_k\sigma_k), \quad \boldsymbol{\gamma} \sim N(\boldsymbol{\gamma}_0, \mathbf{H}_\gamma),\\ \boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{H}_\beta), \quad \boldsymbol{\alpha} \sim N(\boldsymbol{\alpha}_0, \mathbf{H}_\alpha), \quad \psi_k \sim IG(c_{k1}, c_{k2}),\\ \lambda_g \sim Gamma(d_{g1}, d_{g2}), \quad \boldsymbol{\Sigma} \sim IW(\mathbf{R}_0, \rho_0), \end{gathered} \tag{9}$$

where $IG(\cdot,\cdot)$, $Gamma(\cdot,\cdot)$, and $IW(\cdot,\cdot)$ denote the inverse-gamma, gamma, and inverse-wishart distributions, respectively; Λ_k is the vector of unknown factor loadings in the kth row of $\boldsymbol{\Lambda}$; $\Lambda_{k0}, \boldsymbol{\gamma}_0, \boldsymbol{\beta}_0$, and $\boldsymbol{\alpha}_0$ are the hyperparameters representing the means of $\Lambda_k, \boldsymbol{\gamma}, \boldsymbol{\beta}$, and $\boldsymbol{\alpha}$, respectively; $\sigma_k, \mathbf{H}_\gamma$, $\mathbf{H}_\beta$, and $\mathbf{H}_\alpha$ are the hyperparameters representing the variance or covariance matrices of $\Lambda_k, \boldsymbol{\gamma}, \boldsymbol{\beta}$, and $\boldsymbol{\alpha}$, respectively; c_{k1}, c_{k2}, d_{g1}, and d_{g2} are shape and rate hyperparameters for ψ_k and λ_g, respectively; $\mathbf{R}_0$ and ρ_0 are hyperparameters for $\boldsymbol{\Sigma}$.

4.2. Posterior Distribution

Let $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{iJ_i}), \mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n), \mathbf{M}_i = (M_{i1}, \ldots, M_{iJ_i})^T$, $\mathbf{M} = (\mathbf{M}_1, \ldots, \mathbf{M}_n), \mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n), \mathbf{I} = (I_{M_1}, \ldots, I_{M_n})^T, \mathbf{S} = (S_{M_1}, \ldots, S_{M_n})^T, \mathbf{T} = (T_1, \ldots, T_n)^T, \mathbf{U} = (u_0, \ldots, u_G)^T, \mathbf{V}_i = (\nu_{i1}, \ldots, \nu_{iG})^T$, $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_n), \boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$, and $J = \sum_{i=1}^n J_i$. The complete-data likelihood function is derived as follows:

$$\begin{aligned} p(\mathbf{Y},\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta},\mathbf{I},\mathbf{S},\boldsymbol{\theta}|\mathbf{X}) = {} & p(\mathbf{Y}|\boldsymbol{\Lambda},\mathbf{M},\boldsymbol{\Psi})p(\mathbf{I},\mathbf{S}|\mathbf{X},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\Sigma})\\ & p(\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta}|\mathbf{X},\boldsymbol{\alpha},\boldsymbol{\lambda})p(\boldsymbol{\theta}), \end{aligned} \tag{10}$$

where

$$p(\mathbf{Y}|\boldsymbol{\Lambda},\mathbf{M},\boldsymbol{\Psi}) = (2\pi)^{-pJ/2}|\boldsymbol{\Psi}|^{-J/2}\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J_i}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})^T\boldsymbol{\Psi}^{-1}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})\right\},$$

$$p(\mathbf{I},\mathbf{S}|\mathbf{X},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\Sigma}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{w}_i-\mathbf{m}_{wi})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}_i-\mathbf{m}_{wi})\right\}$$

$$p(\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta}|\mathbf{X},\boldsymbol{\alpha},\boldsymbol{\lambda}) = \prod_{i=1}^{n}\prod_{g=1}^{G}\{\lambda_j\exp(m_{\lambda i})\}^{\nu_{ig}\delta_i}\exp\left[-\nu_{ig}\left\{\lambda_g(T_i-u_{g-1})+\sum_{l=1}^{g-1}\lambda_l(u_l-u_{l-1})\right\}\exp(m_{\lambda i})\right],$$

$m_{yij} = I_{M_i} + S_{M_i}t_{ij}, \mathbf{w}_i = (I_{M_i}, S_{M_i})^T, \mathbf{m}_{wi} = (\gamma_0 + \gamma_1 Z_i + \boldsymbol{\gamma}_2^T\mathbf{X}_i, \quad \beta_0 + \beta_1 Z_i + \boldsymbol{\beta}_2^T\mathbf{X}_i)^T$, and $m_{\lambda_i} = \alpha_1 Z_i + \alpha_2(I_{M_i} + S_{M_i} * T_i) + \boldsymbol{\alpha}_3^T\mathbf{X}_i$.

We use an MCMC algorithm that combines the data augmentation, Gibbs sampler, and the Metropolis-Hastings algorithm to implement the posterior sampling. The involved posterior distributions can be derived based on Equation (10) and the prior distributions specified in Section 4.1. The details are presented in the Appendix.

4.3. Estimation of TE(t), DE(t), and IE(t)

The procedure for the Bayesian estimates of $TE(t)$, $DE(t)$, and $IE(t)$ is outlined as follows:

- Randomly select starting values of $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, and $\mathbf{w}^{(0)}$.
- for $k = 1, \ldots, K_0, \ldots, K_0 + K$,
  a. Draw $\boldsymbol{\theta}^{(k)}$ from $p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X},\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$.
  b. Draw $\boldsymbol{\lambda}^{(k)}$ from $p(\boldsymbol{\lambda}|\mathbf{Y},\mathbf{X},\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta},\boldsymbol{\theta}^{(k)},\mathbf{w}^{(k-1)})$.
  c. Draw $\mathbf{w}^{(k)}$ from $p(\mathbf{w}|\mathbf{Y},\mathbf{X},\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta},\boldsymbol{\lambda}^{(k)},\boldsymbol{\theta}^{(k)})$.
  d. Calculate the plug-in estimate of $S_{z,z'}^{(k)}$ by Equation (8).
  e. Calculate $TE(t), DE(t)$, and $IE(t)$.
- Discard burn-in iterations after checking convergence, and compute the posterior mean and standard error estimates of $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, $TE(t)$, $DE(t)$, and $IE(t)$ using posterior samples.

A pseudo algorithm for obtaining the Bayesian estimates of $TE(t)$, $DE(t)$, and $IE(t)$ is given below:

Algorithm 1: MCMC algorithm for obtaining the Bayesian estimates of $TE(t)$, $DE(t)$, and $IE(t)$

Data: $\mathbf{O} = \{\mathbf{Y},\mathbf{X},\mathbf{T},\mathbf{U},\mathbf{V},\boldsymbol{\delta}\}$, K_0, K

1: Randomly initialize $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, $\mathbf{w}$: $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}, \boldsymbol{\lambda} = \boldsymbol{\lambda}^{(0)}, \mathbf{w} = \mathbf{w}^{(0)}$

2: **for** $k = 1$ to $K_0 + K$ **do**

3: Update $\boldsymbol{\theta}^{(k)}$ by sampling from $p(\boldsymbol{\theta}|\mathbf{O},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$, specifically,

(3a): Update $\boldsymbol{\alpha}^{(k)}$ by sampling from $p(\boldsymbol{\alpha}|\mathbf{O},\boldsymbol{\beta}^{(k-1)},\boldsymbol{\gamma}^{(k-1)},\boldsymbol{\Sigma}^{(k-1)},\boldsymbol{\Psi}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$

(3b): Update $\boldsymbol{\beta}^{(k)}$ by sampling from $p(\boldsymbol{\beta}|\mathbf{O},\boldsymbol{\alpha}^{(k-1)},\boldsymbol{\gamma}^{(k-1)},\boldsymbol{\Sigma}^{(k-1)},\boldsymbol{\Psi}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$

(3c): Update $\boldsymbol{\gamma}^{(k)}$ by sampling from $p(\boldsymbol{\gamma}|\mathbf{O},\boldsymbol{\alpha}^{(k-1)},\boldsymbol{\beta}^{(k-1)},\boldsymbol{\Sigma}^{(k-1)},\boldsymbol{\Psi}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$

(3d): Update $\boldsymbol{\Sigma}^{(k)}$ by sampling from $p(\boldsymbol{\Sigma}|\mathbf{O},\boldsymbol{\alpha}^{(k-1)},\boldsymbol{\beta}^{(k-1)},\boldsymbol{\gamma}^{(k-1)},\boldsymbol{\Psi}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$

(3e): Update $\boldsymbol{\Psi}^{(k)}$ by sampling from $p(\boldsymbol{\alpha}|\mathbf{O},\boldsymbol{\alpha}^{(k-1)},\boldsymbol{\beta}^{(k-1)},\boldsymbol{\gamma}^{(k-1)},\boldsymbol{\Sigma}^{(k-1)},\boldsymbol{\lambda}^{(k-1)},\mathbf{w}^{(k-1)})$

Table 1. Bayesian estimates of parameters in simulation study under censoring rate = 30%.

	$\lambda_0(t) = 0.5$				$\lambda_0(t) = t + 1$			
	$N = 500$		$N = 1000$		$N = 500$		$N = 1000$	
CFA model (1)								
Par	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
λ_2	−0.001	0.013	−0.001	0.009	−0.001	0.013	−0.000	0.009
λ_3	−0.001	0.013	−0.001	0.009	−0.001	0.013	−0.001	0.009
ψ_1	0.000	0.019	0.000	0.014	−0.000	0.019	0.000	0.014
ψ_2	−0.000	0.016	0.000	0.012	−0.000	0.015	0.000	0.012
ψ_3	−0.000	0.017	0.000	0.011	−0.000	0.017	0.000	0.012
Latent growth curve model (2)								
Par	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
γ_0	−0.008	0.058	0.001	0.039	−0.005	0.059	0.001	0.039
γ_1	0.006	0.091	0.000	0.065	−0.006	0.097	0.004	0.066
γ_2	0.001	0.045	0.002	0.031	−0.001	0.043	0.002	0.031
γ_3	0.008	0.075	0.004	0.052	0.002	0.072	0.004	0.051
β_0	−0.010	0.058	0.001	0.037	−0.007	0.060	0.001	0.037
β_1	0.002	0.096	−0.001	0.062	−0.007	0.097	−0.000	0.065
β_2	0.003	0.045	0.006	0.032	0.007	0.043	0.000	0.033
β_3	0.000	0.073	0.007	0.053	−0.006	0.074	0.007	0.052
σ_{11}	0.017	0.050	0.008	0.034	0.016	0.050	0.008	0.034
σ_{12}	−0.004	0.040	−0.000	0.026	−0.004	0.040	−0.000	0.026
σ_{22}	0.013	0.050	0.007	0.034	0.013	0.049	0.007	0.034
PH model (3)								
Par	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
α_1	−0.013	0.164	0.001	0.118	−0.046	0.169	0.001	0.118
α_2	−0.001	0.098	0.001	0.073	−0.029	0.103	0.001	0.073
α_3	0.011	0.148	−0.009	0.101	−0.020	0.150	−0.009	0.101
α_4	0.012	0.099	−0.002	0.074	−0.013	0.097	−0.002	0.074
Causal effects								
Par	BIAS	RMSE	BIAS	RMSE	Par	BIAS	RMSE	BIAS	RMSE
TE(0.6)	0.002	0.027	0.000	0.019	*TE*(0.4)	−0.005	0.031	0.001	0.020
DE(0.6)	0.003	0.025	0.000	0.018	*DE*(0.4)	0.001	0.028	0.006	0.020
IE(0.6)	−0.001	0.017	0.000	0.012	*IE*(0.4)	−0.006	0.020	−0.005	0.014
TE(1.2)	0.001	0.023	0.000	0.016	*TE*(0.8)	−0.001	0.028	−0.004	0.019
DE(1.2)	0.003	0.021	0.000	0.015	*DE*(0.8)	0.005	0.024	0.001	0.017
IE(1.2)	−0.002	0.016	0.000	0.011	*IE*(0.8)	−0.006	0.019	−0.005	0.013
TE(1.8)	0.001	0.021	0.000	0.014	*TE*(1.2)	0.005	0.025	0.002	0.016
DE(1.8)	0.002	0.018	0.000	0.013	*DE*(1.2)	0.008	0.022	0.004	0.015
IE(1.8)	−0.001	0.015	0.000	0.010	*IE*(1.2)	−0.003	0.017	−0.002	0.011

BIAS: bias; CFA: confirmatory factor analysis; Par: parameter; PH: proportional hazards; RMSE: root mean square error.

4: Update $\boldsymbol{\lambda}^{(k)}$ by sampling from $p(\boldsymbol{\lambda}|\mathbf{O}, \boldsymbol{\theta}^{(k-1)}, \mathbf{w}^{(k-1)})$
5: Update $\mathbf{w}^{(k)}$ by sampling from $p(\mathbf{w}|\mathbf{O}, \boldsymbol{\theta}^{(k-1)}, \boldsymbol{\lambda}^{(k-1)})$
6: Calculate the plug-in estimate of $S^{(k)}_{z,z'}$ by Equation (8)
7: Calculate $TE(t), DE(t)$, and $IE(t)$
8: **end for**
9: Compute the posterior mean and standard error estimates of $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, $TE(t)$, $DE(t)$, and $IE(t)$ using posterior samples from iterations K_0 to $K_0 + K$.

5. Simulation Study

In this section, we conduct simulation studies to evaluate the empirical performance of the proposed method. The unequally spaced time points t_{ij} are simulated such that the mean for each subject is zero, and increments $t_{ij} - t_{ij-1}$ are independently draws from $U(0, 1]$, where $U(\cdot, \cdot)$ denotes the uniform distribution. The time variable t_{ij} is then standardized to have unit variance. The exposure variable Z_i is randomly generated from a Bernoulli distribution with a success probability of 0.4. The baseline covariates $\mathbf{X}_i = (x_{i1}, x_{i2})^T$, where x_{i1} and x_{i2} are independently generated from $N(0, 1)$ and $U(-1, 1)$, respectively. Next, the random intercept and slope (I_{M_i}, S_{M_i}) are generated based on Equation (2) with $\gamma = (0.5, 0.5, 0.5, 0.5)^T$, $\boldsymbol{\beta} = (0.5, 0.5, 0.5, 0.5)^T$, and $\boldsymbol{\Sigma} = \begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{bmatrix}$. The latent mediator M_{ij} on the time point t_{ij} is calculated from the trajectory model Equation (2). In the CFA model, we set $p = 3$, $\boldsymbol{\Lambda}^T = [1, 0.6, 0.6]$, and $\boldsymbol{\Psi} = \text{diag}(0.36, 0.36, 0.36)$ to obtain the trivariate observations $\mathbf{Y}_{ij}$. As for the PH model, we set $\boldsymbol{\alpha} = (1, 1, 1, 1)^T$ and consider two types of baseline hazard functions, namely, (1) $\lambda_0(t) = 0.5$ and (2) $\lambda_0(t) = t/2 + 0.2$. The censoring time C_i is generated from $U(0, c)$ with an adjusted c to ensure a desired censoring rate (CR). To evaluate the finite-sample performance, we consider four scenarios with $(n, \text{CR}) = (500, 30\%), (500, 50\%), (1000, 30\%)$, and $(1000, 50\%)$ and generate 100 datasets under each scenario.

In conducting estimation, we adopt $G = 5$ time intervals with the cutpoints u_g being the $\frac{g}{G}th$ quantile of the empirical distribution of T_i (Sun et al., 2021). The hyperparameters involved in Equation (9) are assigned as follows (Prior I): $\Lambda_{k0} = 0$, $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_0 = \boldsymbol{\alpha}_0 = \mathbf{0}$, $\sigma_k = 1$, $\mathbf{H}_\gamma = \mathbf{H}_\beta = \mathbf{H}_\alpha = \mathbf{I}$, $c_{k1} = 9$, $c_{k2} = 4$, $d_{g1} = 0.1$, $d_{g2} = 0.2$, $\mathbf{R}_0 = 6\mathbf{I}$, and $\rho_0 = 5$, where $\mathbf{0}$ and $\mathbf{I}$ are the zero vector and identity matrix of appropriate dimensions, respectively. We conduct three test runs starting from different initial values to check the convergence of the MCMC algorithm and find that the three parallel chains mix rapidly within 2,000 iterations. Therefore, we discard

Table 2. Bayesian estimates of parameters in simulation study under censoring rate = 50%.

	$\lambda_0(t) = 0.5$				$\lambda_0(t) = t + 1$			
	$N = 500$		$N = 1000$		$N = 500$		$N = 1000$	
CFA model (1)								
Par	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
λ_2	−0.001	0.013	−0.001	0.009	−0.001	0.013	−0.000	0.009
λ_3	−0.001	0.013	−0.001	0.009	−0.001	0.013	−0.000	0.009
ψ_1	0.000	0.019	0.000	0.014	0.000	0.019	0.000	0.014
ψ_2	−0.000	0.015	0.000	0.012	−0.000	0.015	0.000	0.012
ψ_3	−0.000	0.017	0.000	0.012	−0.000	0.017	0.000	0.011
Latent growth curve model (2)								
Par	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
γ_0	−0.014	0.058	0.001	0.040	−0.007	0.056	0.000	0.039
γ_1	0.014	0.090	0.000	0.063	0.008	0.093	0.002	0.062
γ_2	0.001	0.044	0.002	0.031	0.000	0.042	−0.000	0.030
γ_3	0.007	0.073	0.004	0.051	0.004	0.077	0.002	0.051
β_0	−0.015	0.061	0.001	0.039	−0.010	0.058	0.000	0.038
β_1	0.008	0.088	0.000	0.065	0.002	0.094	0.001	0.063
β_2	0.002	0.045	0.001	0.033	0.002	0.043	−0.000	0.032
β_3	−0.005	0.074	0.005	0.055	−0.003	0.076	0.006	0.052
σ_{11}	0.016	0.050	0.008	0.034	0.017	0.050	0.008	0.034
σ_{12}	−0.004	0.040	−0.000	0.026	−0.004	0.040	−0.000	0.026
σ_{22}	0.013	0.049	0.007	0.034	0.013	0.049	0.007	0.034
PH model (3)								
Par	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
α_1	−0.015	0.191	−0.011	0.130	−0.037	0.190	−0.027	0.128
α_2	−0.002	0.114	−0.000	0.086	−0.018	0.110	−0.014	0.084
α_3	0.007	0.168	−0.015	0.122	−0.012	0.166	−0.026	0.119
α_4	0.006	0.126	−0.004	0.098	0.002	0.116	−0.005	0.090
Causal effects								
Par	BIAS	RMSE	BIAS	RMSE	Par	BIAS	RMSE	BIAS	RMSE
$TE(0.3)$	0.001	0.034	0.002	0.022	$TE(0.3)$	0.014	0.036	0.011	0.024
$DE(0.3)$	0.003	0.032	0.002	0.021	$DE(0.3)$	0.017	0.033	0.014	0.023
$IE(0.3)$	−0.002	0.019	0.000	0.013	$IE(0.3)$	−0.003	0.019	−0.003	0.013
$TE(0.6)$	0.001	0.031	0.001	0.020	$TE(0.6)$	−0.003	0.033	−0.005	0.021
$DE(0.6)$	0.003	0.030	0.001	0.020	$DE(0.6)$	0.005	0.030	0.002	0.020
$IE(0.6)$	−0.002	0.018	0.000	0.011	$IE(0.6)$	−0.008	0.021	−0.007	0.015
$TE(0.9)$	0.001	0.029	0.001	0.013	$TE(0.9)$	−0.001	0.030	−0.003	0.018
$DE(0.9)$	0.003	0.028	0.001	0.019	$DE(0.9)$	0.006	0.027	0.003	0.018
$IE(0.9)$	−0.002	0.017	0.000	0.012	$IE(0.9)$	−0.007	0.019	−0.006	0.014

BIAS: bias; CFA: confirmatory factor analysis; Par: parameter; PH: proportional hazards; RMSE: root mean square error.

2,000 burn-in iterations and collect the subsequent 2,000 posterior samples to conduct posterior inference. Tables 1 and 2 summarize the results of parameter estimation on the basis of 100 replications. The causal effects on the survival probability are functions of t. We calculate their values on 20 time points equally distributed from zero to the maximum observed time. The bias (BIAS) and root mean square error (RMSE) are similar, and thus, we only present the results on three time points in the lower panel of Tables 1 and 2. Overall, the BIAS and RMSE for both parameters and causal effects are close to zero, verifying the satisfactory performance of parameter estimation and the extraordinary ability of the mediation formula approach in estimating the counterfactual outcome. In addition, the performance is improved as either the sample size increases from $N = 500$ to $N = 1000$ or CR decreases from 50% to 30%.

We further investigate the sensitivity of Bayesian results to the prior specification and the normality assumption of the random intercept I_{M_i} and random slope S_{M_i}. We disturb the hyperparameters as follows (Prior II): $\Lambda_{k0} = 2$, $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_0 = \boldsymbol{\alpha}_0 = 2\mathbf{I}$, $\sigma_k = 100$, $\mathbf{H}_\gamma = \mathbf{H}_\beta = \mathbf{H}_\alpha = 100\mathbf{I}$, $c_{k1} = 7$, $c_{k2} = 3$, $d_{g1} = 0.1$, $d_{g2} = 0.2$, $\mathbf{R}_0 = 11\mathbf{I}$, and $\rho_0 = 8$. The results are similar and not reported. The simulation for assessing the normality assumption is conducted under $(n, \text{CR}) = (500, 30\%)$ and $\lambda_0(t) = 0.5$. In this case, the simulation setup is the same as before, except that ϵ_{Ii} and ϵ_{Si} in Equation (2) follow non-normal distributions as follows: (1) $\epsilon_{Ii} \sim Gamma(4, 5)$ and $\epsilon_{Si} \sim t_3$; (2) $\epsilon_{Ii} \sim t_3$ and $\epsilon_{Si} \sim Gamma(4, 5)$. The results reported in Table 3 confirm the robustness of the Bayesian estimation to the violation of the normality assumption.

The python code for the simulation study is on https://github.com/smallpolaris/mcml.

6. Application

Alzheimer's disease (AD) is a progressive and irreversible disease with dementia symptoms, such as memory and cognitive ability loss, gradually worsening over several years. Initially launched in 2004, the Alzheimer's Disease Neuroimaging Initiative (ADNI) is an ongoing longitudinal and clinical-pathologic study to define AD progression and has been extended, until now, to four phases: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. This study has recruited more than 2,000 subjects between the ages of 55 and 90 years and conducted a variety of imaging and clinical assessments upon their consent. Refer to the official website

Table 3. Bayesian estimates of parameters in sensitivity analysis with $(n, CR) = (500, 30\%)$.

$\lambda_0(t) = 0.5$

$\epsilon_{Ii} \sim$ Gamma$(4,5)$, $\epsilon_{Si} \sim t_3$				$\epsilon_{Ii} \sim t_3$, $\epsilon_{Si} \sim$ Gamma$(4,5)$	
CFA model (1)					
Par	BIAS	RMSE		BIAS	RMSE
λ_2	−0.001	0.013		−0.000	0.012
λ_3	−0.001	0.012		0.000	0.013
ψ_1	0.001	0.019		0.001	0.019
ψ_2	−0.000	0.016		−0.000	0.016
ψ_3	0.001	0.017		0.001	0.017
Latent growth curve model (2)					
Par	BIAS	RMSE		BIAS	RMSE
γ_0	−0.006	0.052		−0.002	0.048
γ_1	0.006	0.090		−0.002	0.087
γ_2	0.001	0.046		0.006	0.046
γ_3	−0.011	0.071		0.004	0.078
β_0	0.002	0.050		−0.001	0.050
β_1	−0.009	0.091		0.001	0.088
β_2	0.004	0.045		0.003	0.046
β_3	0.008	0.080		−0.005	0.071
σ_{11}	0.009	0.039		0.011	0.040
σ_{12}	–	–		–	–
σ_{22}	0.014	0.042		0.012	0.041
PH model (3)					
Par	BIAS	RMSE		BIAS	RMSE
α_1	−0.029	0.161		−0.025	0.160
α_2	−0.001	0.108		0.005	0.111
α_3	−0.010	0.154		−0.006	0.156
α_4	0.018	0.110		0.015	0.118
Causal effects					
Par	BIAS	RMSE	Par	BIAS	RMSE
$TE(0.6)$	0.004	0.025	$TE(0.6)$	0.005	0.025
$DE(0.6)$	0.006	0.026	$DE(0.6)$	0.006	0.026
$IE(0.6)$	−0.002	0.015	$IE(0.6)$	−0.001	0.016
$TE(1.2)$	0.003	0.021	$TE(1.2)$	0.003	0.021
$DE(1.2)$	0.005	0.021	$DE(1.2)$	0.004	0.022
$IE(1.2)$	−0.002	0.014	$IE(1.2)$	−0.001	0.015
$TE(1.8)$	0.003	0.019	$TE(1.8)$	0.003	0.018
$DE(1.8)$	0.004	0.019	$DE(1.8)$	0.003	0.018
$IE(1.8)$	−0.001	0.015	$IE(1.8)$	0.000	0.014

BIAS: bias; CFA: confirmatory factor analysis; Par: parameter; PH: proportional hazards; RMSE: root mean square error.

(http://adni.loni.usc.edu/) for more details about this study. Besides the genetic and biochemical biomarkers, multiple cognitive test scores are collected to measure participants' cognitive decline at baseline, 6 months, 12 months, 18 months, 24 months, and every 12 months thereafter. A series of previous studies have shown that the APOE-ϵ4 allele accounts for the vast majority of AD risk (Cramer et al., 2012; Raber et al., 2004). In addition, some researchers (Ali et al., 2018; Bretsky et al., 2003) have pointed out that the APOE-ϵ4 allele may function as a risk factor for cognitive impairment in normal aging across a broad spectrum of domains or exert detectable effects early in a long prodromal AD trajectory. Therefore, we aim to validate and quantify the direct effect of APOE-ϵ4 on the risk of AD and the indirect effect through the longitudinal progressive cognitive decline and then to AD.

In applying the proposed model to the ADNI data analysis, the exposure Z_i was defined as the presence of APOE-ϵ4. To illustrate the trend of cognitive decline comprehensively, three cognitive test scores—including Alzheimer Disease Assessment Scale-Cognitive (ADAS) 11 (ADAS11; $y_{i1}(t_{ij})$), ADAS13 ($y_{i2}(t_{ij})$), and the Mini-Mental State Examination (MMSE, $y_{i3}(t_{ij})$)—were grouped into a longitudinal latent mediator M_{ij}. These highly associated longitudinal scores reflect the cognitive ability on complementary aspects, with high ADAS scores and low MMSE scores implying poor cognitive ability. The failure time of AD (T_i) was the duration from the baseline to the date of the first AD diagnosis or the date of the last visit, whichever came first. Age (X_{i1}), gender (X_{i2}, 1 = female, 0 = male), years of education (X_{i3}), marital status (X_{i4}, 1 = married, 0 = otherwise), race (X_{i5}, 1 = White, 0 = otherwise), hippocampus volume (X_{i6}), and ventricle volume (X_{i7}) at baseline were considered as a minimum set of confounders ($\mathbf{X}_i$) in model Equations (2) and (3). All the continuous variables and the visiting time t_{ij} were standardized prior to analysis. Finally, 656 subjects with mild cognitive impairment (MCI) who had 3 to 14 follow-up visits with complete measurements were selected for the analysis. Among them, 241 subjects were diagnosed with AD during the follow-up period, leading to a CR of approximately 63.3%. The gap between the Kaplan-Meier curves of the two groups shown in Figure 2 strongly supports that APOE-ϵ4 is a risk factor for AD. We also randomly selected an individual from each group and plotted their longitudinal test scores in Figure 3. It is evident that the APOE-ϵ4 carrier has much higher scores in

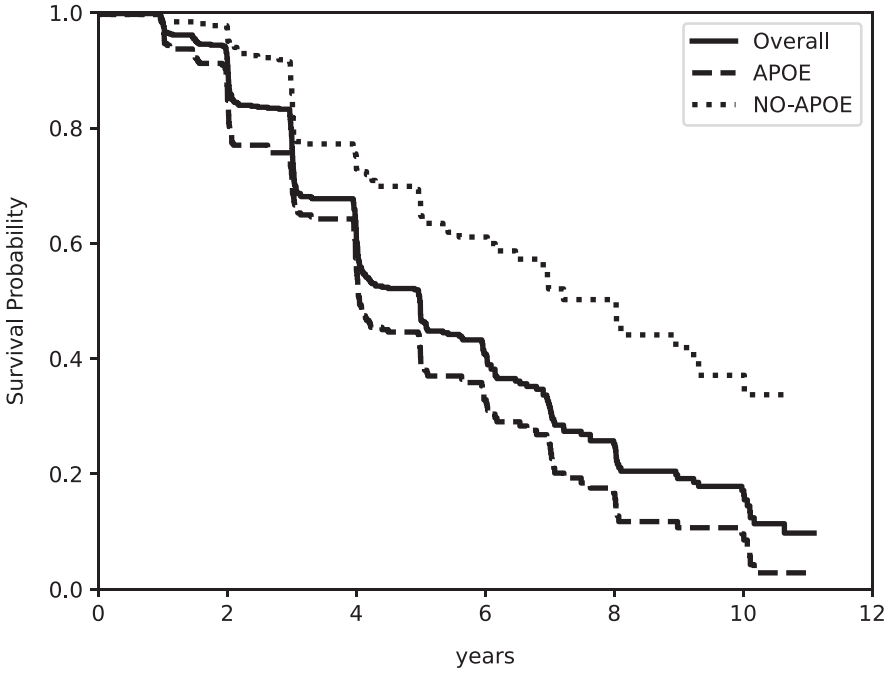

Figure 2. Kaplan-Meier curves for overall, APOE-ϵ4 carrier group, and APOE-ϵ4 noncarrier group.

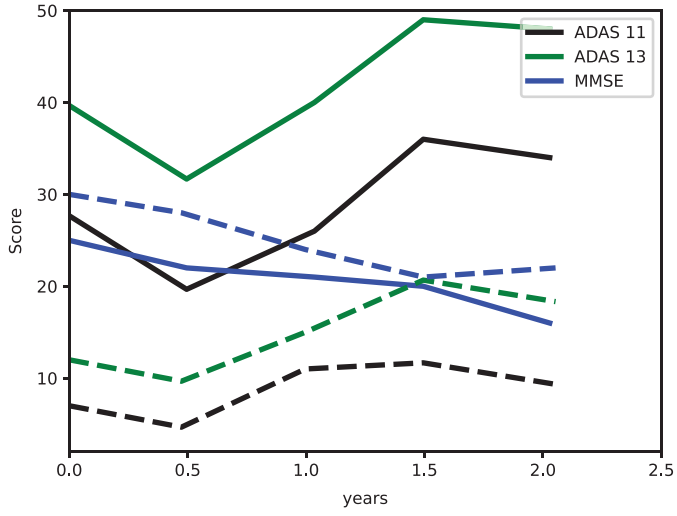

Figure 3. The trajectories of cognitive test scores of two individuals randomly selected from the APOE-ϵ4 carrier (solid line) and noncarrier (dashed line) groups.

ADAS11 and ADAS13 and lower score in MMSE than the APOE-ϵ4 noncarrier, manifesting the significant effect of carrying the APOE–ϵ4 allele on cognitive decline.

To implement the MCMC algorithm, we adopted Prior I in the simulation study and ran several MCMC chains with different initial values to check convergence. The trace plots presented in Figure 4 show that the algorithm converged within 20,000 iterations. After discarding 20,000 burn-in iterations, we collected the subsequent 10,000 posterior samples to obtain parameter estimates and calculate the causal effects on the survival probability of AD. Considering that nearly 90% of failure time lies below 6, we chose $t = \{1, 2, \ldots, 6\}$, covering several percentiles of multiple of 10. Table 4 reports the estimates of the parameters (upper panel) and causal effects (lower panel) together with their 95% credible interval (CI), constructed from 2.5% and 97.5% percentiles of the posterior samples.

First, the factor loading is positive for ADAS13 $[\hat{\lambda}_2 = 1.012(0.995, 1.028)]$ and negative for MMSE $[\hat{\lambda}_3 = -0.904(-0.926, -0.883)]$, implying that a high ADAS13 score or low MMSE score is associated with a high cognitive impairment. Thus, we interpret the dynamic latent mediator as "cognitive decline." Besides, carrying APOE-ϵ4 alleles positively affects the random intercept and slope in the growth curve model $[\hat{\gamma}_1 = 0.359(0.265, 0.431)$, $\hat{\beta}_1 = 0.170(0.132, 0.215)]$, suggesting that APOE-$\epsilon$4 allele carriers tend to have more severe cognitive impairment at baseline and faster progression of cognitive decline than APOE-ϵ4 allele noncarriers. Last, carrying APOE-ϵ4 alleles $[\hat{\alpha}_1 = 0.450(0.186, 0.748)]$ and cognitive decline $[\hat{\alpha}_2 = 0.161 (0.117, 0.206)]$ exhibit significantly positive effects on AD hazards. Likewise, all the causal effects of carrying APOE-ϵ4 alleles on the survival probability of AD are significantly negative, whatever the tested time t is. These findings confirm that APOE-ϵ4 allele carriers or those with rapid cognitive decline have a high AD incidence. Moreover, carrying APOE-ϵ4 alleles affects the survival outcome of interest through both a direct path or aggravated cognitive impairment. Take $t = 4$ as an example; when the exposure changes from APOE-ϵ4 noncarriers to APOE-ϵ4 carriers, the AD survival probability decreases by $0.115([-0.170, -0.062])$ in total with an estimated $0.086([-0.143, -0.035])$ directly affected by carrying APOE-ϵ4 alleles and an estimated $0.029([-0.039, -0.020])$ through cognitive decline. Furthermore, the magnitudes of all the effects increased with time, revealing the enhanced

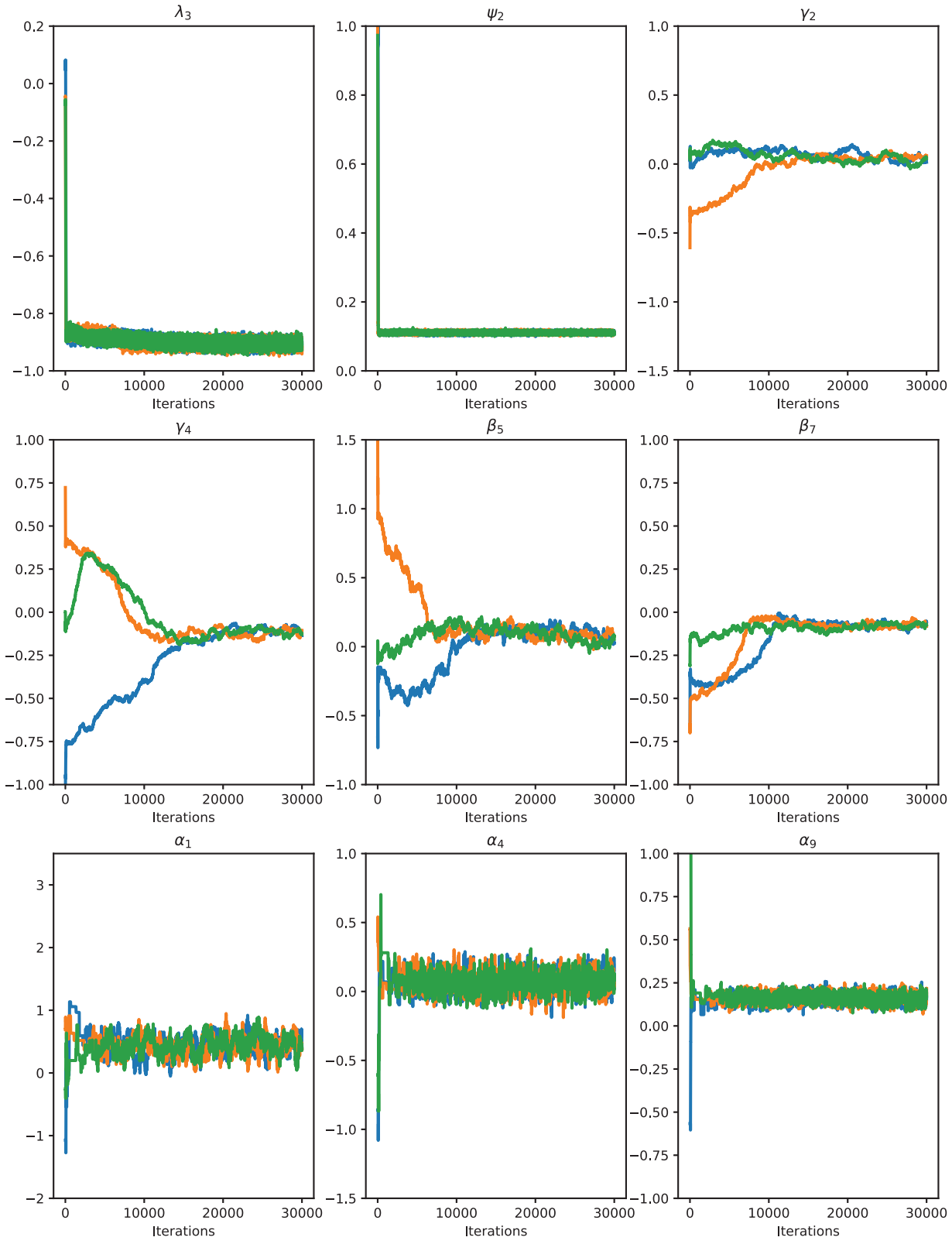

Figure 4. Trace plots of some parameters in Alzheimer's Disease Neuroimaging Initiative study.

Table 4. Bayesian estimates of parameters in ADNI study.

CFA model (1)

Par	Est	95% CI	Par	Est	95% CI	Par	Est	95% CI
λ_2	1.012	(0.995, 1.028)	λ_3	−0.904	(−0.926, −0.883)	ψ_1	0.140	(0.134, 0.146)
ψ_2	0.111	(0.106, 0.116)	ψ_3	0.290	(0.277, 0.303)			

Latent growth curve model (2)

Par	Est	95% CI	Par	Est	95% CI	Par	Est	95% CI
γ_0	−0.431	(−0.599, −0.117)	γ_1	0.359	(0.265, 0.431)	γ_2	0.048	(0.002, 0.126)
γ_3	0.174	(0.102, 0.233)	γ_4	−0.101	(−0.129, −0.070)	γ_5	0.274	(0.092, 0.384)
γ_6	−0.047	(−0.221, 0.076)	γ_7	−0.269	(−0.290, −0.237)	γ_8	0.093	(0.020, 0.166)
β_0	0.023	(−0.094, 0.155)	β_1	0.170	(0.132, 0.215)	β_2	−0.011	(−0.045, 0.026)
β_3	0.112	(0.055, 0.159)	β_4	0.001	(−0.030, 0.029)	β_5	0.109	(0.039, 0.175)
β_6	0.010	(−0.110, 0.123)	β_7	−0.078	(−0.106, −0.051)	β_8	0.045	(0.013, 0.081)
σ_{11}	0.460	(0.411, 0.514)	σ_{12}	0.175	(0.151, 0.200)	σ_{22}	0.138	(0.123, 0.156)

PH model (3)

Par	Est	95% CI	Par	Est	95% CI	Par	Est	95% CI
α_1	0.450	(0.186, 0.748)	α_2	0.161	(0.117, 0.206)	α_3	0.041	(−0.095, 0.196)
α_4	0.025	(−0.336, 0.304)	α_5	0.073	(−0.040, 0.191)	α_6	−0.080	(−0.493, 0.228)
α_7	0.166	(−0.287, 0.589)	α_8	−0.250	(−0.380, −0.109)	α_9	0.270	(0.140, 0.388)

Causal effects on survival probability

Par	Est	95% CI	Par	Est	95% CI	Par	Est	95% CI
TE(1)	−0.018	(−0.029, −0.008)	*DE*(1)	−0.015	(−0.025, −0.005)	*IE*(1)	−0.003	(−0.004, −0.001)
TE(2)	−0.037	(−0.057, −0.017)	*DE*(2)	−0.029	(−0.049, −0.011)	*IE*(2)	−0.008	(−0.010, −0.004)
TE(3)	−0.083	(−0.124, −0.043)	*DE*(3)	−0.063	(−0.105, −0.025)	*IE*(3)	−0.020	(−0.026, −0.013)
TE(4)	−0.115	(−0.170, −0.062)	*DE*(4)	−0.086	(−0.143, −0.035)	*IE*(4)	−0.029	(−0.039, −0.020)
TE(5)	−0.138	(−0.202, −0.075)	*DE*(5)	−0.101	(−0.168, −0.040)	*IE*(5)	−0.037	(−0.049, −0.026)
TE(6)	−0.159	(−0.231, −0.088)	*DE*(6)	−0.114	(−0.190, −0.045)	*IE*(6)	−0.045	(−0.059, −0.032)

BIAS: bias; CFA: confirmatory factor analysis; CI: credible interval; Est: estimate; Par: parameter; PH: proportional hazards; RMSE: root mean square error.

detrimental influence of the APOE-$\epsilon 4$ allele on the incidence of AD.

We also conducted the sensitivity analysis regarding the number of partition intervals G and prior inputs. The results obtained under $G = 10$ and Prior II are similar and not reported.

7. Conclusion

This study conducted Bayesian causal mediation analysis for multivariate longitudinal and survival data based on a joint model framework comprising a linear SEM and PH model. The SEM groups multivariate observations into a latent mediator and captures its trajectory, reducing the dimension of mediators and alleviating misinterpretation of the mediating mechanism. The PH model examines the effects of the exposure, dynamic mediator, and baseline covariates on the survival outcome of interest. Under the potential outcome framework, causal effects were defined to quantify the effects of exposure on the interested survival outcome through direct and indirect paths. A fully Bayesian approach was developed to conduct the estimation. Simulation studies and an application to the ADNI dataset demonstrated the utility of the proposed method.

Several issues merit further research. First, we can replace the PH model with other survival models in the proposed joint modeling framework, such as the accelerated failure time model and mean residual life model. Second, the conditional independence assumption between censoring time C and failure time T given mediator process $\mathbf{M}$ may be too restrictive to reflect the reality. Therefore, relaxing this assumption under the current joint model is of considerable interest. Third, it is unrealistic to exhaust confounders in the application. The appropriate sensitivity analysis techniques may be worth considering. Last, Didelez (2019) proposed an alternative definition of mediated effects, and investigating such effects is also of interest but challenging. The above extensions require substantial efforts in future research.

8. Appendix

8.1. Proof of Theorem 1

$$\begin{aligned} f_{z,z'} &= f(T(z,\mathbf{M}(z'))) = \int f(T(z,\mathbf{m}))dF_{\mathbf{M}(z')}(\mathbf{m}) \\ &= \int f(T(z,I_M=w_1,S_M=w_2))dF_{\mathbf{W}(z')}(w_1,w_2) \\ &= \int f(T(z,\mathbf{W}=\mathbf{w}))dF_{\mathbf{W}(z')}(\mathbf{w}) \\ &\quad [\text{By Assumptions 1 and 2}] \\ &= \int f(T(z,\mathbf{W}=\mathbf{w}|z,\mathbf{w}))dF_{\mathbf{W}(z')}(\mathbf{w}) \\ &\quad [\text{By consistency assumption}] \\ &= \int f(T|z,\mathbf{W}=\mathbf{w})dF_{\mathbf{W}(z')}(\mathbf{w}) \end{aligned} \tag{11}$$

The second line is due to model Equation (2), from which the mediator process $\mathbf{m}$ is uniquely determined by the pair (I_M, S_M).

8.2. Posterior Distribution

The posterior distributions of all parameters are as follows:

$$[\Lambda_k|\cdot] \sim N[\Lambda_k^*, \sigma_k^*], \quad [\psi_k|\cdot] \sim IG(c_{k1}^*, c_{k2}^*), \quad [\lambda_g|\cdot] \sim Gamma(d_{g1}^*, d_{g2}^*), \quad [\boldsymbol{\Sigma}|\cdot] \sim IW(\mathbf{R}^*, \rho^*),$$

$$p(\mathbf{w}_i|\cdot) \propto \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{J_i}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})^T\boldsymbol{\Psi}^{-1}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})-\frac{1}{2}(\mathbf{w}_i-\mathbf{m}_{wi})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}_i-\mathbf{m}_{wi})\right\}$$
$$\times\prod_{g=1}^{G}\{\lambda_j\exp(m_{\lambda i})\}^{\nu_{ig}\delta_i}\exp\left[-\nu_{ig}\left\{\lambda_g(T_i-u_{g-1})+\sum_{l=1}^{g-1}\lambda_l(u_l-u_{l-1})\right\}\exp(m_{\lambda i})\right],$$

$$p(\boldsymbol{\alpha}|\cdot) \propto \prod_{i=1}^{n}\prod_{g=1}^{G}\{\lambda_g\exp(m_{\lambda i})\}^{\nu_{ig}\delta_i}\exp\left[-\nu_{ig}\left\{\lambda_g(T_i-u_{g-1})+\sum_{l=1}^{g-1}\lambda_l(u_l-u_{l-1})\right\}\exp(m_{\lambda i})\right]$$
$$\times\exp\left(-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^T\mathbf{H}_\alpha^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)\right),$$

$$p(\boldsymbol{\gamma}) \propto \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J_i}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})^T\boldsymbol{\Psi}^{-1}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{w}_i-\mathbf{m}_{wi})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}_i-\mathbf{m}_{wi})\right.$$
$$\left.-\frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\mathbf{H}_\gamma^{-1}(\boldsymbol{\gamma}-\boldsymbol{\alpha}_0)\right\}\prod_{i=1}^{n}\prod_{g=1}^{G}\{\lambda_j\exp(m_{\lambda i})\}^{\nu_{ig}\delta_i}\exp\left[-\nu_{ig}\left\{\lambda_g(T_i-u_{g-1})+\sum_{l=1}^{g-1}\lambda_l(u_l-u_{l-1})\right\}\exp(m_{\lambda i})\right],$$

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J_i}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})^T\boldsymbol{\Psi}^{-1}(\mathbf{Y}_{ij}-\boldsymbol{\Lambda}m_{yij})-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{w}_i-\mathbf{m}_{wi})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}_i-\mathbf{m}_{wi})\right.$$
$$\left.-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\mathbf{H}_\beta^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\}\prod_{i=1}^{n}\prod_{g=1}^{G}\left\{\lambda_j\exp(m_{\lambda i})\right\}^{\nu_{ig}\delta_i}\exp\left[-\nu_{ig}\left\{\lambda_g(T_i-u_{g-1})+\sum_{l=1}^{g-1}\lambda_l(u_l-u_{l-1})\right\}\exp(m_{\lambda i})\right],$$

where $\sigma_k^*=(J\psi_k+\sigma_k^{-1})^{-1}$, $\Lambda_k^*=\sum_{i=1}^{n}\sum_{j=1}^{J_i}y_{ijk}(I_{M_i}+S_{M_i}t_{ij})+\sigma_k^{-*}\Lambda_{k0}$, $c_{k1}^*=c_{k1}+J/2$, $c_{k2}^*=c_{k2}+\frac{1}{2}(y_{ijk}-\Lambda_k m_{yij})^2$, $d_{g1}^*=\sum_{i=1}^{n}\nu_{ig}\delta_i+d_{g1}$, $d_{g2}^*=\sum_{i=1}^{n}\exp(m_{\lambda i})[\lambda_g(T_i-u_{g-1})+I(g<G)\sum_{l=g+1}^{G}\nu_{il}(u_l-u_{l-1})]+d_{g2}$, $\mathbf{R}^*=\mathbf{R}+\sum_{i=1}^{n}(\mathbf{w}_i-\mathbf{m}_{wi})^T(\mathbf{w}_i-\mathbf{m}_{wi})$, $\rho^*=\rho+n/2$.

Funding

This work was supported by the Research Grant Council, University Grant Committee (14302220, 14302519) of HKSAR.

ORCID

Xinyuan Song http://orcid.org/0000-0002-4877-3200

References

Aalen, O. O., Stensrud, M. J., Didelez, V., Daniel, R., Røysland, K., & Strohmaier, S. (2020). Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. *Biometrical Journal. Biometrische Zeitschrift*, *62*, 532–549. https://doi.org/10.1002/bimj.201800263

Ali, J. I., Smart, C. M., & Gawryluk, J. R. (2018). Subjective cognitive decline and APOE ε4: A systematic review. *Journal of Alzheimer's Disease : JAD*, *65*, 303–320. https://doi.org/10.3233/JAD-180248

Bretsky, P., Guralnik, J., Launer, L., Albert, M., & Seeman, T. (2003). The role of APOE-ε4 in longitudinal cognitive decline: MacArthur Studies of Successful Aging. *Neurology*, *60*, 1077–1081. https://doi.org/10.1212/01.wnl.0000055875.26908.24

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc.

Cheong, J. (2011). Accuracy of estimates and statistical power for testing meditation in latent growth curve modeling. *Structural Equation Modeling: a Multidisciplinary Journal*, *18*, 195–211. https://doi.org/10.1080/10705511.2011.557334

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 238–262. https://doi.org/10.1207/S15328007SEM1002_5

Cho, S.-H., & Huang, Y.-T. (2019). Mediation analysis with causally ordered mediators using Cox proportional hazards model. *Statistics in Medicine*, *38*, 1566–1581. https://doi.org/10.1002/sim.8058

Cramer, P. E., Cirrito, J. R., Wesson, D. W., Lee, C. Y. D., Karlo, J. C., Zinn, A. E., Casali, B. T., Restivo, J. L., Goebel, W. D., James, M. J., Brunden, K. R., Wilson, D. A., & Landreth, G. E. (2012). ApoE-directed therapeutics rapidly clear β-amyloid and reverse deficits in AD mouse models. *Science (New York, N.Y.)*, *335*, 1503–1506. https://doi.org/10.1126/science.1217697

Didelez, V. (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. *Lifetime Data Analysis*, *25*, 593–610. https://doi.org/10.1007/s10985-018-9449-0

Huang, Y.-T., & Cai, T. (2016). Mediation analysis for survival data using semiparametric probit models. *Biometrics*, *72*, 563–574. https://doi.org/10.1111/biom.12445

Huang, Y.-T., & Yang, H.-I. (2017). Causal mediation analysis of survival outcome with multiple mediators. *Epidemiology (Cambridge, Mass.)*, *28*, 370–378.

Ibrahim, J. G., Chen, M., Sinha, D., Ibrahim, J., & Chen, M. (2001). *Bayesian survival analysis, volume 2*. Springer.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334. https://doi.org/10.1037/a0020761

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Lange, T., & Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology*, *22*, 575–581. https://doi.org/10.1097/EDE.0b013e31821c680c

Lin, S.-H., Young, J. G., Logan, R., & VanderWeele, T. J. (2017). Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Statistics in Medicine*, *36*, 4153–4166. https://doi.org/10.1002/sim.7426

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122. https://doi.org/10.1007/BF02294746

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*, 371–402. https://doi.org/10.1037/1082-989X.2.4.371

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, *66*, 825–852. https://doi.org/10.1146/annurev-psych-010814-015258

Raber, J., Huang, Y., & Ashford, J. W. (2004). ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiology of Aging*, *25*, 641–650. https://doi.org/10.1016/j.neurobiolaging.2003.12.023

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*, 144–164. https://doi.org/10.1080/15427600902911247

Song, X., & Lee, S.-Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. John Wiley & Sons.

Sun, R., Zhou, X., & Song, X. (2021). Bayesian causal mediation analysis with latent mediators and survival outcome. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 778–790. https://doi.org/10.1080/10705511.2020.1863154

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass.)*, *22*, 582–585.

von Soest, T., & Hagtvet, K. A. (2011). Mediation analysis in a latent growth curve modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 289–314. https://doi.org/10.1080/10705511.2011.557344

Zeng, S., Lange, E. C., Archie, E. A., Campos, F. A., Alberts, S. C., & Li, F. (2022). A causal mediation model for longitudinal mediators and survival outcomes with an application to animal behavior. *Journal of Agricultural, Biological and Environmental Statistics*, *27*, 1–22. https://doi.org/10.1007/s13253-022-00490-6

Zeng, S., Rosenbaum, S., Alberts, S. C., Archie, E. A., & Li, F. (2021). Causal mediation analysis for sparse and irregular longitudinal data. *The Annals of Applied Statistics*, *15*, 747–767. https://doi.org/10.1214/20-AOAS1427

Zheng, C., & Liu, L. (2022). Quantifying direct and indirect effect for longitudinal mediator and survival outcome using joint modeling approach. *Biometrics*, *78*, 1233–1243. https://doi.org/10.1111/biom.13475

Zheng, W., & van der Laan, M. (2017). Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *Journal of Causal Inference*, *5*, 20160006. https://doi.org/10.1515/jci-2016-0006

Zhou, X., & Song, X. (2021). Mediation analysis for mixture cox proportional hazards cure models. *Statistical Methods in Medical Research*, *30*, 1554–1572. https://doi.org/10.1177/09622802211003113