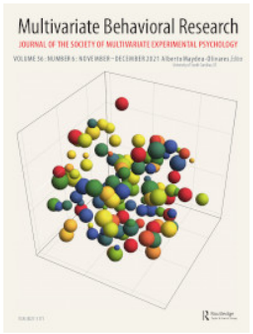

Multivariate Behavioral Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

Joint Hidden Markov Model for Longitudinal and Time-to-Event Data with Latent Variables

Xiaoxiao Zhou, Kai Kang, Timothy Kwok & Xinyuan Song

To cite this article: Xiaoxiao Zhou, Kai Kang, Timothy Kwok & Xinyuan Song (2021): Joint Hidden Markov Model for Longitudinal and Time-to-Event Data with Latent Variables, Multivariate Behavioral Research, DOI: 10.1080/00273171.2020.1865864

To link to this article: https://doi.org/10.1080/00273171.2020.1865864

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2020.1865864

Joint Hidden Markov Model for Longitudinal and Time-to-Event Data with Latent Variables

Xiaoxiao Zhou[a], Kai Kang[a], Timothy Kwok[b], and Xinyuan Song[a]

[a]Department of Statistics, Chinese University of Hong Kong; [b]Department of Medicine & Therapeutics, Prince of Wales Hospital Chinese University of Hong Kong

ABSTRACT

This study develops a new joint modeling approach to simultaneously analyze longitudinal and time-to-event data with latent variables. The proposed model consists of three components. The first component is a hidden Markov model for investigating a longitudinal observation process and its underlying transition process as well as their potential risk factors and dynamic heterogeneity. The second component is a factor analysis model for characterizing latent risk factors through multiple observed variables. The third component is a proportional hazards model for examining the effects of observed and latent risk factors on the hazards of interest. A shared random effect is introduced to allow the longitudinal and time-to-event outcomes to be correlated. A Bayesian approach coupled with efficient Markov chain Monte Carlo methods is developed to conduct statistical inference. The performance of the proposed method is evaluated through simulation studies. An application of the proposed model to a general health survey study concerning cognitive impairment and mortality for Chinese elders is presented.

KEYWORDS

Bayesian methods; hidden Markov model; latent variables; longitudinal response; time-to-event outcome

Introduction

In clinical trials or cohort studies, longitudinal measurements and possibly censored time-to-event outcomes are frequently collected together for each participant. Owing to the superiority to simultaneously reveal the structure of repeated measurements and its association with time-to-event process, joint modeling of longitudinal and time-to-event data has attracted considerable attention in the literature. Cox (1972) conducted a joint analysis of mixed-effect and proportional hazards (PH) models to investigate how the trajectory of longitudinal responses affects the hazards of interest. Gruttola and Tu (1994) and Tsiatis et al. (1995) considered similar joint models to examine the progression of CD4-lymphocyte count and its relationship to survival time. Later on, numerous attempts have been made to exploit the variants of the joint model to accommodate the diverse types of longitudinal and survival data. Larsen (2004) extended the PH model to include a latent class variable as a predictor, where the latent class membership is measured indirectly by multiple, longitudinal, and ordinal indicators. Rizopoulos et al. (2008) developed a two-part joint model to cope with survival and longitudinal binary data with excess zeros. Li and Luo (2017) incorporated functional data to jointly investigate time-to-Alzheimer's disease and longitudinal measurement of a cognitive test. Other relevant works (Andrinopoulou et al., 2018; Barrett & Su, 2017; Hashemi et al., 2003; Ibrahim et al., 2004; Jacqmin-Gadda et al., 2006) are also available in the literature.

Although the aforementioned models shed light on the application of joint modeling of longitudinal and survival data from distinct perspectives, their usual method to process the longitudinal data is establishing a trajectory model either in a parametric or nonparametric framework. The primary goal of majority of the existing works is to investigate the potential risk factors for certain events, thereby neglecting the intrinsic characteristics of the longitudinal measurements, for example, the heterogeneity caused by the mixture distribution. Brown and Ibrahim (2003) proposed a semiparametric model with Dirichlet process priors to partially address this problem. However, they focused only on relaxing the assumptions about the trajectory functions when parametric distribution

CONTACT Xinyuan Song xysong@sta.cuhk.edu.hk Department of Statistics, Chinese University of Hong Kong, Hong Kong, China.
Supplemental data for this article is available online at https://doi.org/10.1080/00273171.2020.1865864.

assumptions are inappropriate. No previous study has ever investigated the dynamic heterogeneity of covariate effects on the longitudinal response in the joint analysis of longitudinal and survival data. Hidden Markov model (HMM) and its variants have been widely recognized as important tools for longitudinal data analysis due to their superiority in simultaneously revealing the longitudinal dependency structure and dynamic heterogeneity of the observed process (Altman, 2007; Bartolucci & Farcomeni, 2009; Maruotti, 2011). Song et al. (2017) further extended HMM to accommodate latent variables that cannot be measured by a single observed variable, but instead should be characterized through multiple observed measurements. Incorporation of the latent variables into survival analyses has also been developed in recent literature (He et al., 2018; Ouyang et al., 2018; Pan et al., 2015). Nevertheless, available methods have seldom incorporated HMM into the joint analysis of longitudinal and time-to-event data. One exception in this direction is the work of Bartolucci and Farcomeni (2019), who introduced a common random effect following a time-continuous Markov process to jointly model the longitudinal response and time-to-event outcome. However, their model neither addressed dynamic covariates effects nor investigated the risk factors that affect the hidden transition process.

This study proposes a new joint modeling approach to address the abovementioned shortcomings. The proposed model comprises three major components. The first component is an HMM to simultaneously investigate a longitudinal observation process and its underlying dynamic transition process. The second component is a confirmatory factor analysis (CFA) model to group multiple observed measurements into latent variables, which may have common or distinct effects on the two types of outcomes. The last component is a PH model to investigate the potential risk factors for the hazards of interest. The longitudinal measurements, the transition process and time-to-event outcomes are allowed to be correlated through a shared subject-specific random effect. The proposed model extends existing joint models by integrating HMM, CFA, and PH models into a unified framework, which enables jointly examining longitudinal and time-to-event outcomes, discovering the dynamic transition and heterogeneous covariate effects in the longitudinal process, and accommodating the presence of latent variables. We develop a full Bayesian approach coupled with Markov chain Monte Carlo (MCMC) algorithms to conduct statistical inference because of its ability in managing complex models and allowing the use of prior information to achieve improved results. A modified deviance information criterion (DIC) is used to determine the number of hidden states (order) of HMM in the proposed model. Furthermore, we extend the prediction of hidden states and dynamic survival probability for newcomers to the proposed joint model.

This research was motivated by a general health survey study, in which community-dwelling men and women aged 65 years or over were invited to attend a health check at the Prince of Wales Hospital from 2001 to 2017 (Auyeung et al., 2011). Diverse characteristics, such as demographic information, physical activity, mental functioning, cognitive assessment, and mortality, were documented through questionnaires, measurements, and self-reports. The primary goal of this health survey is to facilitate cohort study on general health, multiple clinical outcomes, and quality of life for Chinese elders. Cognitive impairment, which was assessed repeatedly over time, is one of the important clinical outcomes reflecting elders' general health status and quality of life. Meanwhile, elders' mortality risk is of major concern in public health. This motivated us to jointly investigate cognitive impairment and mortality, their common or distinct risk factors, and possible heterogeneity in the progression of cognitive impairment. The cognitive function was assessed through the validated Cantonese version of Mini-Mental Status Examination (CMMSE), which is composed of 30 items that assess multiple domains of cognitive function. The mortality status was ascertained by annual death registry search in the Death Registry of Hong Kong Government with a censoring rate (CR) of 44.4%. The survey study also collected participants' demographical information, medical profiles, and multiple measurements, such as physical component summary (PCS) and mental component summary (MCS) calculated from the 12-Item Short Form Health Survey, gait speed measured using the best time in seconds to complete a walk along a straight line 6 m long in distance (6MPACE), maximum of left and right hand grip strength (GRIPMAX), and depression level assessed by Geriatric Depression Scale (GDS). In this study, CMMSE is a longitudinal response reflecting cognitive function, which may progress from normal to impaired state over time. The potential covariate effects on cognitive function may also vary across different cognitive states. Moreover, among the multiple measurements, PCS, 6MPACE, and GRIPMAX all summarize physical function, whereas MCS and GDS both measure mental function. Thus, we consider a

joint model, which includes an HMM for modeling the longitudinal response CMMSE, a factor analysis model for characterizing physical and mental functions through the multiple observed measurements, and a survival model for examining the observed and latent risk factors of the hazards of death. The proposed model perfectly accommodates all the data features. To our knowledge, no existing work can jointly accomplish the aforementioned multiple tasks.

The rest of this article is organized as follows. "Model description" describes the proposed model and discusses the associated model identifiability issues. "Bayesian inference" presents a Bayesian approach for statistical inference. A hybrid algorithm, which combines the Gibbs sampler, Metropolis–Hastings (MH) algorithm, and forward filtering and backward sampling (FFBS) algorithm, is proposed for parameter estimation. A modified DIC is employed to perform order selection of HMM, and a sampling-based procedure is developed for prediction. "Simulation study" investigates the empirical performance of the Bayesian inference through simulation studies. "Application" presents an application of the proposed methodology to a general health survey study concerning the cognitive impairment and mortality of Chinese elders. "Discussion" concludes the paper. Technical details are provided in the Online Appendix.

Model description

Hidden Markov model

Let $v_i(t)$ be a continuous response for subject i at time t. For brevity, we denote $v_{ij} = v_i(t_j)$ for $i = 1, ..., N$ and $j = 1, ..., m_i$, where m_i is the number of measurements for subject i. Let Z_{ij} be a hidden state process, taking values in a finite set $\{1, ..., S\}$ and following the first-order Markov chain. Given hidden state $Z_{ij} = s$, a conditional regression model for the continuous response is defined as follows:

$$\left[v_{ij}|Z_{ij} = s\right] = \rho_s + \boldsymbol{\gamma_s}^T\mathbf{x}_{ij} + \boldsymbol{\phi}_s^T\boldsymbol{\xi}_{ij} + w_i + \epsilon_{ij}^v, \qquad (1)$$

where ρ_s is a state-specific intercept, $\boldsymbol{\gamma}_s = (\gamma_{s1}, ..., \gamma_{sr})^T$ and $\boldsymbol{\phi}_s = (\varphi_{s1}, ..., \varphi_{sl})^T$ are $r \times 1$ and $l \times 1$ vectors of state-specific regression coefficients, respectively; $\mathbf{x}_{ij} = (x_{ij1}, ..., x_{ijr})^T$ is an $r \times 1$ vector of time-variant covariates; $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, ..., \xi_{ijl})^T$ is an $l \times 1$ vector of time-variant latent variables whose information is determined by (3); w_i is a subject-specific random effect, and $w_i \sim N(0, \sigma_w^2)$; ϵ_{ij}^v is a random residual independent of w_i, and $[\epsilon_{ij}^v|Z_{ij} = s] \sim N(0, \sigma_s^2)$. In Equation (1), we assume that v_{ij} is univariate for brevity. An extension to the case with multivariate $\mathbf{v}_{ij}$ is straightforward.

Let p_{ijus} denote the transition probability from state $Z_{ij} = u$ at time t_{j-1} to state $Z_{ij} = s$ at time t_j for subject i. On the basis of the assumption of the first-order Markov chain, the transition probability can be written as follows: $p_{ijus} = P(Z_{ij} = s|Z_{i1}, Z_{i2}, ..., Z_{i,j-1} = u) = P(Z_{ij} = s|Z_{i,j-1} = u)$. The hidden states $s = 1, ..., S$ are assumed to be ranked because they often have natural ranking information in substantive study. For instance, in the application of "Application," the two hidden states that reflect elders' cognitive function can be ranked from bad to good, namely, cognitive impairment and cognitive normal. Let $\vartheta_{ijus} = P(Z_{ij} = s|Z_{ij} \geq s, Z_{i,j-1} = u)$. Then, $\text{logit}(\vartheta_{ijus}) = \log\left(\frac{P(Z_{ij}=s|Z_{i,j-1}=u)}{P(Z_{ij}>s|Z_{i,j-1}=u)}\right) = \log\left(\frac{p_{ijus}}{p_{iju,s+1}+...+p_{ijuS}}\right)$. The effects of potential predictors on the transition probability can be examined through a continuation-ratio logit model as follows:

$$\text{logit}(\vartheta_{ijus}) = \zeta_{us} + \boldsymbol{\alpha}^T\mathbf{d}_{ij} + \alpha_w w_i, \qquad (2)$$

where the left-hand side is the log odds of transition to state s rather than to a state that is higher than s given that $Z_{i,j-1} = u$; ζ_{us} is a transition-specific intercept; $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)^T$ is an $m \times 1$ vector of regression coefficients; $\mathbf{d}_{ij} = (d_{ij1}, ..., d_{ijm})^T$ is an $m \times 1$ vector of covariates; and w_i is the subject-specific random effect defined in Model (1). Similar to the proportional odds assumption in a cumulative logit model, $\boldsymbol{\alpha}$ in (2) is assumed to be independent of u and s to maintain the order of the hidden states. The initial probability of the hidden state at occasion $j = 1$ for subject i is defined as follows: for $s = 1, ..., S-1$, $\log\left(\frac{P(Z_{i1}=s)}{P(Z_{i1}>s)}\right) = \log\left(\frac{p_{i10s}}{p_{i10,s+1}+...+p_{i10,S}}\right) = \tau_s$.

Confirmatory factor analysis (CFA) model

Given hidden state $Z_{ij} = s$, we consider a confirmatory factor analysis model as follows:

$$\left[\mathbf{y}_{ij}|Z_{ij} = s\right] = \boldsymbol{\mu}_s + \boldsymbol{\Lambda}_s\boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \qquad (3)$$

where $\mathbf{y}_{ij} = (y_{ij1}, ..., y_{ijp})^T$ is a $p \times 1$ vector of observed variables that characterize the $q \times 1$ vector of latent variables in $\boldsymbol{\xi}_{ij}$ at time t_j; $\boldsymbol{\Lambda}_s$ is a $p \times q$ $(p > q)$ state-specific matrix of factor loadings; $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, ..., \epsilon_{ijp})^T$ is $p \times 1$ residual vector independent of $\boldsymbol{\xi}_{ij}$; $[\boldsymbol{\xi}_{ij}|Z_{ij} = s] \sim N(0, \boldsymbol{\Phi}_s)$ with a $q \times q$ covariance matrix $\boldsymbol{\Phi}_s$; and $[\boldsymbol{\epsilon}_{ij}|Z_{ij} = s] \sim N(0, \boldsymbol{\Psi}_s)$ with a $p \times p$ diagonal covariance matrix $\boldsymbol{\Psi}_s$. The elements of $\boldsymbol{\Lambda}_s$ reflect the associations between the observed variables and their corresponding latent variables. In a CFA model, q (the dimension of $\boldsymbol{\xi}_{ij}$) and the structure of $\boldsymbol{\Lambda}_s$ are

predetermined according to subject knowledge, experts' suggestion, and/or existing literature. For example, in the application of "Application," the existing medical literature (Auyeung et al., 2011) suggests that {PCS, 6MPACE, GRIPMAX} and {MCS, GDS} represent physical and mental functions, respectively, from different perspectives. Hence, $q = 2$ and $\mathbf{\Lambda}_s$ has a non-overlapping structure, in which the factor loadings relating {PCS, GRIPMAX, 6MPACE} to physical function and {GDS, MCS} to mental function are nonzero, and other factor loadings are fixed to zero. If such information is unavailable, one can conduct an exploratory factor analysis to determine q and the structure of $\mathbf{\Lambda}_s$ based on the data (Song & Lee, 2012).

Proportional hazards model

For subject i, let T_i be the time to death for the ith subject, $\mathbf{f}_i$ be an $h \times 1$ vector of baseline observed covariates, $\boldsymbol{\xi}_{i1}$ be an $l \times 1$ vector of latent variables at baseline, C_i be the censoring time independent of T_i given $\mathbf{f}_i$ and $\boldsymbol{\xi}_{i1}$. Let $U_i = \min(T_i, C_i)$ be the observed time and $\delta_i = I(T_i \leq C_i)$ be the failure indicator, where $I(\cdot)$ is the indicator function. The PH model specifies that, given $\mathbf{f}_i$, $\boldsymbol{\xi}_{i1}$ and w_i, the hazard function of T_i takes the form of

$$\lambda(t|\mathbf{f}_i, \boldsymbol{\xi}_{i1}, w_i) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^T \mathbf{f}_i + \boldsymbol{\eta}^T \boldsymbol{\xi}_{i1} + \eta_w w_i\right), \quad (4)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function; $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are the $h \times 1$ and $l \times 1$ vectors of unknown regression parameters, respectively; and w_i is the subject-specific random effect involved in Models (1) and (2). Notably, incorporating the trajectory of dynamic latent variable $\boldsymbol{\xi}_{ij}$ into (4) provides additional information for the latent risk factors of hazard function $\lambda(t|\cdot)$. However, characterizing the trajectory of $\boldsymbol{\xi}_{ij}$ requires sufficient number of measurements, which are unavailable for some subjects in the motivating example presented in "Application." In Models (1, 2), and (4), a univariate random effect w_i (together with coefficients α_w and η_w) is introduced to address possible correlations among longitudinal observations, dependency between the observation and transition processes, and association between the longitudinal and time-to-event processes. An extension to the case where the three models include distinct but correlated random effects can be considered without much difficulty.

Equations (1)–(4) define a new joint model for longitudinal and survival data through the shared subject-specific random effect w_i and shared latent risk factors $\boldsymbol{\xi}_{ij}$. The proposed model is advantageous over the conventional joint analyses in several aspects. First, unlike the existing joint models that focused on the longitudinal trajectory of responses, the proposed model jointly investigates the effects of potential risk factors on the longitudinal response and time-to-event variable of interest, and simultaneously models the dynamic heterogeneity of the longitudinal observations under the HMM framework. Second, the CFA model groups highly correlated observed variables into relatively independent latent variables, thereby eliminating the multicollinearity problem and reducing model dimensionality. Finally, the shared subject-specific random effect accounts for the association between the longitudinal and time-to-event processes and the effects of possibly omitted covariates in the conditional regression, transition, and PH models (1, 2), and (4).

Model identifiability

There are two model indeterminacies in the proposed joint model. One stems from the label switching problem elicited by the likelihood function in an HMM being invariant to a random permutation of the state labels, which results in multi-modal posterior distribution under the symmetric priors of the parameters in different states. We address this issue by imposing constraint $\rho_1 < \cdots < \rho_S$ in implementing posterior sampling. Notably, this constraint is inapplicable when v_{ij} is multivariate. In this situation, $\rho_1, ..., \rho_S$ are vectors, and the permutation sampler proposed by Frühwirth-Schnatter (2001) can be used to address the problem. The permutation sampler consists of two stages: the first stage implements the MCMC agorithm in an unconstrained parameter space to select a suitable constraint, and the second stage proceeds with the MCMC algorithm in the constrained parameter space to conduct statistical inference. The other indeterminacy originates from the invariance of the covariance matrix of the latent variables under orthogonal transformation. In this study, we follow a common practice in latent variable modeling literature (e.g., Song & Lee, 2012) to fix certain elements of $\mathbf{\Lambda}_s$ at preassigned values (see "Simulation study" and "Application") to address this problem.

Bayesian inference

Bayesian estimation

For the unknown baseline hazard function $\lambda_0(t)$, we assume a piecewise exponential distribution (Ibrahim et al., 2014). Let $0 < u_1 < ... < u_G$ be a finite partition of the time axis with $u_G > U_i$ for all $i = 1, ..., N$. In

the gth interval, let $\lambda_0(t) = \lambda_g$ for $t \in (u_{g-1}, u_g]$ and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_G)^T$. Define $\nu_{ig} = 1$ if $U_i \in (u_{g-1}, u_g]$.

We first specify prior distributions for the unknown parameters. For the parameters involved in the conditional regression model (1), we consider priors as follows:

$$p(\rho_s) \overset{D}{=} N(\rho_{s0}, \sigma^2_{\rho s0}), \quad p(\boldsymbol{\gamma}_s) \overset{D}{=} N(\boldsymbol{\gamma}_{s0}, \boldsymbol{\Sigma}_{\gamma s0}), \quad p(\boldsymbol{\phi}_s) =^D N(\boldsymbol{\phi}_{s0}, \boldsymbol{\Sigma}_{\phi s0})$$
$$p(\sigma_s^{-2}) =^D \text{Gamma}(\tilde{\alpha}_{s0}, \tilde{\beta}_{s0}), \qquad p(\sigma_w^{-2}) =^D \text{Gamma}(\tilde{\alpha}_{w0}, \tilde{\beta}_{w0}),$$

where "$p(\cdot) \overset{D}{=}$" denotes "the distribution $p(\cdot)$ is equal to," and $\rho_{s0}, \sigma^2_{\rho s0}, \boldsymbol{\gamma}_{s0}, \boldsymbol{\Sigma}_{\gamma s0}, \boldsymbol{\phi}_{s0}, \boldsymbol{\Sigma}_{\phi s0}, \tilde{\alpha}_{s0}, \tilde{\beta}_{s0}, \tilde{\alpha}_{w0}$, and $\tilde{\beta}_{w0}$ are hyperparameters whose values are preassigned.

For the parameters involved in the transition model (2), we assign normal priors:

$$p(\zeta_{us}) \overset{D}{=} N(\zeta_{us0}, \sigma^2_{\zeta us0}), \quad p(\boldsymbol{\alpha}) \overset{D}{=} N(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}0}),$$
$$p(\alpha_w) \overset{D}{=} N(\alpha_{w0}, \sigma^2_{\alpha w0}), \quad p(\tau_s) \overset{D}{=} N(\tau_{s0}, \sigma^2_{\tau s0}),$$

where $\zeta_{us0}, \sigma^2_{\zeta us0}, \boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}0}, \alpha_{w0}, \sigma^2_{\alpha w0}, \tau_{s0}$, and $\sigma^2_{\tau s0}$ are hyperparameters whose values are preassigned.

For the parameters involved in the CFA model (3), we consider priors as follows:

$$p(\boldsymbol{\Lambda}_{sk}|\psi_{sk}) \overset{D}{=} N(\boldsymbol{\Lambda}_{sk0}, \psi_{sk}\boldsymbol{\Sigma}_{\lambda sk}), p(\psi_{sk}^{-1}) \overset{D}{=} \text{Gamma}(a_{sk}, b_{sk}), \quad k = 1, ..., p,$$
$$p(\boldsymbol{\mu}_s) \overset{D}{=} N(\boldsymbol{\mu}_{s0}, \boldsymbol{\Sigma}_{\mu s0}), \qquad p(\boldsymbol{\Phi}_s^{-1}) =^D \text{Wishart}(\mathbf{R}_s, r_s),$$

where $\boldsymbol{\Lambda}_{sk}$ and ψ_{sk} are the kth row and kth diagonal element of $\boldsymbol{\Lambda}_s$ and $\boldsymbol{\Psi}_s$, respectively; $\boldsymbol{\Lambda}_{sk0}, \boldsymbol{\Sigma}_{\lambda sk}, a_{sk}, b_{sk}, \boldsymbol{\mu}_{s0}, \boldsymbol{\Sigma}_{\mu s0}, \mathbf{R}_s$, and r_s are hyperparameters with preassigned values.

For the parameters involved in the PH model (4), we follow Ibrahim et al. (2014) to specify gamma and normal priors as follows:

$$p(\lambda_g) \overset{D}{=} \text{Gamma}(\tilde{\alpha}_{\lambda g0}, \tilde{\beta}_{\lambda g0}), \quad p(\boldsymbol{\beta}) \overset{D}{=} N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\beta 0}),$$
$$p(\boldsymbol{\eta}) \overset{D}{=} N(\boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\eta 0}), \quad p(\eta_w) \overset{D}{=} N(\eta_{w0}, \sigma^2_{\eta w0}),$$

where $\tilde{\alpha}_{\lambda g0}, \tilde{\beta}_{\lambda g0}, \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\beta 0}, \boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\eta 0}, \eta_{w0}$ and $\sigma^2_{\eta w0}$ are hyperparameters with preassigned values.

Let $\boldsymbol{\theta}$ be the vector that contains all the unknown parameters, $\mathbf{v}_i = (v_{i1}, ..., v_{im_i})^T$, $\mathbf{V} = (\mathbf{v}_1, ..., \mathbf{v}_N)^T$, $\mathbf{Y}_i = (\mathbf{y}_{i1}, ..., \mathbf{y}_{im_i})^T$, $\mathbf{Y} = (\mathbf{Y}_1, ..., \mathbf{Y}_N)^T$, $\mathbf{X}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{im_i})^T$, $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_N)^T$, $\mathbf{D}_i = (\mathbf{d}_{i1}, ..., \mathbf{d}_{im_i})^T$, $\mathbf{D} = (\mathbf{D}_1, ..., \mathbf{D}_N)^T$, $\boldsymbol{\Xi}_i = (\boldsymbol{\xi}_{i1}, ..., \boldsymbol{\xi}_{im_i})^T$, $\boldsymbol{\Xi} = (\boldsymbol{\Xi}_1, ..., \boldsymbol{\Xi}_N)^T$, $\mathbf{Z}_i = (Z_{i1}, ..., Z_{m_i})^T$, $\mathbf{Z} = (\mathbf{Z}_1, ..., \mathbf{Z}_N)^T$, $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_N)^T$, $\boldsymbol{\delta} = (\delta_1, ..., \delta_N)^T$, $\mathbf{U} = (U_1, ..., U_N)^T$, $\boldsymbol{w} = (w_1, ..., w_N)^T$, and $\mathbf{Q} = (\mathbf{V}, \mathbf{Y}, \mathbf{X}, \mathbf{D}, \mathbf{F}, \boldsymbol{\delta}, \mathbf{U})$. In Bayesian estimation, the major task is sampling from $p(\boldsymbol{\theta}|\mathbf{Q})$. However, this posterior distribution involves high-dimensional integrals with respect to latent quantities $\{\mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{w}\}$. We utilize the idea of data augmentation to work on the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{w}|\mathbf{Q})$. This joint posterior distribution is still intractable because of its extremely complex form. Thus, we develop a MCMC algorithm, which combines the Gibbs sampler, MH algorithm, and FFBS algorithm, to facilitate the posterior sampling. The Gibbs sampler is implemented by iteratively sampling (i) $\mathbf{Z}$ from $p(\mathbf{Z}|\mathbf{Q}, \boldsymbol{\Xi}, \boldsymbol{w}, \boldsymbol{\theta})$, (ii) $\boldsymbol{\Xi}$ from $p(\boldsymbol{\Xi}|\mathbf{Q}, \mathbf{Z}, \boldsymbol{w}, \boldsymbol{\theta})$, (iii) $\boldsymbol{w}$ from $p(\boldsymbol{w}|\mathbf{Q}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{\theta})$, and (iv) $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\mathbf{Q}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{w})$. The details of the full conditional distributions involved in the MCMC algorithm are provided in the Online Appendix. The convergence of the MCMC algorithm can be claimed if several MCMC chains starting from different initial values mix adequately. After convergence, we collect a large number of posterior samples $\{(\boldsymbol{\theta}^{(k)}, \mathbf{Z}^{(k)}, \boldsymbol{\Xi}^{(k)}, \boldsymbol{w}^{(k)}) : k = 1, ..., K\}$ to obtain the Bayesian estimate of $\boldsymbol{\theta}$ through the posterior mean of $\{\boldsymbol{\theta}^{(k)} : k = 1, ..., K\}$.

Moreover, the hidden states can be estimated as follows:

$$\hat{Z}_{ij} = \arg \max_{s \in \{1, ..., S\}} P(Z_{ij} = s|\mathbf{v}_i, \boldsymbol{\theta})$$
$$\approx \arg \max_{s \in \{1, ..., S\}} \frac{1}{K} \sum_{k=1}^{K} I(Z_{ij}^{(k)} = s), \tag{5}$$

where $Z_{ij}^{(k)}$ denotes the hidden state of v_{ij} at the kth iteration, and $\arg\max_{s \in \{1, ..., S\}} \frac{1}{K} \sum_{k=1}^{K} I(Z_{ij}^{(k)} = s)$ is the posterior mean of the hidden states of v_{ij} drawn from the MCMC iterations.

Order selection

Order selection (determination of the number of hidden states) of HMM is an indispensable task in the analysis of the proposed model. While Bayes factor is a common Bayesian statistic for model selection, its computation is highly challenging and time consuming under the present modeling framework. As a good alternative to Bayes factor, DIC (Spiegelhalter et al., 2002) is frequently used for model selection in Bayesian analysis. However, the use of DIC in the context of mixture-type models is problematic. Celeux et al. (2006) explored several variants of DIC and ascertained that a modified DIC showed prominent performance among all the evaluated versions of DIC. The modified DIC is defined as follows:

$$\text{DIC} = E_{\boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{w}}\left[-2 \log p(\mathbf{Q}, \mathbf{Z}, \boldsymbol{\Xi}, \boldsymbol{w}|\boldsymbol{\theta})|\mathbf{Q}\right] + p_D, \tag{6}$$

where the first term of the right-hand side is the posterior mean deviance for assessing the model fit; p_D is the effective number of parameters for evaluating model complexity, $p_D = E_{\boldsymbol{\theta}, \boldsymbol{\Xi}, \boldsymbol{w}}$,

$\mathbf{Z}[-2\log p(\mathbf{Q},\mathbf{Z},\boldsymbol{\Xi},\boldsymbol{w}|\boldsymbol{\theta})|\mathbf{Q}] + 2E_{\mathbf{Z},\boldsymbol{\Xi},\boldsymbol{w}}[\log p(\mathbf{Q},\mathbf{Z},\boldsymbol{\Xi},\boldsymbol{w})|E_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{Q},\mathbf{Z},\boldsymbol{\Xi},\boldsymbol{w}])|\mathbf{Q}]$; and $\log p(\mathbf{Q},\mathbf{Z},\boldsymbol{\Xi},\boldsymbol{w}|\boldsymbol{\theta})$ is the complete-data log-likelihood function. The specific form of $\log p(\mathbf{Q},\mathbf{Z},\boldsymbol{\Xi},\boldsymbol{w}|\boldsymbol{\theta})$ and the computation of DIC in (6) are provided in the Online Appendix.

Posterior prediction

Assume that we have collected longitudinal measurements $\mathcal{M}_i^{\{t\}} = \{\mathbf{y}_{ij}, v_{ij}, \mathbf{d}_{ij}, \mathbf{x}_{ij}; j = 1, ..., m_i, t_{m_i} \le U_i\}$ and baseline covariates $\mathbf{f}_i$. Two components are worthy of investigation. One is the hidden states in $\mathbf{Z}_i$ conveying the progression of the status that we care about, and the other is the survival probability on the relevant time frame $(t, t+\Delta]$, both of which are indispensable for the follow-up intervention. Let $\mathcal{Q}_n = \{\mathbf{V}_i, \mathbf{Y}_i, \mathbf{X}_i, \mathbf{D}_i, \mathbf{F}_i, \delta_i, U_i, i = 1, ..., n\}$ denote the sample, based on which the proposed model is fitted and the prediction is made. In the Bayesian framework, $\mathbf{Z}_i$ can be directly sampled from its posterior distribution (see the Online Appendix). Then, we concentrate on the conditional probability of survival time $t+\Delta$ given survival up to t, which is calculated as follows:

$$\begin{aligned}\pi_i(t+\Delta|t) &= \Pr(T_i \ge t+\Delta | T_i \ge t, \mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathcal{Q}_n)\\ &= \int \Pr(T_i \ge t+\Delta | T_i \ge t, \mathcal{M}_i^{\{t\}}, \mathbf{f}_i; \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{Q}_n) d\boldsymbol{\theta}.\end{aligned} \tag{7}$$

The latter of the integrand, which is the posterior distribution of the parameter based on the observed data, can be well approximated by $N(\hat{\boldsymbol{\theta}}, \hat{\mathcal{H}})$ with $\hat{\boldsymbol{\theta}}$ being the Bayesian estimate of $\boldsymbol{\theta}$ and $\hat{\mathcal{H}} = \widehat{\mathrm{var}}(\hat{\boldsymbol{\theta}})$, according to the arguments of standard asymptotic Bayesian theory (Cox & Hinkley, 1979) and the assumption of sufficiently large sample size. The former of the integrand can be written as

$$\begin{aligned}&\Pr(T_i \ge t+\Delta | T_i \ge t, \mathcal{M}_i^{\{t\}}, \mathbf{f}_i; \boldsymbol{\theta})\\ &= \int\int\int \Pr(T_i \ge t+\Delta | T_i \ge t, \mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathbf{Z}_i, \xi_i, w_i; \boldsymbol{\theta}) p(\mathbf{Z}_i, \xi_i, w_i | T_i \ge t, \mathcal{M}_i^{\{t\}}, \mathbf{f}_i) d\mathbf{Z}_i d\xi_i dw_i\\ &= \int\int\int \frac{\Pr(T_i \ge t+\Delta | \mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathbf{Z}_i, \xi_i, w_i; \boldsymbol{\theta})}{\Pr(T_i \ge t | \mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathbf{Z}_i, \xi_i, w_i; \boldsymbol{\theta})} p(\mathbf{Z}_i, \xi_i, w_i | T_i \ge t, \mathcal{M}_i^{\{t\}}, \mathbf{f}_i) d\mathbf{Z}_i d\xi_i dw_i,\end{aligned} \tag{8}$$

where $\Pr(T_i \ge t | \mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathbf{Z}_i, \xi_i, w_i; \boldsymbol{\theta}) = \exp\{-\sum_{g=1}^{G} \nu_{ig} \sum_{k=1}^{g} \int_{u_{k-1}}^{min(u_k, t)} \lambda_k \exp(\varpi_i)\}$, denoted as $S_i(t|\mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathbf{Z}_i, \boldsymbol{\xi}_i, w_i; \boldsymbol{\theta})$. Substituting (8) into (7), the Monte Carlo estimates of $\mathbf{Z}_i$ and $\pi_i(t+\Delta|t)$ can be obtained using the following scheme:

- Step 1: Draw $\boldsymbol{\theta}^{(b)} \sim N(\hat{\boldsymbol{\theta}}, \hat{\mathcal{H}})$.
- Step 2: Draw $\mathbf{Z}_i^{(b)} \sim p(\mathbf{Z}_i | T_i > t, \mathbf{Q}_i, w_i^{(b-1)}, \boldsymbol{\xi}_i^{(b-1)}, \boldsymbol{\theta}^{(b)})$.
- Step 3: Draw $\boldsymbol{\xi}_i^{(b)} \sim p(\boldsymbol{\xi}_i | T_i > t, \mathbf{Q}_i, w_i^{(b-1)}, \mathbf{Z}_i^{(b)}, \boldsymbol{\theta}^{(b)})$.
- Step 4: Draw $w_i^{(b)} \sim p(w_i | T_i > t, \mathbf{Q}_i, \boldsymbol{\xi}_i^{(b)}, \mathbf{Z}_i^{(b)}, \boldsymbol{\theta}^{(b)})$.
- Step 5: Compute $\pi_i^{(b)}(t+\Delta|t) = S_i(t+\Delta|\mathcal{M}_i^{\{t+\Delta\}}, \mathbf{f}_i, \mathbf{Z}_i^{(b)}, \boldsymbol{\xi}_i^{(b)}, w_i^{(b)}; \boldsymbol{\theta}^{(b)}) \times [S_i(t|\mathcal{M}_i^{\{t\}}, \mathbf{f}_i, \mathbf{Z}_i^{(b)}, \boldsymbol{\xi}_i^{(b)}, w_i^{(b)}; \boldsymbol{\theta}^{(b)})]^{-1}$.
- Step 6: For each subject i, repeat Steps 1–5 B times ($B = 100$ in this study).

Finally, $\mathbf{Z}_i$ can be estimated using (5), and $\pi_i(t+\Delta|t)$ can be estimated by

$$\hat{\pi}_i(t+\Delta|t) = B^{-1} \sum_{b=1}^{B} \pi_i^{(b)}(t+\Delta|t). \tag{9}$$

Under this machinery, the subject-specific survival probability can be dynamically updated with the latest longitudinal information (Rizopoulos, 2011). The other inherent characteristic of the model is to distinguish between patients who are going to experience the event within the given time frame from those who are not (Andrinopoulou et al., 2018). We utilize the time-dependent area under the receiver operating characteristic curve (AUC) to assess the dynamic discrimination ability. Specifically, given a randomly chosen pair of subjects (i_1, i_2), AUC is defined as $\mathrm{AUC}(t,\Delta) = \Pr(\pi_{i1}(t+\Delta|t) < \pi_{i2}(t+\Delta|t) | \{T_{i_1} \in (t, t+\Delta]\} \cap \{T_{i_2} > t+\Delta\})$. The proposed model is expected to assign higher survival probability for subject i_2 if subject i_1 experiences the event within the specific period $(t, t+\Delta]$, whereas subject i_2 does not. A large AUC value indicates superior discrimination ability. We follow Andrinopoulou et al. (2018) to account for the problem of not fully observed survival time T_i to estimate $\mathrm{AUC}(t,\Delta)$. In addition, we assess dynamic prediction error using Brier score (BS) (Graf et al., 1999), which is defined as $\mathrm{BS}(t,\Delta) = E[\{\boldsymbol{I}(T > t+\Delta | T > t) - \hat{\pi}(t+\Delta|t)\}^2]$. We also follow Emura et al. (2018) to accommodate the presence of censoring in calculating BS.

Simulation study

In this section, Simulation 1 focuses on estimation, Simulation 2 investigates the performance of the modified DIC in order selection, and Simulation 3 conducts out-of-sample prediction.

Simulation 1

We generate datasets from the joint model defined by (1)–(4) with S = 2. In the CFA model (3), $p = 6$, $q = 2$, $\mathbf{y}_{ij} = (y_{ij1}, ..., y_{ij6})^T$, $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \xi_{ij2})^T$, and $\boldsymbol{\Lambda}_s$ has a non-overlapping structure of

$$\boldsymbol{\Lambda}_s = \begin{bmatrix} 1^* & \lambda_{s,21} & \lambda_{s,31} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 1^* & \lambda_{s,52} & \lambda_{s,62} \end{bmatrix}, \qquad s = 1, 2,$$

where the elements with asterisk are fixed to obtain

an identified mode and clear interpretation of the latent variables. In the conditional model (1), $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2})^T$, where x_{ij1} and x_{ij2} are generated from $N(0, 1)$ and $U[-1, 1]$, respectively, and $U[a, b]$ denotes the uniform distribution defined in $[a, b]$, the random effect w_i is generated from $N(0, \sigma_w^2)$, and $[\epsilon_{ij}^v | Z_{ij} = s] \sim N(0, \sigma_s^2)$. In the transition model (2), we assume that $\boldsymbol{d}_{ij} = (d_{ij1}, d_{ij2})^T$, where d_{ij1} and d_{ij2} are independently drawn from Bernoulli(0.6), the Bernoulli distribution with a success probability of 0.6, and $N(0, 1)$, respectively. In the PH model (4), $\mathbf{f}_i = (f_{i1}, f_{i2})^T$, where f_{i1} and f_{i2} are both drawn from $N(0, 1)$.

Three types of baseline hazard functions, $\lambda_0(t) = 0.5, \lambda_0(t) = t + 0.2,$ and $\lambda_0(t) = t^2 + 0.3,$ are considered. The censoring time C_i is generated independently from uniform distribution $U[c_1, c_2]$, where c_1 and c_2 are taken to achieve CR of 30% and 50%. The true population values of the state-specific parameters are set as follows: At state 1, $\rho_1 = \gamma_{11} = \gamma_{12} = \varphi_{11} = \varphi_{12} = -1, \sigma_1^2 = 0.25, \zeta_{11} = -1, \boldsymbol{\mu}_1 = (\mu_{11}, ..., \mu_{16})^T = (-2, ..., -2)^T, \lambda_{1,21} = \lambda_{1,31} = \lambda_{1,52} = \lambda_{1,62} = 0.8,$ $\boldsymbol{\Psi}_1 = 0.25\mathbf{I}_6,$ and $\{\phi_{1,11}, \phi_{1,12}, \phi_{1,22}\} = \{1, 0.3, 1\},$ where $\mathbf{I}_6$ is the 6×6 identity matrix, and $\phi_{1,11}, \phi_{1,22},$ and $\phi_{1,12}$ are the main- and off-diagonal elements of Φ_1. At state 2, $\rho_2 = \gamma_{21} = \gamma_{22} = \varphi_{21} = \varphi_{22} = 1, \sigma_2^2 = 0.36, \zeta_{21} = 1, \boldsymbol{\mu}_2 = (\mu_{21}, ..., \mu_{26})^T = (2, ..., 2)^T, \lambda_{2,21} = \lambda_{2,31} = \lambda_{2,52} = \lambda_{2,62} = 0.9,$ $\boldsymbol{\Psi}_2 = 0.36\mathbf{I}_6,$ and $\{\phi_{2,11}, \phi_{2,12}, \phi_{2,22}\} = \{1, 0.2, 1\}$. The true population values of state-invariant parameters are set as $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T = (1, -1)^T, \alpha_w = 0.5, \tau_0 = 0, \boldsymbol{\beta} = (\beta_1, \beta_2)^T = (-0.8, -0.8)^T, \boldsymbol{\eta} = (\eta_1, \eta_2)^T = (0.8, -0.8)^T, \eta_w = 0.8,$ and $\sigma_w^2 = 1$.

To evaluate the finite sample performance of Bayesian estimation, we consider four scenarios with $(N, \text{CR}) = (200, 30\%), (500, 30\%), (200, 50\%),$ and $(500, 50\%)$. For each scenario, we conduct 100 replications, in which the last observation time t_{m_i} is smaller than T_i and varies from 3 to 6 for $i = 1, ..., N$. The hyperparameters of the prior distributions specified in "Bayesian estimation" are assigned as follows (Prior I): $\rho_{s0}, \tau_{s0}, \zeta_{us0},$ and the elements of $\gamma_{s0}, \boldsymbol{\phi}_{s0}, \boldsymbol{\alpha}_0, \boldsymbol{\Lambda}_{sk0}, \boldsymbol{\mu}_{s0}, \boldsymbol{\beta}_0,$ and $\boldsymbol{\eta}_{s0}$ are 0; $\alpha_{w0} = \eta_{w0} = 0; \sigma_{\alpha w0}^2 = \sigma_{\eta w0}^2 = \sigma_{\rho s0}^2 = \sigma_{\zeta us0}^2 = \sigma_{\tau s0}^2 = 1, \boldsymbol{\Sigma}_{\gamma s0} = \boldsymbol{\Sigma}_{\varphi s0} = \boldsymbol{\Sigma}_{\alpha 0} = \boldsymbol{\Sigma}_{\lambda sk} = \boldsymbol{\Sigma}_{\mu s0} = \boldsymbol{\Sigma}_{\beta 0} = \boldsymbol{\Sigma}_{\eta 0} = \mathbf{I},$ where $\mathbf{I}$ is an identity matrix of appropriate dimension; $\tilde{\alpha}_{s0} = \tilde{\alpha}_{w0} = a_{sk} = 9, \tilde{\beta}_{s0} = \tilde{\beta}_{w0} = b_{sk} = 4; \tilde{\alpha}_{\lambda g0} = 0.2, \tilde{\beta}_{\lambda g0} = 0.4;$ $\mathbf{R}_s = 5\mathbf{I},$ and $r_s = 4$. We first conduct three test runs starting from different initial values to check convergence of the MCMC algorithm. Figure S1 in the Online Appendix shows the trace plots of the three parallel chains of several randomly selected parameters in $(\gamma_s, \boldsymbol{\phi}_s, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta})$. The MCMC chains that start from different initial values mix within 1000 iterations, suggesting a quick convergence of the algorithm. Thus, we discard 1000 burn-in iterations and collect the subsequent 5000 posterior samples to conduct posterior inference. The obtained results are summarized on the basis of 100 replications. Tables 1–4 present the averaged values of the bias (Bias) and root mean square error (RMS) between the Bayesian estimates and true population values of the unknown parameters. The Bias and RMS for most of the parameters are close to zero, indicating a satisfactory performance of parameter estimation. However, those involved in the transition and PH models perform relatively worse due to their nonlinearity. The performance is improved as either the sample size increases from $N = 200$ to $N = 500$ or CR decreases from 50% to 30%. The hidden states $\{Z_{ij}, i = 1, ..., N, j = 1, ..., m_i\}$ are estimated on the basis of (5). The correct classification rate is approximately 99% under each scenario and replication.

To assess the sensitivity of Bayesian estimation to prior inputs, we disturb the hyperparameters as follows (Prior II): $\rho_{s0}, \tau_{s0}, \zeta_{us0}, \alpha_{w0}, \eta_{w0},$ and the elements of $\gamma_{s0}, \boldsymbol{\phi}_{s0}, \boldsymbol{\alpha}_0, \boldsymbol{\Lambda}_{sk0}, \boldsymbol{\mu}_{s0}, \boldsymbol{\beta}_0,$ and $\boldsymbol{\eta}_{s0}$ are 2; $\sigma_{\alpha w0}^2 = \sigma_{\eta w0}^2 = \sigma_{\rho s0}^2 = \sigma_{\zeta us0}^2 = \sigma_{\tau s0}^2 = 100, \boldsymbol{\Sigma}_{\gamma s0} = \boldsymbol{\Sigma}_{\varphi s0} = \boldsymbol{\Sigma}_{\alpha 0} = \boldsymbol{\Sigma}_{\lambda sk} = \boldsymbol{\Sigma}_{\mu s0} = \boldsymbol{\Sigma}_{\beta 0} = \boldsymbol{\Sigma}_{\eta 0} = 100\mathbf{I};$ $\tilde{\alpha}_{s0} = \tilde{\alpha}_{w0} = a_{sk} = 10, \tilde{\beta}_{s0} = \tilde{\beta}_{w0} = b_{sk} = 3; \tilde{\alpha}_{\lambda g0} = 0.5, \tilde{\beta}_{\lambda g0} = 1;$ $\mathbf{R}_s = 5\mathbf{I},$ and $r_s = 3$. The Bayesian estimates under Prior II are similar to those in Tables 1–4 and omitted.

We further investigate the sensitivity of Bayesian results to the normality assumption of the latent variables and random residuals. The setting is the same as before, except that $\boldsymbol{\xi}_{ij}, \epsilon_{ij}^v$ or $\boldsymbol{\epsilon}_{ij}$ follows non-normal distributions as follows: Case (1): $\boldsymbol{\xi}_{ij} \sim I(Z_{ij} = 1)\{\frac{2}{3}N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \frac{1}{3}N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)\} + I(Z_{ij} = 2)\{\frac{2}{3}N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2) + \frac{1}{3}N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\},$ where $\boldsymbol{\mu}_1 = [-0.5, -0.5]^T, \boldsymbol{\mu}_2 = [1, 1]^T,$ $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix},$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix};$ Case (2): $\xi_{ijq} \sim I(Z_{ij} = 1)t_{(3)} + I(Z_{ij} = 2)t_{(4)},$ where $t_{(a)}$ denotes the t distribution with degree of freedom a; Case (3): $\epsilon_{ij}^v \sim \frac{2}{5}N(-0.3, 0.5) + \frac{3}{5}N(0.2, 0.5);$ Case (4): $\epsilon_{ij}^v \sim Gamma(5, 4),$ where $Gamma(\cdot, \cdot)$ denotes the Gamma distribution. Case (5): $\epsilon_{ijq} \sim \frac{2}{5}N(-0.3, 0.5) + \frac{3}{5}N(0.2, 0.5);$ Case (6): $\epsilon_{ijq} \sim Beta(3, 1),$ where $Beta(\cdot, \cdot)$ denotes the beta distribution. The estimation results under the scenario $(N, \text{CR}) = (200, 30\%)$ are presented in Tables S1 and S2 of the Online Appendix. The parameter estimates, especially the regression coefficient estimates, perform similarly to

those presented in Table 1. We also check the standard error (SE) estimates and the 95% credible interval (C.I.) constructed based on the 2.5% and 97.5% percentiles of the posterior samples. Results (not reported) show that they are likewise robust to the violation of the normality assumption of $\boldsymbol{\xi}_{ij}$, ϵ^{v}_{ij}, and $\boldsymbol{\epsilon}_{ij}$.

Simulation 2

To examine the performance of the modified DIC in the order selection of HMM, we consider five competing models M_2 to M_6, in which M_2 is the true model, whereas M_3 to M_6 are defined in the same manner as M_2 except that $S = 3$, 4, 5, and 6, respectively. We calculate the modified DIC under M_S $(S = 2, ..., 6)$ based on the 100 datasets generated in the setting of $\lambda_0(t) = 0.5$, $N = 200$, CR $= 30\%$, and Prior I in Simulation 1. Figure 1 depicts the boxplots of the DIC values under all competing models, which apparently support M_2. A further check indicates that the DIC value of M_2 is consistently the smallest in each replication, thereby indicating a good performance of DIC in order selection.

Simulation 3

The proposed joint model is also compared with the following two simpler models in terms of prediction accuracy:

Table 1. Summary of the Bayesian estimates under (N, CR) = (200, 30%).

	$\lambda_0(t) = 0.5$				$\lambda_0(t) = t + 0.2$				$\lambda_0(t) = t^2 + 0.3$			
Parameters in the conditional regression model (1)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
ρ	−.003	.127	−.037	.143	−.024	.138	−.050	.161	−.009	.132	−.027	.150
γ_1	.010	.047	−.010	.058	.015	.051	−.008	.060	.007	.050	−.005	.058
γ_2	.004	.083	.006	.102	.019	.083	−.006	.103	.003	.080	.012	.096
φ_1	.010	.065	.019	.084	.001	.065	.017	.082	.012	.061	.018	.077
φ_2	.010	.065	.016	.078	.012	.069	.020	.084	.013	.065	.024	.079
σ^2_s	.066	.081	.046	.084	.073	.088	.050	.089	.065	.080	.033	.072
Parameters in the transition model (2)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
τ	−.006		.197		.017		.191		−.014		.190	
α_1	−.000		.244		.022		.276		−.010		.247	
α_2	.010		.143		−.002		.146		.003		.143	
α_w	.014		.142		−.009		.155		.007		.137	
ζ_{11}	−.013		.230		−.019		.249		.007		.230	
ζ_{21}	−.023		.241		−.018		.242		.032		.254	
Parameters in the CFA model (3)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
μ_1	.005	.067	−.017	.074	.010	.067	−.013	.078	.010	.066	−.002	.078
μ_2	.005	.057	−.014	.070	.008	.058	−.013	.074	.008	.065	−.010	.074
μ_3	.006	.057	−.012	.069	.007	.061	−.017	.079	.014	.057	−.005	.073
μ_4	.002	.064	−.008	.074	.004	.065	−.012	.078	.005	.065	−.009	.075
μ_5	.002	.055	−.011	.069	.006	.058	−.013	.076	.003	.054	−.013	.074
μ_6	.003	.057	−.017	.069	.007	.053	−.017	.075	.003	.054	−.011	.073
λ_{21}	.009	.046	.011	.065	.001	.043	.017	.063	.008	.042	.015	.062
λ_{31}	.004	.044	.017	.066	.007	.047	.017	.065	.004	.041	.011	.064
λ_{52}	.005	.043	.015	.061	.009	.044	.024	.065	.015	.045	.022	.062
λ_{62}	.011	.044	.019	.063	.006	.043	.013	.063	.013	.041	.019	.061
ψ_1	.015	.033	.014	.049	.015	.033	.013	.047	.011	.031	.015	.047
ψ_2	.012	.029	.006	.043	.015	.031	.009	.047	.012	.029	.009	.045
ψ_3	.014	.029	.003	.041	.013	.030	.005	.045	.011	.029	.008	.044
ψ_4	.015	.033	.014	.050	.020	.037	.018	.048	.020	.036	.016	.049
ψ_5	.015	.030	.006	.043	.014	.030	.002	.043	.011	.027	.011	.044
ψ_6	.008	.027	.006	.043	.011	.030	.007	.047	.010	.028	.008	043
ϕ_{11}	−.028	.102	−.010	.132	−.014	.109	−.026	.131	−.017	.100	−.028	.125
ϕ_{12}	−.010	.067	−.006	.077	−.004	.076	.005	.078	−.013	.068	−.002	.074
ϕ_{22}	−.022	.108	−.024	.127	−.019	.106	−.021	.138	−.034	.102	−.029	.130
Parameters in the PH model (4)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
β_1	−.047		.167		−.006		.151		.002		.153	
β_2	−.035		.162		.028		.149		.005		.145	
η_1	.061		.182		−.005		.168		.011		.172	
η_2	−.067		.198		−.011		.169		.009		.166	
η_w	−.068		.176		.006		.169		.014		.163	
σ^2_w	−.064		.153		−.058		.150		−.053		.150	

Par: parameter; RMS: root mean square errors.

M_{ind} : the same as M_2 except that CFA Model (3) is excluded, and the multiple observed measurements $y_{ij1}, \ldots, y_{ij6}$ are directly included in models (1) and (4) as follows:

$$\left[v_{ij}|Z_{ij}=s\right]=\rho_s+\gamma_{s1}x_{ij1}+\gamma_{s2}x_{ij2}+\varphi_{s1}y_{ij1}+\cdots+\varphi_{s6}y_{ij6}+w_i+\epsilon^v_{ij}, \tag{10}$$

$$\lambda(t|\mathbf{f}_i,\mathbf{y}_{i1},w_i)=\lambda_0(t)\exp\left(\beta_1 f_{i1}+\beta_2 f_{i2}+\eta_1 y_{i11}+\cdots+\eta_6 y_{i16}+\eta_w w_i\right). \tag{11}$$

M_{fix} : the same as M_2 except that random effect w_i is excluded in models (1, 2) and (4):

$$\left[v_{ij}|Z_{ij}=s\right]=\rho_s+\gamma_{s1}x_{ij1}+\gamma_{s2}x_{ij2}+\varphi_{s1}\xi_{ij1}+\varphi_{s2}\xi_{ij2}+\epsilon^v_{ij}, \tag{12}$$

$$\text{logit}(\vartheta_{ijus})=\zeta_{us}+\boldsymbol{\alpha}^T\mathbf{d}_{ij}, \tag{13}$$

$$\lambda(t|\mathbf{f}_i,\boldsymbol{\xi}_{i1})=\lambda_0(t)\exp\left(\beta_1 f_{i1}+\beta_2 f_{i2}+\eta_1\xi_{i11}+\eta_2\xi_{i12}\right). \tag{14}$$

For each dataset generated in the setting of $\lambda_0(t)=0.5$, $N=500$, CR $=30\%$ in Simulation 1, we fit the three competing models with randomly selected 300

Table 2. Summary of the Bayesian estimates under (N, CR) = (500, 30%).

	$\lambda_0(t)=0.5$				$\lambda_0(t)=t+0.2$				$\lambda_0(t)=t^2+0.3$			
Parameters in the conditional regression model (1)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
ρ	−.003	.082	−.012	.094	−.008	.086	−.016	.096	−.002	.084	−.029	.100
γ_1	.006	.030	−.002	.037	.005	.031	−.004	.038	.007	.031	−.009	.037
γ_2	.005	.054	−.003	.063	.011	.056	−.008	.067	.018	.052	−.011	.066
φ_1	.016	.041	−.017	.052	.013	.043	.007	.048	.018	.042	.011	.048
φ_2	.016	.040	.014	.047	.012	.042	.007	.050	.018	.041	.006	.047
σ_s^2	.062	.068	.052	.068	.065	.073	.048	.069	.070	.076	.049	.067
Parameters in the transition model (2)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
τ	.001		.116		−.006		.128		.003		.116	
α_1	.012		.160		−.032		.160		−.021		.156	
α_2	.008		.089		.024		.093		.018		.091	
α_w	−.002		.088		.007		.094		.001		.084	
ζ_{11}	−.020		.147		−.035		.150		.024		.138	
ζ_{21}	−.040		.169		.033		.152		−.022		.153	
Parameters in the CFA model (3)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
μ_1	.005	.049	−.004	.051	.003	.042	−.010	.051	.003	.040	−.008	.050
μ_2	.007	.040	−.004	.046	.008	.037	−.009	.049	.005	.033	−.013	.045
μ_3	.006	.035	−.002	.047	.005	.038	−.015	.050	.001	.033	−.016	.044
μ_4	.006	.035	−.001	.048	.002	.041	−.003	.053	.005	.039	−.004	.050
μ_5	.005	.041	−.005	.046	.005	.036	−.005	.047	.007	.034	−.006	.046
μ_6	.006	.035	−.002	.046	.004	.037	−.003	.046	.004	.032	−.007	.049
λ_{21}	.002	.026	.011	.038	.007	.028	.014	.040	.004	.026	.006	.038
λ_{31}	.003	.027	.010	.039	.004	.027	.012	.039	.008	.026	.006	.039
λ_{52}	.005	.027	.007	.037	.006	.027	.010	.040	.005	.027	.009	.038
λ_{62}	.006	.028	.016	.039	.007	.028	.008	.039	.003	.027	.006	.035
ψ_1	.004	.022	.007	.030	.004	.022	.005	.031	.004	.020	.009	.032
ψ_2	.007	.019	.002	.028	.008	.019	.003	.029	.007	.017	.002	.028
ψ_3	.006	.018	.005	.028	.006	.018	.000	.029	.005	.017	.005	.028
ψ_4	.004	.021	.010	.034	.005	.020	.009	.032	.005	.020	.009	.030
ψ_5	.006	.019	.001	.028	.004	.018	−.000	.029	.007	.018	−.002	.030
ψ_6	.006	.018	.000	.027	.006	.018	.000	.029	.007	.019	.001	028
ϕ_{11}	−.011	.070	−.010	.082	−.017	.074	−.009	.086	−.019	.071	−.006	.078
ϕ_{12}	.001	.044	.009	.051	−.007	.049	.010	.055	.008	.044	.013	.051
ϕ_{22}	.006	.063	−.007	.083	−.010	.068	.004	.086	−.011	.067	.000	.081
Parameters in the PH model (4)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
β_1	−.003		.098		.028		.095		.012		.094	
β_2	−.011		.094		.021		.089		.022		.089	
η_1	.032		.110		−.034		.106		−.030		.104	
η_2	−.021		.105		.022		.103		.026		.107	
η_w	−.023		.108		−.037		.109		−.022		.102	
σ_w^2	−.019		.092		−.035		.100		−.033		.094	

Par: parameter; RMS: root mean square errors.

Table 3. Summary of the Bayesian estimates under (N, CR) = (200, 50%).

	$\lambda_0(t) = 0.5$				$\lambda_0(t) = t + 0.2$				$\lambda_0(t) = t^2 + 0.3$			
Parameters in the conditional regression model (1)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
ρ	−.028	.141	−.031	.148	−.029	.132	−.041	.159	−.036	.135	−.041	.153
γ_1	.005	.049	−.006	.060	.003	.048	−.014	.058	.013	.051	−.017	.065
γ_2	.008	.085	−.000	.105	.009	.085	−.008	.107	.011	.082	.001	.100
φ_1	.007	.063	.014	.082	.016	.064	.013	.081	.025	.063	.013	.077
φ_2	.008	.066	.019	.081	.004	.062	.016	.085	.003	.061	.013	.079
σ_s^2	.078	.091	.050	.086	.063	.079	.049	.082	.076	.090	.056	.087
Parameters in the transition model (2)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
τ	−.017		.201		.003		.188		.005		.192	
α_1	.010		.260		−.005		.261		.016		.236	
α_2	.009		.153		−.008		.154		−.000		.138	
α_w	−.003		.149		−.027		.155		−.001		.138	
ζ_{11}	−.001		.224		.015		.239		.004		.225	
ζ_{21}	−.008		.245		.027		.281		−.005		.242	
Parameters in the CFA model (3)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
μ_1	.011	.067	−.015	.075	.009	.070	−.013	.081	.018	.062	−.020	.076
μ_2	.012	.056	−.010	.074	.009	.056	−.011	.073	.016	.053	−.016	.071
μ_3	.011	.059	−.016	.074	.003	.058	−.015	.079	.015	.056	−.019	.074
μ_4	.015	.074	−.003	.075	.023	.067	−.016	.081	.018	.065	−.000	.080
μ_5	.015	.062	−.006	.071	.015	.055	−.017	.077	.017	.055	−.003	.076
μ_6	.012	.064	−.009	.073	.020	.057	−.014	.078	.017	.055	−.006	.077
λ_{21}	.008	.044	.064	.071	.008	.042	.014	.063	.005	.043	.014	.062
λ_{31}	.011	.043	.066	.072	.005	.044	.011	.061	.004	.042	.023	.061
λ_{52}	.011	.043	.062	.080	.010	.046	.017	.064	.010	.042	.012	.064
λ_{62}	.014	.045	.064	.078	.009	.043	.014	.060	.004	.041	.008	.059
ψ_1	.017	.035	.012	.048	.020	.036	.010	.049	.013	.033	.010	.045
ψ_2	.011	.027	.010	.045	.013	.031	.010	.045	.013	.029	.007	.045
ψ_3	.012	.029	.006	.046	.013	.033	.006	.045	.013	.028	.008	.046
ψ_4	.018	.036	.010	.049	.018	.036	.014	.046	.015	.034	.011	.048
ψ_5	.012	.030	.006	.045	.011	.028	.010	.045	.010	.028	.010	.044
ψ_6	.012	.029	.009	.046	.014	.030	.008	.044	.013	.029	.004	043
ϕ_{11}	−.025	.112	−.014	.133	−.009	.110	−.011	.131	−.013	.104	−.016	.130
ϕ_{12}	−.016	.073	.011	.085	.004	.068	.008	.078	−.007	.068	.001	.078
ϕ_{22}	−.036	.111	−.019	.131	−.016	.107	−.024	.134	−.029	.110	−.016	.127
Parameters in the PH model (4)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
β_1	−.040		.173		−.032		.176		−.027		.166	
β_2	−.056		.193		.018		.180		−.027		.161	
η_1	.069		.207		.033		.197		.046		.204	
η_2	−.057		.201		−.029		.207		−.060		.193	
η_w	−.072		.203		−.054		.198		.060		.196	
σ_w^2	−.033		.154		−.078		.154		−.074		.155	

Par: parameter; RMS: root mean square errors.

samples and conduct prediction on the remaining 200 samples. The hidden states of the test data are estimated based on (5). The AUC(t, Δ) and BS(t, Δ) are calculated at follow-up times $t = \{1, 1.5\}$ with $\Delta = \{1, 2\}$. Table 6 presents the average accuracy of estimated states and the means of AUC and BS values under each combination of t and Δ based on the 100 replicated datasets. M_2 performs better than M_{ind} and M_{fix} in terms of discrimination and prediction, and the prediction errors of M_2 are smaller than or comparable to those of M_{ind} and M_{fix}. The lowest estimation accuracy of Z_{ij}s under M_{ind} may arise from the multicollinearity of the predictors in (10), which plays a deterministic role in state estimation.

The computer codes for conducting parameter estimation and prediction are available at http://www.sta.cuhk.edu.hk/xysong/JHMM/.

Application

We applied the proposed method to the general health survey study described in the Introduction. The 14-year prospective health survey, commencing in 2001, gathered the participants' information regarding the

Table 4. Summary of the Bayesian estimates under (N, CR) = (500, 50%).

	$\lambda_0(t) = 0.5$				$\lambda_0(t) = t + 0.2$				$\lambda_0(t) = t^2 + 0.3$			
Parameters in the conditional regression model (1)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
ρ	−.004	.084	−.021	.096	−.010	.084	−.021	.096	−.009	.081	−.025	.097
γ_1	.006	.032	−.005	.039	−.000	.030	−.003	.039	.004	.030	−.012	.039
γ_2	.014	.052	−.012	.066	.008	.056	.001	.068	.010	.050	−.004	.065
φ_1	.018	.045	−.011	.050	.017	.042	.013	.055	.002	.043	.015	.052
φ_2	.022	.044	.003	.052	.017	.042	.004	.050	.014	.042	.006	.051
σ_s^2	.070	.076	.056	.072	.066	.072	.044	.064	.068	.074	.043	.065
Parameters in the transition model (2)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
τ	−.008		.129		.001		.124		−.001		.120	
α_1	−.024		.163		−.013		.165		−.021		.157	
α_2	.015		.098		.017		.095		.017		.091	
α_w	−.005		.089		.007		.094		.012		.086	
ζ_{11}	.024		.153		.022		.168		.022		.143	
ζ_{21}	−.021		.157		−.005		.186		−.015		.169	
Parameters in the CFA model (3)	State 1		State 2		State 1		State 2		State 1		State 2	
Par	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS	Bias	RMS
μ_1	.002	.042	−.005	.050	.002	.047	−.007	.057	.006	.041	−.006	.051
μ_2	.002	.036	−.008	.048	.004	.036	−.009	.047	.006	.035	−.011	.052
μ_3	.005	.037	−.009	.048	.005	.035	−.009	.048	.005	.034	−.009	.049
μ_4	.007	.044	−.008	.051	.008	.041	−.004	.048	.003	.041	−.004	.048
μ_5	.006	.037	−.006	.047	.005	.036	−.003	.046	.005	.035	−.010	.047
μ_6	.006	.038	−.008	.047	.009	.036	−.005	.047	.005	.036	−.008	.045
λ_{21}	.004	.029	.015	.042	.007	.028	.015	.041	.005	.028	.015	.040
λ_{31}	.004	.027	.013	.042	.005	.029	.016	.043	.004	.026	.014	.040
λ_{52}	.004	.027	.013	.040	.004	.028	.012	.040	.004	.028	.016	.039
λ_{62}	.007	.027	.012	.041	.005	.028	.010	.037	.004	.027	.010	.039
ψ_1	.007	.022	.011	.033	.006	.021	.010	.033	.005	.021	.008	.030
ψ_2	.007	.019	−.000	.028	.004	.018	.003	.031	.006	.018	−.000	.027
ψ_3	.007	.019	.004	.029	.007	.020	.005	.031	.007	.018	.006	.029
ψ_4	.007	.022	.006	.032	.004	.020	.004	.031	.005	.021	.007	.031
ψ_5	.007	.018	−.002	.028	.007	.019	.003	.029	.007	.018	−.001	.029
ψ_6	.004	.018	.003	.028	.003	.018	.001	.029	.005	.018	.004	027
ϕ_{11}	−.012	.074	−.004	.084	−.016	.067	−.012	.084	−.013	.069	−.021	.083
ϕ_{12}	−.003	.044	.009	.055	−.005	.044	.009	.052	−.007	.045	.013	.051
ϕ_{22}	−.001	.067	−.008	.081	−.013	.069	.003	.084	−.002	.067	−.013	.081
Parameters in the PH model (4)												
Par	Bias		RMS		Bias		RMS		Bias		RMS	
β_1	−.026		.106		.003		.106		.006		.102	
β_2	−.013		.108		.002		.103		.003		.101	
η_1	.042		.127		−.004		.121		−.015		.112	
η_2	−.034		.127		.018		.122		.005		.118	
η_w	−.032		.117		.010		.125		−.005		.114	
σ_w^2	−.025		.095		−.035		.101		−.032		.101	

Par: parameter; RMS: root mean square errors.

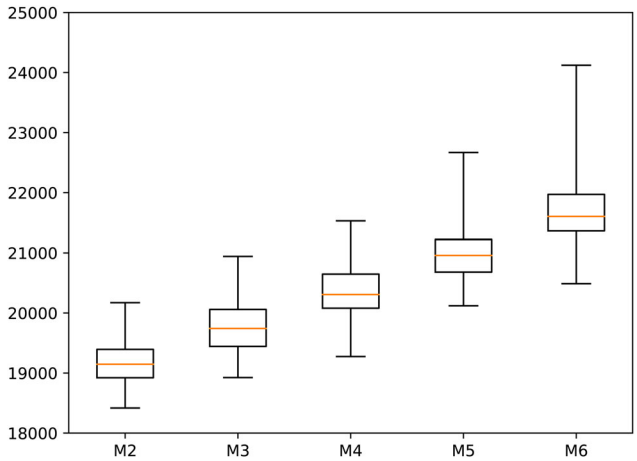

Figure 1. DIC values for comparison in Simulation 2.

general and medical information, cognitive assessment, and physical function through questionnaires, measurements, and self-reports. The primary goal of this survey is to investigate health-related issues, such as osteoporosis, cognitive impairment, and death, for Chinese elderly people, so as to facilitate effective prevention strategies. In this study, we aimed to jointly investigate cognitive impairment and mortality, their common or distinct risk factors, and possibly heterogeneous predictor effects on the progression of cognitive impairment. Cognitive impairment was assessed through CMMSE score ranging from 0 to 30 with a lower CMMSE score representing more dementia-related cognitive impairment. The time to death was

Table 5. Summary of the Bayesian estimates in real data.

Par	Est	SE	95% HPD	Par	Est	SE	95% HPD
Parameters in the conditional regression model (16)							
State 1				State 2			
ρ_1	−0.178	0.053	(−0.277, −0.072)	ρ_2	0.291	0.044	(0.207, 0.373)
γ_{11}	0.103	0.039	(0.024, 0.175)	γ_{21}	0.014	0.029	(−0.041, 0.070)
γ_{12}	−0.032	0.040	(−0.110, 0.046)	γ_{22}	0.040	0.044	(−0.042, 0.126)
φ_{11}	0.430	0.093	(0.261, 0.619)	φ_{21}	0.154	0.081	(−0.003, 0.313)
φ_{12}	0.116	0.061	(−0.003, 0.237)	φ_{22}	0.526	0.240	(0.040, 0.987)
σ_1^2	0.770	0.067	(0.657, 0.921)	σ_2^2	0.198	0.021	(0.164, 0.253)
Parameters in the transition model (17)							
τ_1	0.686	0.178	(0.274, 0.981)	ζ_{11}	1.651	0.672	(0.281, 2.901)
ζ_{21}	−2.329	0.450	(−3.298, −1.511)	α_1	1.160	0.497	(0.203, 1.197)
α_2	0.983	0.309	(0.421, 1.571)	α_3	0.735	0.571	(−0.368, 1.770)
α_4	−0.400	0.509	(−1.358, 0.633)	α_5	0.365	0.391	(−0.341, 1.141)
α_6	0.497	0.364	(−0.201, 1.130)	α_7	−0.351	0.362	(−1.069, 0.319)
α_8	0.615	0.403	(−0.114, 1.397)	α_9	1.123	0.489	(0.204, 2.125)
α_w	−0.334	0.429	(−1.100, 0.655)				
Parameters in the CFA model (15)							
	State 1				State 2		
μ_{11}	−0.219	0.039	(−0.294, −0.141)	μ_{21}	0.431	0.068	(0.294, 0.570)
μ_{12}	−0.279	0.048	(−0.374, −0.187)	μ_{22}	0.547	0.047	(0.443, 0.633)
μ_{13}	−0.147	0.037	(−0.220, −0.075)	μ_{23}	0.293	0.074	(0.150, 0.444)
μ_{14}	−0.218	0.054	(−0.336, −0.118)	μ_{24}	0.416	0.060	(0.378, 0.456)
μ_{15}	0.307	0.061	(0.188, 0.432)	μ_{25}	−0.591	0.023	(−0.637, −0.547)
$\lambda_{1,21}$	0.572	0.084	(0.398, 0.736)	$\lambda_{2,21}$	0.156	0.072	(0.011, 0.296)
$\lambda_{1,31}$	0.715	0.111	(0.505, 0.925)	$\lambda_{2,31}$	0.675	0.226	(0.278, 1.191)
$\lambda_{1,52}$	−0.989	0.210	(−1.380, −0.647)	$\lambda_{2,52}$	−0.415	0.149	(−0.705, −0.117)
ψ_{11}	0.468	0.073	(0.327, 0.610)	ψ_{21}	0.443	0.073	(0.310, 0.592)
ψ_{12}	0.997	0.061	(0.877, 1.117)	ψ_{22}	0.214	0.029	(0.163, 0.278)
ψ_{13}	0.670	0.053	(0.564, 0.773)	ψ_{23}	0.825	0.086	(0.686, 1.030)
ψ_{14}	0.675	0.145	(0.391, 0.943)	ψ_{24}	0.070	0.007	(0.057, 0.086)
ψ_{15}	0.568	0.135	(0.293, 0.797)	ψ_{25}	0.116	0.011	(0.094, 0.139)
$\phi_{1,11}$	0.506	0.081	(0.350, 0.661)	$\phi_{2,11}$	0.329	0.072	(0.192, 0.473)
$\phi_{1,12}$	−0.023	0.037	(−0.101, 0.048)	$\phi_{2,12}$	−0.007	0.012	(−0.031, 0.017)
$\phi_{1,22}$	0.655	0.149	(0.403, 0.941)	$\phi_{2,22}$	0.052	0.006	(0.040, 0.066)
Parameters in the PH model (18)							
β_1	−0.067	0.056	(−0.178, 0.037)	β_2	−0.104	0.066	(−0.235, 0.021)
β_3	−1.122	0.167	(−1.470, −0.841)	β_4	0.346	0.071	(0.210, 0.483)
β_5	0.151	0.210	(−0.236, 0.530)	β_6	−0.127	0.170	(−0.427, 0.209)
β_7	0.146	0.132	(−0.093, 0.416)	β_8	−0.019	0.164	(−0.338, 0.284)
β_9	−0.063	0.170	(−0.301, 0.250)	β_{10}	0.434	0.136	(0.181, 0.684)
β_{11}	0.173	0.164	(−0.144, 0.486)	η_1	−0.073	0.137	(−0.343, 0.202)
η_2	0.110	0.143	(−0.156, 0.411)	η_w	−0.302	0.143	(−0.564, −0.037)
σ_w^2	0.293	0.034	(0.229, 0.362)				

calculated as the period from enrollment to the date of the clinical endpoint or the documented mortality record from the Hong Kong Death Registry, whichever came first. We excluded the participants whose information is partially missing or whose education level is below primary school by considering the clinical validness range of CMMSE, yielding a cohort of 484 participants with a broad range of information for statistical analysis. The number of follow-up visits varied from 2 to 4, corresponding to visiting times at baseline, 4th year, 7th year, and 14th year. The CR of mortality is approximately 44.4%. The dataset includes a number of variables, such as age, gender (1 = female), smoking habit (1 = current smoker), alcohol intake (1 = drinker), marital status (1 = married), socioeconomic status assessed by self-mark on a ladder with 10 rungs, which is coded by a dummy variable "soce" (1 = status not less than 5), education level (1 = high level), hypertension history (1 = ever had hypertension), diabetes history (1 = ever had diabetes), body mass index (BMI), physical activity levels assessed by the Physical Activity Scale of Elderly (PASE), and multiple measurements, namely, PCS, 6MPACE, GRIPMAX, MCS, and GDS, for measuring physical and mental functions.

We considered a CFA model to group $\{y_{ij1}, y_{ij2}, y_{ij3}\} = \{\text{6MPACE, PCS, GRIPMAX}\}$ into a latent variable "physical function (ξ_{ij1})" and $\{y_{ij4}, y_{ij5}\} = \{\text{MCS, GDS}\}$ into a latent variable "mental health (ξ_{ij2})" based on their meanings and experts' suggestions as follows:

$$\left[\mathbf{y}_{ij}|Z_{ij}=s\right] = \boldsymbol{\mu}_s + \boldsymbol{\Lambda}_s\boldsymbol{\xi}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad \text{with} \quad \boldsymbol{\Lambda}_s^T = \begin{bmatrix} 1^* & \lambda_{s,21} & \lambda_{s,31} & 0^* & 0^* \\ 0^* & 0^* & 0^* & 1^* & \lambda_{s,52} \end{bmatrix}, \tag{15}$$

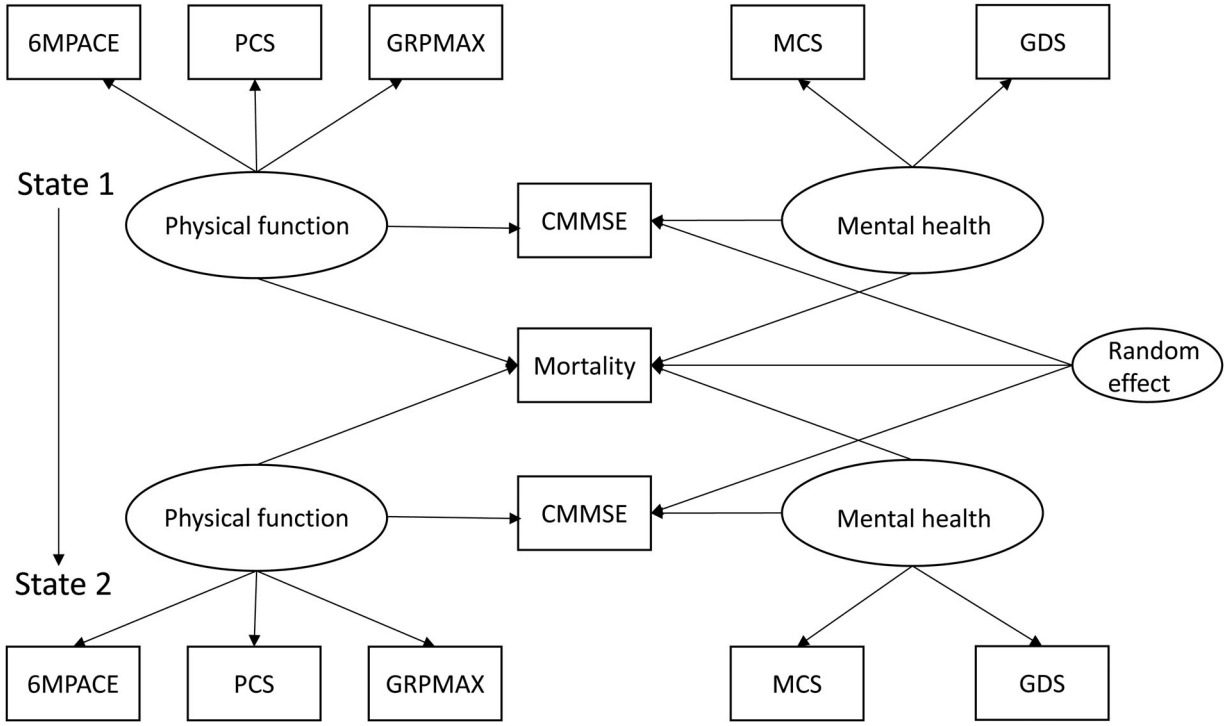

Figure 2. Path diagram of the proposed model: The rectangles represent the observed responses or fixed covariates, and the ellipses denote the unobserved latent factors or random effects. The arrows identify the direct effect.

where $\mathbf{y}_{ij} = (y_{ij1}, ..., y_{ij5})^T$, $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \xi_{ij2})^T$, and the elements with asterisk in $\boldsymbol{\Lambda}_s$ were fixed to obtain an identified model and clear interpretation of the latent variables.

Then, we assessed state-specific covariate effects on cognitive impairment as follows:

$$[v_{ij}|Z_{ij} = s] = \rho_s + \gamma_{s1}x_{ij1} + \gamma_{s2}x_{ij2} + \varphi_{s1}\xi_{ij1} + \varphi_{s2}\xi_{ij2} + w_i + \epsilon^v_{ij}, \tag{16}$$

where v_{ij} = CMMSE score, x_{ij1} = PASE, x_{ij2} = BMI, ξ_{ij1} = physical function, and ξ_{ij2} = mental health.

We included baseline individual characteristics, such as gender (d_{ij1}), age at enrollment (d_{ij2}), smoking habit (d_{ij3}), alcohol intake (d_{ij4}), soce (d_{ij5}), marital status (d_{ij6}), education level (d_{ij7}), hypertension history (d_{ij8}), and diabetes history (d_{ij9}), into the transition model as follows:

$$\text{logit}(\vartheta_{ijus}) = \zeta_{us} + \alpha_1 d_{ij1} + \alpha_2 d_{ij2} + \cdots + \alpha_9 d_{ij9} + \alpha_w w_i. \tag{17}$$

Lastly, we examined the effects of all the aforementioned variables at baseline on the hazards of mortality as follows:

$$\lambda(t|\mathbf{f}_i, \boldsymbol{\xi}_{i1}, w_i) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\mathbf{f}_i + \boldsymbol{\eta}^T\boldsymbol{\xi}_{i1} + \eta_w w_i), \tag{18}$$

where $\mathbf{f}_i$ = (PASE, BMI, gender, age, smoking habit, alcohol intake, soce, marital status, education level, hypertension history, diabetes history)^T, $\boldsymbol{\xi}_{i1}$ = (physical function, mental health)^T, and all the elements of $\mathbf{f}_i$ and $\boldsymbol{\xi}_{i1}$ are taken as the observations at baseline.

Figure 2 presents the path diagram of the proposed model to visualize the complex associations between the observed and latent variables. For clarity, the covariates in $\mathbf{x}_{ij}, \mathbf{d}_{ij}$, and $\mathbf{f}_i$ are excluded in Figure 2. In the data analysis, the hyperparameters involved in the prior distributions of the unknown parameters were assigned in the same manner as the Prior I of Simulation 1. We first conducted the order selection of HMM using the modified DIC. Let M_S denote the proposed model of order S. In the posterior sampling, we imposed the identifiability constraint $\rho_1 < \cdots < \rho_S$ to solve the label switching problem. Figure S2 of the Online Appendix indicated that the MCMC algorithm converged rapidly. To be conservative, we discarded 6000 burn-in iterations and collected the subsequent 6000 posterior samples to obtain Bayesian results. The DIC values corresponding to M_1 to M_6 were 24,151, 22,424, 23,219, 25,089, 25,164, and 25,372, respectively. Hence, M_2 with the smallest DIC was selected. Table 5 presents the Bayesian estimates (Est), SEs, and the 95% highest posterior density

(HPD) intervals (Chen & Shao, 1999) of the unknown parameters in M_2. We obtained the following findings.

In the conditional regression model (16), the estimates of ρ_1 and ρ_2 $[\hat{\rho}_1 = -0.178(0.053), \hat{\rho}_2 = 0.291(0.044)]$ are ranked in an ascending order, implying that subjects in state 2 have higher CMMSE scores than those in state 1. Hence, states 1 and 2 can be interpreted as cognitive impairment and cognitive normal, respectively. Moreover, physical activity is positively associated with cognitive function in state 1 $[\hat{\gamma}_{11} = 0.103(0.039)]$, which in lines with the previous finding that regular physical activity could represent an important and potent protective factor of cognitive decline for patients with cognitive impairment (Laurin et al., 2001; Lautenschlager et al., 2008). Mental health is positively associated with cognitive function in state 2 $[\widehat{\varphi_{22}} = 0.526(0.240)]$. Published reports (e.g., Rock et al., 2014) also showed that keeping good mental health would enhance cognitive ability for elders of cognitive normal. Furthermore, physical function exhibits a positive effect on CMMSE score in state 1 $[\hat{\varphi}_{11} = 0.430(0.093)]$, which agrees with previous discoveries and suggests that physical frailty is associated with increased prevalence and incidence of cognitive impairment (Auyeung et al., 2008; Feng et al., 2017).

In the transition model (17), gender, age, and diabetes history are all positively associated with the probability of transition to a state rather than a better one $[\hat{\alpha}_1 = 1.160(0.497), \hat{\alpha}_2 = 0.983(0.309), \hat{\alpha}_9 = 1.123(0.489)]$. Therefore, female, elder people, and those who ever had diabetes are at higher risk of progression to cognitive impairment. Such gender effect and the accelerated cognitive dysfunction contributed by the diabetes have been demonstrated by existing studies (Roberts et al., 2014; Yaffe et al., 2012; Zhang, 2006). Our results also show that smoking habit, alcohol intake, social ladder, married status, education level, and hypertension history have insignificant effects on the transition probability $[\hat{\alpha}_3 = 0.735(0.571), \hat{\alpha}_4 = -0.400(0.509), \hat{\alpha}_5 = 0.365(0.391), \hat{\alpha}_6 = 0.497(0.364), \hat{\alpha}_7 = -0.351(0.362), \hat{\alpha}_8 = 0.615(0.403)]$. Besides, the random effect $[\hat{\alpha}_w = -0.334(0.429)]$ is nonsignificant, indicating that the transition process does not exhibit heterogeneity.

In the CFA model (15), the mean levels of y_1, y_2, and y_3 in state 1 $[\hat{\mu}_{11} = -0.219(0.039), \hat{\mu}_{12} = -0.279(0.048), \hat{\mu}_{13} = -0.147(0.037)]$ are significantly smaller than those in state 2 $[\hat{\mu}_{21} = 0.431(0.068), \hat{\mu}_{22} = 0.547(0.047), \hat{\mu}_{23} = 0.293(0.074)]$. On the average, elders of cognitive impairment have lower scores (worse capability) than elders of cognitive normal in various physical functions. Meanwhile, the mean level of y_4 in state 1 $[\hat{\mu}_{14} = -0.218(0.054)]$ is significantly smaller than that in state 2 $[\hat{\mu}_{24} = 0.416(0.060)]$, whereas the mean level of y_5 in state 1 $[\hat{\mu}_{15} = 0.307(0.061)]$ is significantly larger than that in state 2 $[\hat{\mu}_{25} = -0.591(0.023)]$. This result indicates that elders of cognitive impairment tend to have worse mental health and more serious depression compared with elders of cognitive normal. Furthermore, all the estimated factor loadings are significantly different from zero in both states. Thus, physical and mental functions are properly characterized by their observed indictors in both states. However, the magnitudes of the estimated factor loadings are larger in state 1 than in state 2, thereby implying that {6MPACE, PCS, GRIPMAX} and {MCS, GDS} play more pronounced roles in the characterization of "physical function" and "mental health" in cognitive impairment state than in cognitive normal state.

In the PH model (18), age and hypertension history $[\hat{\beta}_4 = 0.346(0.071), \hat{\beta}_{10} = 0.434(0.136)]$ have significantly positive effects on the hazards of mortality. Elder people and those who ever had hypertension are at higher risk of death, which is consistent with the previous finding (Dahlöf et al., 2002). By contrast, gender $[\hat{\beta}_3 = -1.123(0.167)]$ exerts negative effects on the hazards of mortality, suggesting that female have lower mortality risk than male. These results agree with the published reports (Gove, 1973). The significant random effect $[\hat{\eta}_w = -0.302(0.143)]$ also illustrates a negative association between the longitudinal and time-to-event outcomes.

Furthermore, the variance of the random effect is significantly different from zero $[\hat{\sigma}_w^2 = 0.293(0.034)]$, thereby reconfirming the dependency of the longitudinal observations.

To check whether the estimation results are sensitive to the prior inputs, we reconducted the preceding analysis by using the Prior II of Simulation 1. The obtained results are similar and not reported.

We also compared the proposed joint model M_2 with two simpler models in terms of the out-of-sample predictive performance. The two simpler model M_{ind} and M_{fix} were defined in the same manner as in Simulation 3. We focused on the time frame with $t = 5$ and 6 years and $\Delta = 1$ and 2 years in predicting the survival probability $\hat{\pi}(t + \Delta | t)$. The full dataset was randomly split into a training set with 300 subjects and a test set with 184 subjects. The random split was repeated 100 times. For each split, we fitted M_2, M_{ind}, and M_{fix} to the training set and then

Table 6. The average accuracy of estimated states and means of integrated AUC and BS under M_2, M_{ind}, and M_{fix} in Simulation 3 and the real data example.

Results in Simulation 3									
	(1)		(1, 2)		(1.5,1)		(1.5,2)		Acc
	AUC	BS	AUC	BS	AUC	BS	AUC	BS	
M_2	0.814	0.114	0.829	0.215	0.796	0.163	0.815	0.283	91.94%
M_{ind}	0.771	0.100	0.808	0.215	0.775	0.160	0.807	0.299	44.31%
M_{fix}	0.749	0.128	0.767	0.232	0.733	0.173	0.753	0.296	91.91%
Results in real data example									
	(5, 1)		(5, 2)		(6, 1)		(6, 2)		Acc
	AUC	BS	AUC	BS	AUC	BS	AUC	BS	
M_2	0.752	0.017	0.714	0.060	0.704	0.046	0.715	0.073	
M_{ind}	0.644	0.050	0.628	0.093	0.630	0.076	0.632	0.106	
M_{fix}	0.724	0.018	0.681	0.062	0.661	0.048	0.655	0.074	

calculated AUC and BS based on the test set. The accuracy of the estimated hidden states cannot be compared because elders' diagnosed cognitive status is unavailable in the dataset. Table 6 presents the averages of the AUC and BS values on the basis of 100 random splits for the three competing models. The AUC values of M_2 are consistently higher than those of M_{ind} and M_{fix} in all combinations of t and Δ, indicating that the proposed model has better discriminative capability than the simpler models M_{ind} and M_{fix} in the prediction of death. Meanwhile, the values of BS of M_2 are smaller than those of M_{ind} and M_{fix}, suggesting a more stable prediction performance of M_2 than the simpler models. We also noticed from Table 6 that the AUC values of M_2 and M_{ind} did not differ remarkably in certain cases. This result is not surprising because the existing literature (Kutner et al., 2005; Weiss & Hassett, 1999) revealed that the presence of multicollinearity in predictors may not affect the model's ability to accurately predict the response variable.

Discussion

We proposed a novel joint model to simultaneously analyze longitudinal and time-to-event data in the presence of latent variables. The proposed model encompasses the factor analytic technique, non-homogeneous HMM, and the PH model into an integrated framework. The longitudinal response and time-to-event outcome are allowed to be correlated through a shared subject-specific random effect and shared latent risk factors. We developed a fully Bayesian approach with a hybrid MCMC algorithm to conduct parameter estimation, order selection of HMM, and out-of-sample predication.

This study has limitations. First, the proposed model specified the parametric forms of predictor effects a priori. This parametric assumption may be restrictive in substantive studies. Generalizing the current model to a semiparametric or nonparametric context can enhance model flexibility and analytic power, but its development requires further investigation. Second, considering that missing data are common in longitudinal settings, accommodating missing data, especially missing not at random data, in the context of the proposed model is of our future research interest. Last, the participants in this cohort study were community-dwelling instead of being recruited by random population sampling. The selection bias should be taken into consideration owing to the relatively non-frail and more health-conscious sample (Auyeung et al., 2011). However, the aforementioned extensions require substantial efforts in the future.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was supported by GRF grants 14301918, 14302519, and Earmarked Grant CUHK4101/02 M from the Research Grant Council of the Hong Kong Special Administrative Region, the National Institutes of Health R01 grant AR04943901A1, and the direct grants of the Chinese University of Hong Kong.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design

and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank the Editor, associate editor, and two referees for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institution is not intended and should not be inferred.

References

Altman, R. M. (2007). Mixed hidden markov models: An extension of the hidden markov model to the longitudinal data setting. *Journal of the American Statistical Association*, *102*(477), 201–210. https://doi.org/10.1198/016214506000001086

Andrinopoulou, E.-R., Eilers, P. H., Takkenberg, J. J., & Rizopoulos, D. (2018). Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using p-splines. *Biometrics*, *74*(2), 685–693. https://doi.org/10.1111/biom.12814

Auyeung, T. W., Kwok, T., Lee, J., Leung, P. C., Leung, J., & Woo, J. (2008). Functional decline in cognitive impairment–the relationship between physical and cognitive function. *Neuroepidemiology*, *31*(3), 167–173. https://doi.org/10.1159/000154929

Auyeung, T. W., Lee, J., Kwok, T., & Woo, J. (2011). Physical frailty predicts future cognitive decline—A four-year prospective study in 2737 cognitively normal older adults . *The Journal of Nutrition, Health & Aging*, *15*(8), 690–694. https://doi.org/10.1007/s12603-011-0110-9

Barrett, J. K., & Su, L. (2017). Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Statistics in Medicine*, *36*(9), 1447–1460. https://doi.org/10.1002/sim.7209

Bartolucci, F., & Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent markov heterogeneity structure. *Journal of the American Statistical Association*, *104*(486), 816–831. https://doi.org/10.1198/jasa.2009.0107

Bartolucci, F., & Farcomeni, A. (2019). A shared-parameter continuous-time hidden markov and survival model for longitudinal data with informative dropout. *Statistics in Medicine*, *38*(6), 1056–1073. https://doi.org/10.1002/sim.7994

Brown, E. R., & Ibrahim, J. G. (2003). A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, *59*(2), 221–228. https://doi.org/10.1111/1541-0420.00028

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*(4), 651–673. https://doi.org/10.1214/06-BA122

Chen, M.-H., & Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, *8*(1), 69–92. https://doi.org/10.2307/1390921

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.2517-6161.1972.tb00899.x.

Cox, D. R., & Hinkley, D. V. (1979). *Theoretical statistics*. Chapman and Hall/CRC.

Dahlöf, B., Devereux, R. B., Kjeldsen, S. E., Julius, S., Beevers, G., de Faire, U., Fyhrquist, F., Ibsen, H., Kristiansson, K., Lederballe-Pedersen, O., Lindholm, L. H., Nieminen, M. S., Omvik, P., Oparil, S., & Wedel, H. (2002). Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (life): A randomised trial against atenolol. *The Lancet*, *359*(9311), 995–1003. https://doi.org/10.1016/S0140-6736(02)08089-3

Emura, T., Nakatochi, M., Matsui, S., Michimae, H., & Rondeau, V. (2018). Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: Meta-analysis with a joint model. *Statistical Methods in Medical Research*, *27*(9), 2842–2858. https://doi.org/10.1177/0962280216688032

Feng, L., Nyunt, M. S. Z., Gao, Q., Feng, L., Lee, T. S., Tsoi, T., Chong, M. S., Lim, W. S., Collinson, S., & Yap, P. (2017). Physical frailty, cognitive impairment, and the risk of neurocognitive disorder in the singapore longitudinal ageing studies. *The Journals of Gerontology: Series A*, *72*(3), 369–375. https://doi.org/10.1093/gerona/glw050

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, *96*(453), 194–209. https://doi.org/10.1198/016214501750333063

Gove, W. R. (1973). Sex, marital status, and mortality. *American Journal of Sociology*, *79*(1), 45–67. https://doi.org/10.1086/225505

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, *18*(17-18), 2529–2545. https://doi.org/10.1002/(sici)1097-0258(19990915/30)18:17/18¡2529::aid-sim274¿3.0.co;2-5

Gruttola, D., & Tu, X. M. (1994). Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, *50*(4), 1003–1014. https://doi.org/10.2307/2533439

Hashemi, R., Jacqmin-Gadda, H., & Commenges, D. (2003). A latent process model for joint modeling of events and marker. *Lifetime Data Analysis*, *9*(4), 331–343. https://doi.org/10.1023/B:LIDA.0000012420.36627.a6

He, H., Pan, D., Song, X., & Sun, L. (2018). Additive mean residual life model with latent variables under right censoring. *Statistica Sinica*, *29*, 47–66. https://doi.org/10.5705/ss.202015.0369

Ibrahim, J. G., Chen, M., & Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, *14*(3), 863–883.

Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2014). *Bayesian survival analysis*. Wiley Online Library. https://doi.org/10.1002/9781118445112.stat06003

Jacqmin-Gadda, H., Commenges, D., & Dartigues, J.-F. (2006). Random change point model for joint modeling

of cognitive decline and dementia. *Biometrics*, *62*(1), 254–260. https://doi.org/10.1111/j.1541-0420.2005.00443.x

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models, volume 5*. McGraw-Hill Irwin.

Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*, *60*(1), 85–92. doi:j.0006-341X.2004.00141.x. https://doi.org/10.1111/j.0006-341X.2004.00141.x

Laurin, D., Verreault, R., Lindsay, J., MacPherson, K., & Rockwood, K. (2001). Physical activity and risk of cognitive impairment and dementia in elderly persons. *Archives of Neurology*, *58*(3), 498–504. https://doi.org/10.1001/archneur.58.3.498

Lautenschlager, N. T., Cox, K. L., Flicker, L., Foster, J. K., van Bockxmeer, F. M., Xiao, J., Greenop, K. R., & Almeida, O. P. (2008). Effect of physical activity on cognitive function in older adults at risk for alzheimer disease: A randomized trial. *JAMA*, *300*(9), 1027–1037. https://doi.org/10.1001/jama.300.9.1027

Li, K., & Luo, S. (2017). Functional joint model for longitudinal and time-to-event data: An application to Alzheimer's disease. *Statistics in Medicine*, *36*(22), 3560–3572. https://doi.org/10.1002/sim.7381

Maruotti, A. (2011). Mixed hidden markov models for longitudinal data: An overview. *International Statistical Review*, *79*(3), 427–454. https://doi.org/10.1111/j.1751-5823.2011.00160.x

Ouyang, M., Wang, X., Wang, C., & Song, X. (2018). Bayesian semiparametric failure time models for multivariate censored data with latent variables. *Statistics in Medicine*, *37*(28), 4279–4297. https://doi.org/10.1002/sim.7916

Pan, D., He, H., Song, X., & Sun, L. (2015). Regression analysis of additive hazards model with latent variables. *Journal of the American Statistical Association*, *110*(511), 1148–1159. https://doi.org/10.1080/01621459.2014.950083

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, *67*(3), 819–829. https://doi.org/10.1111/j.1541-0420.2010.01546.x

Rizopoulos, D., Verbeke, G., Lesaffre, E., & Vanrenterghem, Y. (2008). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, *64*(2), 611–619. doi:j.1541-0420.2007.00894.x. https://doi.org/10.1111/j.1541-0420.2007.00894.x

Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., Baertlein, L., Boeve, B. F., Tangalos, E. G., Ivnik, R. J., Mielke, M. M., & Petersen, R. C. (2014). Association of diabetes with amnestic and nonamnestic mild cognitive impairment. *Alzheimer's & Dementia*, *10*(1), 18–26. https://doi.org/10.1016/j.jalz.2013.01.001

Rock, P., Roiser, J., Riedel, W., & Blackwell, A. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. https://doi.org/10.1017/S0033291713002535

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

Song, X., Xia, Y., & Zhu, H. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics*, *73*(1), 313–323. https://doi.org/10.1111/biom.12536

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353

Tsiatis, A. A., Degruttola, V., & Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and cd4 counts in patients with AIDS. *Journal of the American Statistical Association*, *90*(429), 27–37. https://doi.org/10.1080/01621459.1995.10476485

Weiss, N. A., & Hassett, M. J. (1999). *Introductory statistics*. Addison-Wesley.

Yaffe, K., Falvey, C., Hamilton, N., Schwartz, A. V., Simonsick, E. M., Satterfield, S., Cauley, J. A., Rosano, C., Launer, L. J., Strotmeyer, E. S., & Harris, T. B. (2012). Diabetes, glucose control, and 9-year cognitive decline among older adults without dementia. *Archives of Neurology*, *69*(9), 1170–1175. https://doi.org/10.1001/archneurol.2012.1117

Zhang, Z. (2006). Gender differentials in cognitive impairment and decline of the oldest old in China. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *61*(2), S107–S115. https://doi.org/10.1093/geronb/61.2.S107