SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
2019, Vol. 28(7) 2112–2124

DOI: 10.1177/0962280217748675
journals.sagepub.com/home/smm
SAGE

Bayesian hidden Markov models for delineating the pathology of Alzheimer's disease

Kai Kang,[1] Jingheng Cai,[2] Xinyuan Song[1,3] and Hongtu Zhu[4]

Abstract

Alzheimer's disease is a firmly incurable and progressive disease. The pathology of Alzheimer's disease usually evolves from cognitive normal, to mild cognitive impairment, to Alzheimer's disease. The aim of this paper is to develop a Bayesian hidden Markov model to characterize disease pathology, identify hidden states corresponding to the diagnosed stages of cognitive decline, and examine the dynamic changes of potential risk factors associated with the cognitive normal–mild cognitive impairment–Alzheimer's disease transition. The hidden Markov model framework consists of two major components. The first one is a state-dependent semiparametric regression for delineating the complex associations between clinical outcomes of interest and a set of prognostic biomarkers across neurodegenerative states. The second one is a parametric transition model, while accounting for potential covariate effects on the cross-state transition. The inter-individual and inter-process differences are taken into account via correlated random effects in both components. Based on the Alzheimer's Disease Neuroimaging Initiative data set, we are able to identify four states of Alzheimer's disease pathology, corresponding to common diagnosed cognitive decline stages, including cognitive normal, early mild cognitive impairment, late mild cognitive impairment, and Alzheimer's disease and examine the effects of hippocampus, age, gender, and APOE-ϵ4 on degeneration of cognitive function across the four cognitive states.

Keywords

Bayesian P-splines, correlated random effects, hidden Markov models, MCMC methods, semiparametric models

1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease that usually starts slowly and worsens over time. The most common early symptom of AD is short-term memory loss, also referred to mild cognitive impairment (MCI). Patients at MCI state have high likelihood to transit to dementia or AD within a few years.[1] Despite an increasing attention to its growing public threat, the cause of AD remains poorly understood. Thus, it is of great interest to discover or validate prognostic biomarkers that may identify subjects at great risk for future cognitive decline and investigate the functional effects of various biomarkers on the conversion from cognitive normal (CN) to AD.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) study began in 2004 and it collected imaging, generic, clinical, and cognitive data from subjects under CN controls and subjects with MCI or AD in order to delineate the complex associations among various characteristics of the clinical spectrum of AD. The ADNI-1 recruited approximately 800 subjects according to its initial aim and has been extended by three follow-up studies, namely, ADNI-GO, ADNI-2, and ADNI-3. More information on ADNI can be obtained in the official website (www.adni-info.org). Functional assessment questionnaire (FAQ), an assessment of abilities to function independently in

[1]Department of Statistics, Chinese University of Hong Kong, Hong Kong, China
[2]Department of Statistics, Sun Yat-sen University, Guangzhou, China
[3]Shenzhen Research Institute, Chinese University of Hong Kong, Hong Kong, China
[4]MD Anderson Cancer Center, University of Texas, Houston, USA

Corresponding author:
Xinyuan Song, Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China.
Email: xysong@sta.cuhk.edu.hk

daily life, is widely used to monitor the decline of cognitive ability over time. Patients with higher FAQ scores have lower cognitive ability. The FAQ scores of each subject were obtained at baseline and then every 6 months across 9 years multiple study phases. Figure 1 plots the individual trajectories of FAQ scores for 20 randomly selected samples, which were initially diagnosed as CN, MCI, and AD, respectively. The cognitive decline patterns are apparently distinct over the groups, suggesting at least three (and probably more) distinct neurodegenerative states existent underlying the observations of FAQ score. Thus, for this longitudinal study, several central questions are naturally raised as follows: (1) How many hidden pathophysiological states exist in the progression of AD? (2) Which factors should contribute to the neurodegenerative pathology from one state (e.g., MCI) to another (e.g., AD)? (3) Whether the identified risk factors are equally good predictors of cognitive decline at each state? Given these questions, there is a particular need for the development of statistical models that delineate cognitive decline in terms of the pathophysiological states of AD.

Hidden Markov models (HMMs) are well suited to the characterization of longitudinal data in terms of a set of hidden states.[2–4] HMMs consist of two components: a transition model to describe the dynamic transition of hidden states and a conditional regression model to examine state-specific covariate effects on responses. Owing to their ability to simultaneously reveal the longitudinal association structure and dynamic heterogeneity of the observed process, HMMs and their variants have attracted significant attention from the medical, behavioral, social, environmental, and psychological sciences.[4–9] In particular, HMMs have previously been applied to investigate diseases progression to identify latent pathophysiological states. For instance, Albert et al.[10] used HMMs to analyze multiple sclerosis disease across relapse and remission states (see also Altman and Petkau;[11] Altman[12]). Ip et al.[13] identified 10 disable states on the basis of a 10-year follow-up study of late-life disability in elder adults, and examined the patterns and risk factors for transition among disable states. Song et al.[14] revealed the dynamic change of treatment effectiveness in preventing cocaine use across three cocaine addiction states.

Despite the rapid development and wide applications of HMMs, existing literature has mainly focused on parametric HMMs, in which the forms of covariate effects on responses and on transition probabilities are pre-specified. One problem of parametric models is that they may be too restrictive to reflect correctly the reality

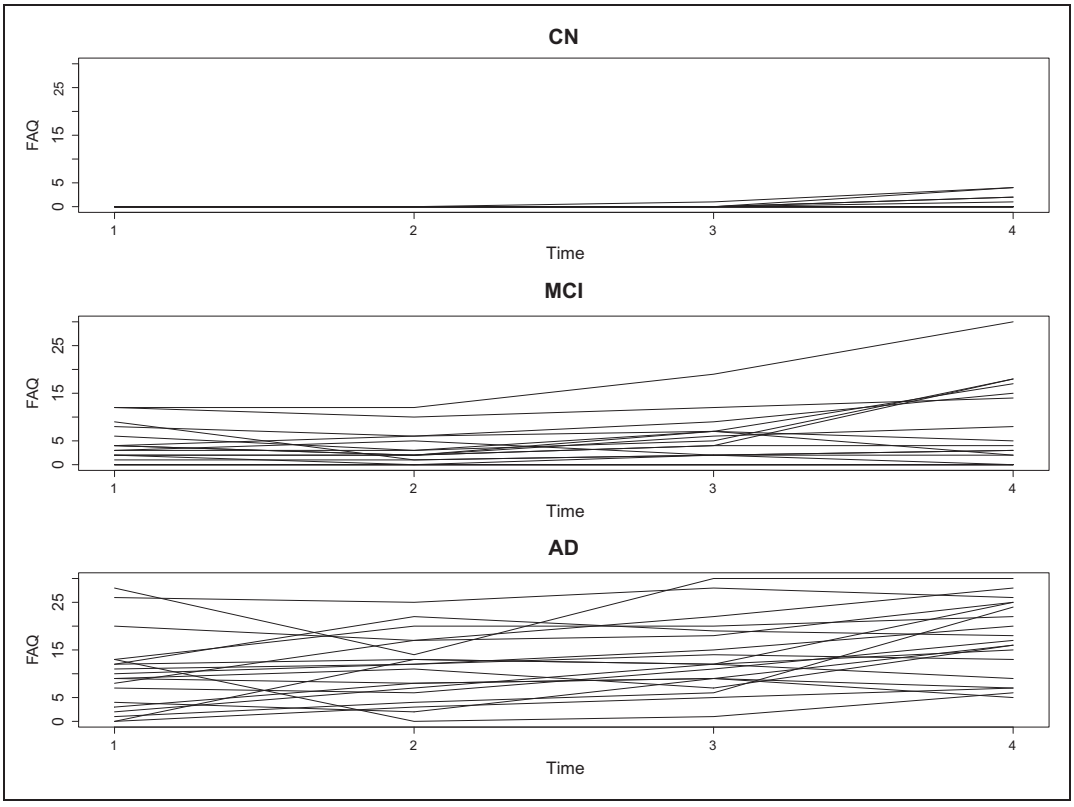

Figure 1. ADNI-1 data analysis results: individual trajectories of functional assessment questionnaire scores for 20 randomly selective samples whose baseline states are cognitive normal, mild cognitive impairment, and Alzheimer's disease, respectively.

because the complex relationships among variables are seldom known a priori, and a pre-specified parametric form tends to overlook the subtle pattern of a function. A more comprehensive analysis can be performed by incorporating nonparametric functions into HMMs so that the functional effects of interest can be discovered. One of a few exceptions in this direction is the development of Yau et al.,[15] which considered a Bayesian nonparametric HMM, in which the sampling distribution of the observations arising in each state was assumed unknown and formulated through a mixture of Dirichlet process (DP) model. Although such an approach enables the nonparametric modeling of the distribution of an observed process, it cannot reveal the functional effects of potential predictors on the outcome of interest.

In this study, we propose a Bayesian HMM to analyze the ADNI-1 data set. Similar to conventional HMMs, the proposed model consists of two major components. The first component is a state-dependent semiparametric regression to investigate linear and nonlinear covariate effects on the clinical outcome of cognitive decline (e.g., FAQ score). The second component is a mixed continuation-ratio logit transition model to examine covariate effects on the probabilities of transitioning among neurodegenerative states. We introduce a random effect in both models in order to account for interindividual differences and allow the random effects to be dependent by assigning a joint distribution for them. Such joint random effects enable the model to accommodate the situation where some omitted factors influence both the observed process and the hidden transition process.[16,17] We develop a full Bayesian approach, along with Bayesian P-splines procedure and Markov chain Monte Carlo (MCMC) methods, for statistical inference. Available works in Bayesian HMMs include, but are not limited to, Scott et al.[6] for investigating the health state of schizophrenia and comparing the effectiveness of two antipsychotic medications for schizophrenia, Yau et al.[15] for relaxing the distribution assumption of the observation process through a DP model, Scott[18] for introducing recursive computing and MCMC sampling into HMMs, and Johnson and Willsky[19] for analyzing sequential data in an unsupervised setting with the use of a hierarchical DP hidden semi-Markov model. Nevertheless, these studies were developed either in the parametric setting or for the purpose of extending the distributional assumptions on observed and/or hidden processes. To our knowledge, this study is the first to investigate the neurodegenerative pathology of AD and examine the possible functional effects of its potential risk factors in each pathophysiological state of AD.

The rest of this article is organized as follows. Section 2 defines the model and discusses model identifiability issues. Section 3 introduces the Bayesian inference procedure. Section 4 illustrates the use of the proposed model in the analysis of the ADNI data set. Section 5 demonstrates the empirical performance of the proposed methodology through a simulation study. Section 6 concludes the paper. Technical details are provided in the online Appendix.

2 Model description

The proposed model consists of two major components, including a conditional semiparametric regression model and a continuation-ratio logit transition model, as detailed below.

2.1 Conditional semiparametric regression model

Let y_{it} with subject $i = 1, \ldots, n$ at $t = 1, \ldots, T$ be the observation process. The hidden state process, Z_{it}, which takes values in $\{1, \ldots, S\}$, is assumed to follow a first-order Markov chain. Given the hidden state $Z_{it} = s$, the conditional semiparametric regression model is defined as follows

$$[y_{it}|Z_{it} = s] = \mu_s + \boldsymbol{\gamma}_s^T \mathbf{c}_{it} + \sum_{j=1}^{q} f_{sj}(x_{it,j}) + w_{i1} + \delta_{it} \tag{1}$$

where μ_s is a state-specific intercept, $\boldsymbol{\gamma}_s = (\gamma_1, \ldots, \gamma_r)$ is a state-specific vector of fixed effect of discrete covariates, $f_{sj}(\cdot)$s are state-specific unknown smoothing functions, $\mathbf{c}_{it} = (c_{it,1}, \ldots, c_{it,r})^T$ and $\mathbf{x}_{it} = (x_{it,1}, \ldots, x_{it,q})^T$ are $r \times 1$ vector of discrete covariates and $q \times 1$ vector of continuous covariates, respectively, w_{i1} is a subject-specific random effect, δ_{it} is a random residual independent of y_{it}, and $[\delta_{it}|Z_{it} = s] \sim N[0, \psi_s]$.

The conditional model defined by equation (1) extends the parametric regression to allow the additive nonparametric functions of covariates, so that the functional effects of interest can be discovered. Such nonparametric modeling provides great flexibility in fitting nonlinear effects whose forms need not be specified a priori. When used as an exploratory tool, the proposed model is able to help users to visually examine and interpret the functional effects of potential predictors on the response of interest. Moreover, the subject-specific

random effect w_{i1} permits additional dependencies elicited from other sources and thus avoids a large number of hidden states caused by possible residual correlation among responses.

2.2 Continuation-ratio logit transition model

Let p_{itus} denote the transition probability from state $Z_{i,t-1} = u$ at occasion $t-1$ to state $Z_{it} = s$ at occasion t for individual i. Based on the assumption of the first-order Markov chain, we have

$$p_{itus} = P(Z_{it} = s|Z_{i1}, Z_{i2}, \ldots, Z_{i,t-1} = u) = P(Z_{it} = s|Z_{i,t-1} = u) \tag{2}$$

The initial distribution of Z_{i1} is assumed to be a multinomial with probabilities $(\tau_1, \ldots, \tau_S)^T$ such that $\tau_s \geq 0$ and $\sum_{s=1}^{S} \tau_s = 1$. The distribution of $\{Z_{it}\}_{t=1}^{T}$ is then fully determined by the transition probabilities and the distribution of the initial state.

In the study of disease progression, the hidden states can often be naturally ranked (e.g., CN, MCI, and AD can be ranked from the best to the worst cognitive condition). Thus, we assume that the hidden states $\{1, \ldots, S\}$ are ordered and $\vartheta_{itus} = P(Z_{it} = s|Z_{it} \geq s, Z_{i,t-1} = u)$. Then, the transition across the ordered states can be described by continuation-ratio logits[20] as follows: For $t = 2, \ldots, T$, $s = 1, \ldots, S-1$, and $u = 1, \ldots, S$

$$\log\left(\frac{P(Z_{it} = s|Z_{i,t-1} = u)}{P(Z_{it} > s|Z_{i,t-1} = u)}\right) = \log\left(\frac{p_{itus}}{p_{itu,s+1} + \cdots + p_{ituS}}\right) = \text{logit}(\vartheta_{itus}) \tag{3}$$

The parameterization in equation (3) is intended to facilitate the interpretation of transition to a state rather than to a better one. To examine the effects of potential predictors on the transition probabilities, we consider a continuation-ratio logit transition model as follows

$$\text{logit}(\vartheta_{itus}) = \zeta_{us} + \boldsymbol{\alpha}^T \mathbf{d}_{it} + w_{i2} \tag{4}$$

where ζ_{us} is a state-specific intercept, $\mathbf{d}_{it}$ is a $h \times 1$ vector of covariates, $\boldsymbol{\alpha}$ is a $h \times 1$ vector of regression coefficients that can be interpreted as conditional log odds ratios in a logistic regression, w_{i2} is a subject-specific random effect that is distinct from but correlated with w_{i1}, and $\mathbf{w}_i = (w_{i1}, w_{i2})^T$ is assumed to follow a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Phi})$. Similar to the proportional assumption in a cumulative logit model, $\boldsymbol{\alpha}$ in equation (4) is assumed to be independent of u and s in order to maintain the order of the hidden states and avoid a tedious transition model, in which every transition elicits a set of parameters for all possible states of origination and destination. This outcome, in turn, greatly reduces the complexity and enhances the interpretability of the transition model. Moreover, unlike the existing literature that usually treats w_{i1} and w_{i2} separately, we accommodate their possible correlation by assigning a joint distribution for $\mathbf{w}_i = (w_{i1}, w_{i2})^T$. Consequently, the possible correlation between the heterogeneities existent within the two stochastic processes can be appropriately addressed and examined through the covariance matrix $\boldsymbol{\Phi}$.

2.3 Model identification

The proposed model is not identifiable because of the following three model indeterminacies. The first is caused by the additive nonparametric functions involved in equation (1), in which each unknown function is not identifiable up to a constant. To address this problem, we need to impose constraints on the unknown functions to enforce their integrations in the ranges of predictors to zero[21,22] as follows

$$\int_{\mathcal{X}_j} f_{sj}(x)\mathrm{d}x = 0, \quad \text{for} \quad s = 1, \ldots, S, \quad j = 1, \ldots, q \tag{5}$$

where $\mathcal{X}_j$ is the domain of $\mathbf{x}_j$. Let γ_{sm} and $c_{it,m}$ denote the mth element of $\boldsymbol{\gamma}_s$ and $\mathbf{c}_{it}$ in equation (1), respectively. The second model determinacy is caused by the fact that if $c_{it,m}$ is a constant, then μ_s, γ_{sm}, and w_{i1} are not simultaneously identifiable. Given that the mean of $\mathbf{w}_i$ is assumed to be $\mathbf{0}$, we can absorb γ_{sm} into μ_s to address the problem. Then, model (1) becomes an identifiable random intercept model with a nested error structure.[23–25] The last model determinacy is the label switching problem elicited by the invariance of the likelihood function to a random permutation of the state labels, which results in a multi-modal posterior under a symmetric prior specification. We follow Frühwirth-Schnatter[26] to conduct a permutation sampler to address this issue.

3 Bayesian inference

3.1 Nonparametric modeling

The first critical issue in the Bayesian analysis of the proposed model is to estimate the nonparametric functions involved in equation (1). We consider the use of Bayesian P-splines.[27–29] The basic idea is to estimate the unknown smoothing functions through a sum of B-splines basis functions[30] given a large number of knots in the domains of predictors. Specifically, $f_{sj}(x_{it,j})$, the functional effect of the jth covariate at state s for subject i at time t, can be approximated as follows

$$f_{sj}(x_{it,j}) = \sum_{l=1}^{L_{sj}} \beta_{sj,l} B_l(x_{it,j}) = \boldsymbol{\beta}_{sj}^T \mathbf{B}(x_{it,j}) \tag{6}$$

where L is the number of splines determined by the number of knots, $\boldsymbol{\beta}_{sj} = (\beta_{sj,1}, \ldots, \beta_{sj,L_{sj}})^T$ is the vector of the unknown parameters, $B(\cdot)$s' are cubic B-splines basis functions, and $\mathbf{B}(x_{it,j}) = (B_1(x_{it,j}), \ldots, B_{L_{sj}}(x_{it,j}))^T$. Usually, L_{sj} taking a value from 10 to 30 provides sufficient flexibility in fitting most smooth functions.

One problem of applying equation (6) to approximate an unknown smooth function is the overfitting caused by the use of a large number of knots. Eilers and Marx[31] suggested the penalization of the coefficients of adjacent B-splines basis functions to prevent the overfitting. Such penalization can be implemented in the Bayesian framework by applying random walk priors to $\boldsymbol{\beta}_{sj}$.[22,28,29]

3.2 Prior distributions

We assign a truncated Gaussian priors for $\boldsymbol{\beta}_{sj}$ as follows

$$p(\boldsymbol{\beta}_{sj}|\nu_{sj}) = \left(\frac{1}{2\pi\nu_{sj}}\right)^{L_{sj}/2} \exp\left\{-\frac{1}{2\nu_{sj}} \boldsymbol{\beta}_{sj}^T \mathbf{K}_{sj} \boldsymbol{\beta}_{sj}\right\} \mathrm{I}(\mathbf{1}_{n_s}^T \mathbf{B}_{sj} \boldsymbol{\beta}_{sj} = 0) \tag{7}$$

where ν_{sj} is a smoothing parameter for controlling the amount of penalty, $\mathbf{K}_{sj}$ is a penalty matrix derived according to the random walk penalties proposed, L_{sj} is the rank of $\mathbf{K}_{sj}$, $\mathbf{1}_{n_s}$ is an $n_s \times 1$ vector with all elements equal to 1, n_s is the sample size at state s, $\mathbf{B}_{sj}$ is the sub-matrix of $\mathbf{B}_j = [B_l(x_{it,j})]_{nT \times L}$ without the rows where $Z_{it} \neq s$, and the truncation term incorporates the identifiability constraint (equation (5)) into the splines approximation (equation (6)).

For the smoothing parameters ν_{sj}, we assign a highly dispersed but proper inverse gamma prior as follows

$$p(\nu_{sj}^{-1}) \stackrel{D}{=} \mathrm{Gamma}[\nu_1, \nu_2] \tag{8}$$

where ν_1 and ν_2 are hyperparameters whose values are pre-specified. In this study, we set $\nu_1 = 1$ and $\nu_2 = 0.005$ to follow a common practice.[22,29]

For the parameters involved in the conditional model (1), conjugate-type priors are assigned as follows: for $s = 1, \ldots, S$

$$\begin{aligned} p(\mu_s) &\stackrel{D}{=} N[\mu_{s0}, \sigma_{\mu s0}^2], \quad p(\boldsymbol{\gamma}_s) \stackrel{D}{=} N[\boldsymbol{\gamma}_{s0}, \textstyle\sum_{s0}], \\ p(\Phi^{-1}) &\stackrel{D}{=} \mathrm{Wishart}[\mathbf{R}_0, \rho_0], \quad p(\psi_s^{-1}) \stackrel{D}{=} \mathrm{Gamma}[\tilde{\alpha}_{s0}, \tilde{\beta}_{s0}] \end{aligned} \tag{9}$$

where μ_{s0}, $\sigma_{\mu s0}^2$, $\boldsymbol{\gamma}_{s0}$, Σ_{s0}, $\tilde{\alpha}_{s0}$, $\tilde{\beta}_{s0}$, $\mathbf{R}_0$, and ρ_0 are hyperparameters with preassigned values.

For the parameters involved in the transition model (4), we consider Gaussian priors

$$p(\zeta_{us}) \stackrel{D}{=} N[\zeta_{us0}, \sigma_{\zeta 0}^2], \quad p(\boldsymbol{\alpha}) \stackrel{D}{=} N[\boldsymbol{\alpha}_0, \mathbf{H}_{\alpha 0}], \quad p(\tau_s) \stackrel{D}{=} N[\tau_{s0}, \sigma_{\tau 0}^2] \tag{10}$$

where $\zeta_{us0}, \sigma_{\zeta 0}^2, \boldsymbol{\alpha}_0, \mathbf{H}_{\alpha 0}, \tau_{s0}$, and $\sigma_{\tau 0}^2$ are hyperparameters with preassigned values.

3.3 Posterior computation

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})^T$, $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iT})^T$, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$, $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)$, and $\boldsymbol{\theta}$ be the vector of the unknown parameters. The Bayesian estimate of $\boldsymbol{\theta}$ is obtained by drawing samples from $p(\boldsymbol{\theta}|\mathbf{Y})$, which is intractable because of the existence of latent states and random effects. We instead work on $p(\boldsymbol{\theta}, \mathbf{Z}, \mathbf{W}|\mathbf{Y})$ and use the Gibbs sampler to implement the posterior simulation. Owing to the nonlinearity of the continuation-logit transition model and the existence of the nonparametric functions in the conditional regression, some full conditional distributions, especially those related to the transition model, have complex forms. MCMC methods, such as the forward filtering and backward sampling algorithm[2] and the Metropolis–Hastings (MH) algorithm,[32,33] are employed to sample from them. The details are provided in the online Appendix.

With the use of posterior samples, the hidden states can be estimated as follows

$$\hat{Z}_{it} = \arg\max_{s\in\{1,\ldots,S\}} P(Z_{it} = s|\mathbf{y}_i, \boldsymbol{\theta}) \approx \arg\max_{s\in\{1,\ldots,S\}} \frac{1}{M}\sum_{m=1}^{M} \mathrm{I}(Z_{it}^{(m)} = s) \tag{11}$$

where $Z_{it}^{(m)}$ denotes the latent allocation of y_{it} at the mth iteration, and $\frac{1}{M}\sum_{m=1}^{M} \mathrm{I}(Z_{it}^{(m)} = s)$ is the posterior mean of the latent allocations of y_{it} drawn from the MCMC iterations.

3.4 Determination of the number of hidden states

In the applications of the proposed model to the ADNI data set, the states of the Markov chain can often naturally be interpreted as proxies for the neurodegenerative states, although a one-to-one correspondence between nominal HMM states and the clinical cognitive stages diagnosed by doctors is unnecessary. In this regard, a relevant question is how to determine the number of hidden states in the analysis of ADNI data. Celeux et al.[34] explored several versions of deviance information criteria (DIC)[35] for model comparison in the presence of incomplete data. They pointed out that DIC_4 is the best performing one among all the evaluated versions of DIC. In this study, we use DIC_4 for model comparison as follows

$$\mathrm{DIC} = \overline{D(\boldsymbol{\theta})} + p_D \tag{12}$$

where $\overline{D(\boldsymbol{\theta})} = E_{\boldsymbol{\theta},\mathbf{W},\mathbf{Z}}[-2\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}|\boldsymbol{\theta})|\mathbf{Y}]$ is the posterior mean deviance to measure the fit of the model, p_D is the effective number of parameters to reflect model complexity, and $p_D = E_{\boldsymbol{\theta},\mathbf{W},\mathbf{Z}}[-2\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}|\boldsymbol{\theta})|\mathbf{Y}] + 2E_{\mathbf{W},\mathbf{Z}}[\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z})|E_{\boldsymbol{\theta}}[\boldsymbol{\theta}|\mathbf{Y}, \mathbf{W}, \mathbf{Z}])|\mathbf{Y}]$. The expectations involved in equation (12) can be approximated using the posterior samples collected through MCMC methods.[34,35] The model with the smallest value of DIC is selected.

4 ADNI data analysis

We focused on 633 subjects who were obtained from the ADNI-I and followed up at baseline, 6 months, 12 months, and 24 months. For each subject, we included his/her clinical and genetic variables at the four time points. The clinical characteristics include gender (0 = male; 1 = female), age at baseline, and FAQ score. We considered apolipoprotein E (APOE)-ε4 as a covariate because it has been identified as a risk factor for early onset of AD.[36] APOE-ε4 is coded as 0, 1, and 2, denoting the number of APOE-ε4 alleles. Furthermore, the logarithm of the ratio of hippocampal volume over whole brain volume was included as a covariate because published reports[37–39] revealed that the atrophy of the hippocampal formation was a significant diagnostic marker of clinical dementia. The aims of this ADNI data analysis are (I) to identify the hidden states of the neurodegenerative pathology, (II) to reveal potential covariates that influence the between-states transition, and (III) to investigate the linear and/or functional covariate effects on cognitive decline across the hidden states of the AD progression.

We fitted the proposed model with the FAQ score as the response y_{it}, age and hippocampus that are directly associated with the FAQ score as covariates in $\mathbf{x}_{it}$, and typically individual characteristics, such as gender and APOE-ε4, as covariates in $\mathbf{d}_{it}$. Three continuous variables, FAQ score, hippocampus, and age, were standardized prior to analysis. We first determined the number of hidden states. We considered five competing models M_k, $k = 1, \ldots, 5$, where M_k represents a model with k states. A total of 24 equidistant knots were used to construct cubic P-splines, and the second-order random walk penalties were used for the Bayesian P-splines to estimate the unknown smooth functions. Given the lack of prior information, we assign the hyperparameters in equations (9) and (10) to reflect vague prior information as follows: $\mu_{s0} = \zeta_{us0} = \tau_{s0} = 0$, $\sigma_{\mu s0} = \sigma_{\zeta 0}^2 = \sigma_{\tau 0}^2 = 1$, $\tilde{\alpha}_{s0} = 9$,

$\tilde{\beta}_{s0} = 4$, $\rho_0 = 7$, $\mathbf{R}_0 = 4\mathbf{I}_2$, {$\boldsymbol{\alpha}_0$ is a vector with all elements being zero, and $\mathbf{H}_\alpha = \mathbf{I}_3$, where $\mathbf{I}_r$ is a r-dimensional identity matrix. We used the random permutation sampler to search for a suitable identifiability constraint to solve the label switching problem. The MCMC algorithm converged within 2000 iterations for all competing models. We collected a total of 10,000 observations after discarding 10,000 burn-in iterations to calculate DIC. The values of $\overline{D(\boldsymbol{\theta})}$, p_D, and DIC corresponding to M_1 to M_5 are reported in Table 1. The four-state model M_4 with the smallest DIC value was selected. Two things are noteworthy. First, directly comparing the DIC values between M_1 and M_s ($s > 1$) may be inappropriate because the values of $\overline{D(\boldsymbol{\theta})}$ under M_1 and M_s ($s > 1$) are not computed on exactly the same observed data; the latter involves additional covariates $\mathbf{d}_i$ in the transition model. Nevertheless, an apparent heterogeneity of the data can be revealed by Figure 1. Second, the value of $\overline{D(\boldsymbol{\theta})}$ is greater for M_5 than for M_4. After checking the results further, we found that one of the states in M_5 included only five individuals. Such an extremely small-sized state leads to unreliable results and implies nonexistence of five states in this application.

To examine the necessity of the random effects in the proposed model, we considered another competing model M_N: a four-state model without random effects. The DIC value suggests an evident advantage of the proposed mixed effect model in the presence of high dependency/heterogeneity in longitudinal observations. Thus, M_4 was selected for the subsequent analysis. The estimation results obtained under M_4 are reported in Table 2 (parametric part) and Figure 2 (nonparametric part).

We have the following observations. First, μ_1, μ_2, μ_3, and μ_4 are ranked in an ascending order, indicating that patients in state 1 got the lowest score of FAQ, whereas those in state 4 got the highest. That is, patients' ability to function independently in daily life steadily deteriorated from state 1 to state 4. According to the existing literature,[40] state 1 to state 4 can be explained as CN, early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD, respectively.

Second, the functional effect of hippocampus on FAQ exhibits a descending trend as hippocampus grows regardless of states. Specifically, in CN state, people with bigger hippocampus volume tend to have slightly

Table 1. Summary of DIC values in the analysis of ADNI data set.

Competing model	$\overline{D(\theta)}$	p_D	DIC
M_1	6322	27	6349
M_2	1650	16	1666
M_3	897	153	1050
M_4	860	178	1038
M_5	1219	236	1455
M_N	1025	72	1097

ADNI: Alzheimer's Disease Neuroimaging Initiative; DIC: deviance information criteria.

Table 2. ADNI-1 data analysis results: parameter estimation results.

State 1			State 2			State 3			State 4		
Par.	Est	SE	Par.	Est	SE	Par.	Est	SE	Par.	Est	SE
Parameters in conditional regression model											
μ_1	−0.547	0.008	μ_2	0.138	0.039	μ_3	1.125	0.055	μ_4	2.551	0.079
ψ_1	0.013	0.001	ψ_2	0.123	0.013	ψ_3	0.190	0.022	ψ_4	0.400	0.054
Parameters in probability transition model											
α_1	−0.597	0.149	α_2	−1.105	0.203	α_3	0.122	0.139			
τ_1	0.707	0.020	τ_2	0.140	0.020	τ_3	0.117	0.016	τ_4	0.036	0.009
ζ_{11}	2.631	0.149	ζ_{21}	−1.838	0.311	ζ_{31}	−3.344	0.461	ζ_{41}	−3.241	0.514
ζ_{12}	2.562	0.392	ζ_{22}	1.373	0.248	ζ_{32}	−2.138	0.386	ζ_{42}	−3.178	0.511
ζ_{13}	1.968	0.721	ζ_{23}	2.735	0.485	ζ_{33}	1.382	0.254	ζ_{43}	−1.978	0.458
Covariance matrix of random effects											
φ_{11}	0.022	0.002	φ_{22}	0.225	0.059	φ_{12}	−0.006	0.005			

ADNI: Alzheimer's Disease Neuroimaging Initiative; Est: Estimate; Par.: parameter; SE: standard error.

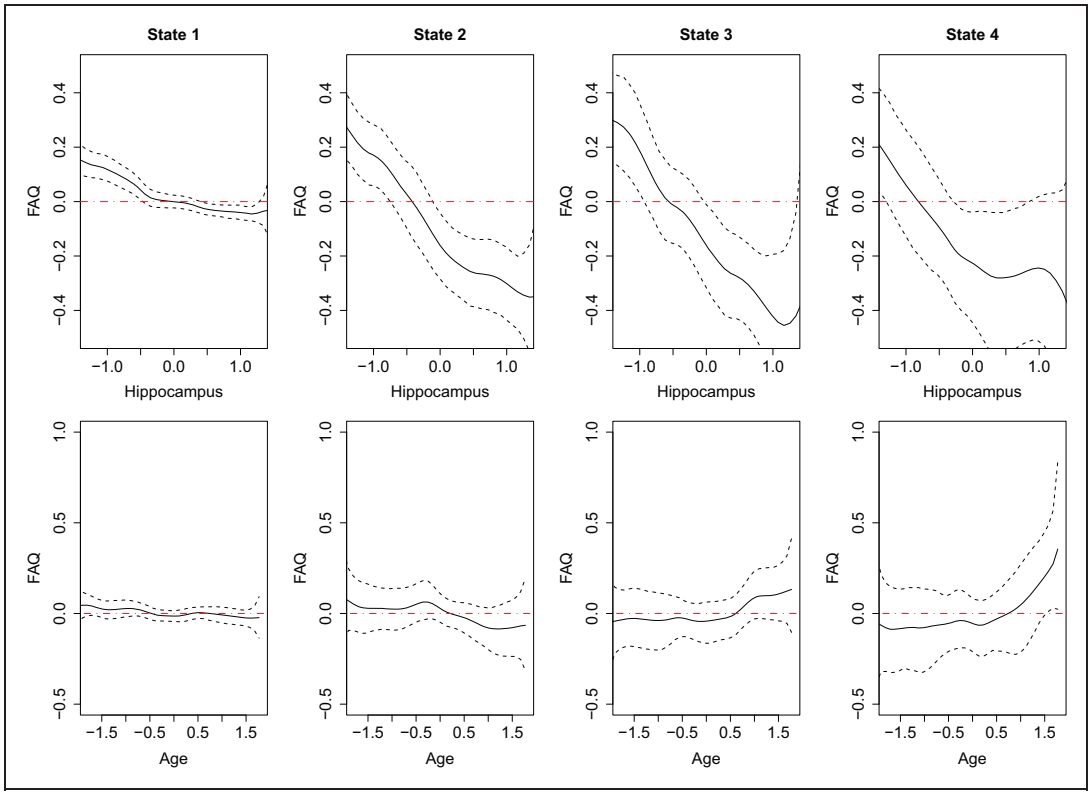

Figure 2. ADNI-1 data analysis results: estimates of the unknown smooth functions. The solid curves represent the pointwise mean curves, and the dashed curves represent the 2.5%- and 97.5%-pointwise quantiles. Line $y = 0$ has been shown on each picture by red dot-dash to illustrate the range of significant effect for each variable.

better memory. This finding is in line with the common sense that hippocampus plays an important role in the consolidation of information from short-term memory to long-term memory. In EMCI, LMCI, and AD states, the descending trend of the functional effect of hippocampus on FAQ becomes much more pronounced, implying that atrophy in hippocampal volume increasingly impairs patients' cognitive ability during the progression of EMCI–LMCI–AD. The published reports[37–39] also indicate that the volume loss of the hippocampus is greatly associated with clinical dementia.

Third, the effect of age on FAQ is nonsignificant in CN, EMCI, and LMCI states, implying that age affects cognitive function mainly in AD state. Elderly AD patients (say, over 85 years old) suffer from more increasing neurodegeneration than relatively young patients. Such age effect was also discovered by previous studies.[41,42]

Fourth, in the transition model, APOE-ε4 alleles have negative effects on the probability of transitioning from a state to a better one, indicating that APOE-ε4 alleles are important risk factors for the development of AD. This result agrees with those obtained by previous studies.[43] However, the effect of gender on the transition probability is inapparent.

Fifth, the variances of the two random effects are significant, reconfirming the necessity of the random effects proposed. However, the covariance between the two random effects is nonsignificant, showing that some omitted clinical or genetic indicators influenced outcomes of the observation process or probabilities of the transition process but did not affect the two processes simultaneously.

Moreover, we estimated the hidden states of all patients at four time points based on equation (11). Around 98.1% posterior transition patterns are from a state to a severer one, which is in line with the common knowledge of irreversibility of AD. Table 3 reports patients' estimated hidden states and their diagnosed status by doctors. For CN, LMCI, and AD states, a majority of the estimated states are consistent with those diagnosed by doctors. For EMCI state, however, 809 (65%) EMCI patients diagnosed by doctors were classified into CN state by our procedure. Such vague demarcation between CN and EMCI was also found and discussed in the literature (e.g., Petersen[44]).

Table 3. ADNI data analysis results: estimated hidden states versus diagnosed status.

Diagnosis \ Estimates	CN	EMCI	LMCI	AD	Total
CN	839	21	2	0	862
EMCI	809	266	147	21	1243
LMCI	7	21	46	16	90
AD	18	54	116	149	337
Total	1673	362	311	186	2532

ADNI: Alzheimer's Disease Neuroimaging Initiative; CN: cognitive normal; EMCI: early mild cognitive impairment; LMCI: late mild cognitive impairment; AD: Alzheimer's disease.

5 Simulation study

We conduct Monte Carlo simulations to assess the empirical performance of the proposed method in estimation of the nonparametric functions and model parameters.

5.1 Model setup

We consider a model with four hidden states ($S=4$), a continuous response y_{it}, three discrete covariates $\mathbf{d}_{it}=(d_{it,1},d_{it,2},d_{it,3})^T$ ($h=3$) and two continuous covariates $\mathbf{x}_{it}=(x_{it,1},x_{it,2})^T$ ($q=2$) to mimic the scenario of the ADNI study. For $i=1,\ldots,700$ and $t=1,\ldots,9$, $d_{it,1}$, $d_{it,2}$, and $d_{it,3}$ are all generated from the Bernoulli distribution with the probability of success 0.5, and $x_{it,1}$ and $x_{it,2}$ are generated from $U(-1,1)$ and $N(0,1)$, respectively. The conditional model is defined as

$$[y_{it}|Z_{it}=s]=\mu_s+f_{s1}(x_{it,1})+f_{s2}(x_{it,2})+w_{i1}+\delta_{it} \tag{13}$$

where $f_{11}(x_{it,1})=-1.185+\exp(x_{it,1})$, $f_{12}(x_{it,2})=\sin(1.5x_{it,2})+x_{it,2}$, $f_{21}(x_{it,1})=-0.04-\log((1+x_{it,1})/(1-x_{it,1}))$, $f_{22}(x_{it,2})=0.125+x_{it,2}^3$, $f_{31}(x_{it,1})=-0.8x_{it,1}$, $f_{32}(x_{it,2})=-0.175+\cos(2x_{it,2})+0.5x_{it,2}$, $f_{41}(x_{it,1})=-x_{it,1}^3$ and $f_{42}(x_{it,2})=0.05+1.5x_{it,2}$.

The transition model is defined as

$$\text{logit}(\vartheta_{itus})=\zeta_{us}+\alpha_1 d_{it,1}+\alpha_2 d_{it,2}+\alpha_3 d_{it,3}+w_{i2} \tag{14}$$

The true population values of the unknown parameters are set as $\boldsymbol{\mu}=(\mu_1,\mu_2,\mu_3,\mu_4)=(-5,-1,3,7)$, $\boldsymbol{\tau}=(\tau_1,\tau_2,\tau_3,\tau_4)=(0.27,0.27,0.23,0.23)$, $\zeta_{11}=\zeta_{21}=\zeta_{31}=\zeta_{41}=-1$, $\zeta_{12}=\zeta_{22}=\zeta_{32}=\zeta_{42}=-1/2$, $\zeta_{13}=\zeta_{23}=\zeta_{33}=\zeta_{43}=1/2$, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3)^T=(-0.5,0.5,1)$, $\boldsymbol{\psi}=(\psi_1,\psi_2,\psi_3,\psi_4)=(1,0.64,0.36,0.25)$, and $\boldsymbol{\Phi}$ is a correlation matrix with off diagonal elements -0.5. Based on the above setup, we simulate 100 data sets for analysis.

5.2 Simulation results

We used a total of 24 equidistant knots to construct the cubic B-splines of the covariates. Again, the second-order random walk penalties were used for the Bayesian P-splines to estimate the unknown smooth functions. The prior inputs in equations (9) and (10) were assigned as follows: $\mu_{s0}=\zeta_{us0}=\tau_{s0}=0$, $\sigma_{\mu s0}=\sigma_{\zeta 0}^2=\sigma_{\tau 0}^2=1$, $\tilde{\alpha}_{s0}=9$, $\tilde{\beta}_{s0}=4$, $\rho_0=7$, $\mathbf{R}_0=4\mathbf{I}_2$, and $\boldsymbol{\alpha}_0$ is a vector with all elements being zero, and $\mathbf{H}_\alpha=\mathbf{I}_3$. We conducted a few test runs to decide the number of burn-in iterations required for convergence and found that the MCMC algorithm converged within 2000 iterations. Therefore, we obtain Bayesian results using 10,000 observations after discarding 10,000 burn-in iterations. The performance of the Bayesian estimates is assessed through the bias (BIAS) and the root mean square errors (RMSE) between the Bayesian estimates and the true population values of the parameters.

Table 4 summarizes the result of parameter estimation based on the 100 data sets. The BIAS and RMSE for most of the parameters are close to zero, indicating a satisfactory performance of Bayesian estimation regarding the parametric part. Figure 3 depicts the averages of the pointwise posterior means of the nonparametric

functions, along with their 2.5%- and 97.5%- pointwise quantiles. The posterior means of the nonparametric functions are close to their true curves and all the ranges of the two pointwise quantiles are relatively small, indicating that the estimated functions can correctly recover the true functional relationships between the response and covariates. In this simulation, the average of the correct classification rates calculated through equation (11) based on the 100 data sets is 91%. Considering the complexity of proposed model, this result is satisfactory.

Table 4. Bayesian estimates of the parameters in the simulation study.

State 1			State 2			State 3			State 4		
Par.	BIAS	RMSE	Par.	BIAS	RMSE	Par.	BIAS	RMSE	Par.	BIAS	RMSE
Parameters in conditional regression model											
μ_1	0.014	0.058	μ_2	−0.011	0.069	μ_3	0.012	0.156	μ_4	0.018	0.126
ψ_1	0.010	0.038	ψ_2	0.042	0.076	ψ_3	0.017	0.036	ψ_4	0.010	0.031
Parameters in probability transition model											
α_1	−0.041	0.092	α_2	0.003	0.085	α_3	−0.022	0.086			
τ_1	0.001	0.021	τ_2	0.006	0.021	τ_3	−0.003	0.015	τ_4	−0.004	0.017
ζ_{11}	0.047	0.109	ζ_{21}	0.062	0.120	ζ_{31}	0.049	0.124	ζ_{41}	0.040	0.122
ζ_{12}	0.047	0.115	ζ_{22}	0.066	0.133	ζ_{32}	0.034	0.116	ζ_{42}	0.031	0.115
ζ_{13}	0.023	0.152	ζ_{23}	0.048	0.156	ζ_{33}	0.050	0.130	ζ_{43}	0.045	0.143
Covariance matrix of random effects											
φ_{11}	0.077	0.111	φ_{22}	0.027	0.061	φ_{12}	−0.018	0.098			

ADNI: Alzheimer's Disease Neuroimaging Initiative; Par.: parameter; RMSE: root mean square error.

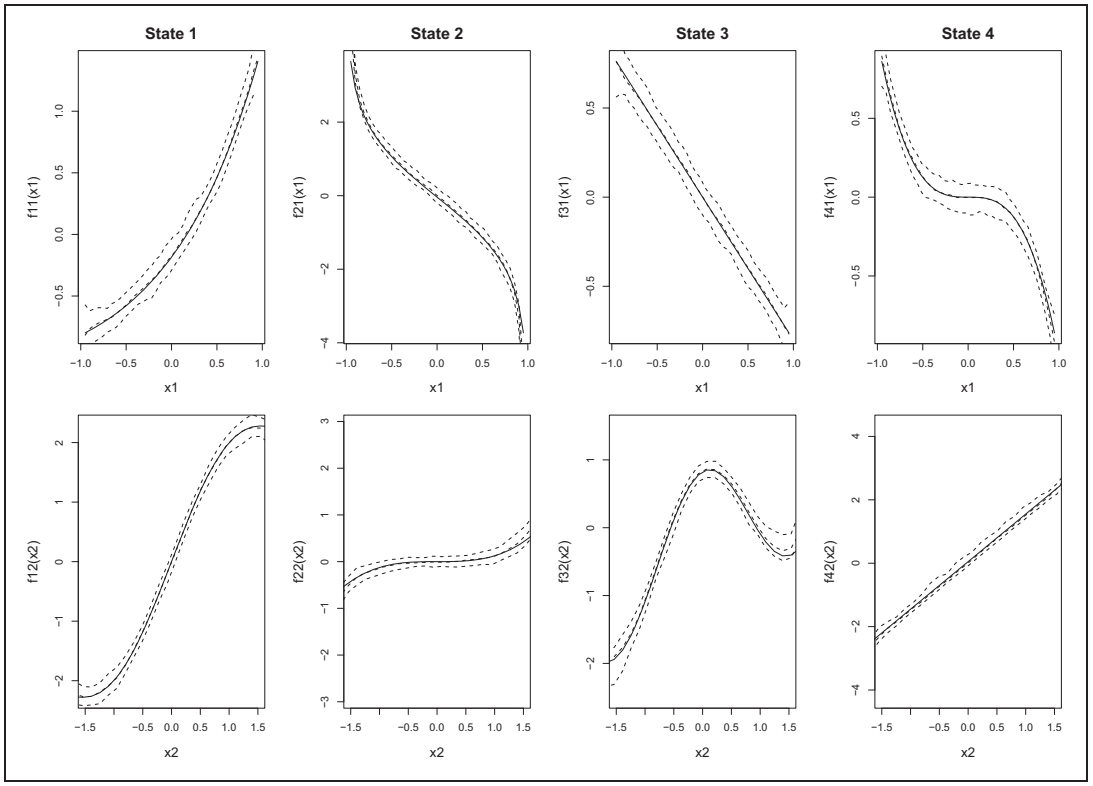

Figure 3. Estimates of the unknown smooth functions in the simulation study: the solid curves represent the true curves, and the dashed curves represent the estimated posterior means and the 2.5%- and 97.5%-pointwise quantiles based on 100 replications, respectively.

To reveal the sensitivity of the Bayesian estimates to the input of priors, we disturbed the prior inputs as follows: $\mu_{s0} = \zeta_{us0} = \tau_{s0} = 2,\ \sigma_{\mu s0} = \sigma^2_{\zeta 0} = \sigma^2_{\tau 0} = 2,\ \tilde{\alpha}_{s0} = 3,\ \tilde{\beta}_{s0} = 2,\ \rho_0 = 4,\ \mathbf{R}_0 = 2\mathbf{I}_2$, and $\boldsymbol{\alpha}_0$ are vectors with all elements being two, and $\mathbf{H}_\alpha = 2\mathbf{I}_3$. The obtained results are similar and not reported.

The computer code for conducting the preceding analyses is written in R and is freely available at www.sta.cuhk.edu.hk/xysong/codes/BHMMs.

6 Discussion

The proposed model was developed and successfully applied to the ADNI data analysis. Although HMMs and their variants have already been extensively used for longitudinal data analysis, a majority of applications restrict analysis in a parametric framework. Nonetheless, examples of using HMMs to characterize the neurodegenerative states of AD pathology are not prevalent, especially in a semiparametric context. In this study, we extended parametric HMMs to accommodate the functional effects of hippocampus and age on cognitive decline across four neurodegenerative states. Our model incorporates correlated random effects to account for individual and/or contextual differences in the progression of cognitive decline and in between-state transition. As we demonstrated in the ADNI study, accounting for such differences can dramatically improve model fit, as evidenced by an apparent improvement in DIC value between models with and without random effects. In addition, the correlation between the random effects enhances the model capability of accommodating the situation where some omitted covariates influence both the state-dependent observation process and the hidden-state transition process. Another appealing feature of this study is that it implements a full Bayesian analysis along with efficient MCMC methods. The sampling-based Bayesian approach is not only applicable to the current parameter-rich model but also possesses potential to address highly complex problems with which huge challenges are confronted by ML-based procedures.

This study can be extended in several directions: First, we considered the nonparametric modeling only in the conditional model. Generalizing the parametric transition model to a semiparametric/nonparametric one can further enhance model flexibility and analytic power. However, the statistical analysis of such comprehensive models can be challenging because the computational burden and sample size often limit the complexity of candidate models. Thus, the feasibility of this extension requires further investigation. Second, in the application to the ADNI data set, a highly comprehensive characterization of cognitive function is to group the FAQ, Alzheimer's Disease Assessment Scale, and Mini-Mental State Examination into an integrated latent construct through multivariate techniques such as factor analysis. Finally, this study did not consider missing data. Given that missingness is very common in longitudinal settings, accommodation of missing responses and/or missing covariates in the context of the proposed model framework is both of scientific interest and of practical value. These advances certainly require substantial efforts for further investigation.

Acknowledgement

The authors thank the editor, the associate editor, and the two anonymous referees for their valuable comments, which substantially improved the paper.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by GRF grants 14303017 and 14601115 from the Research Grant Council of the Hong Kong Special Administrative Region and NSFC 11471277 from the National Natural Science Foundation of China.

Supplemental material

Supplemental material for this article is available online.

References

1. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimers disease: recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7**: 270–279.
2. Cappé O, Moulines E and Rydén T. *Inference in hidden Markov models*. New York: Springer, 2005.
3. Maruotti A. Mixed hidden Markov models for longitudinal data: an overview. *Int Stat Rev* 2011; **79**: 427–454.
4. Bartolucci F, Farcomeni A and Pennoni F. *Latent Markov models for longitudinal data*. Florida: Chapman & Hall/CRC, Taylor and Francis Group, 2013.
5. Vermunt JK, Langeheine R and Böckenholt U. Latent Markov models with time-constant and time-varying covariates. *J Educ Behav Stat* 1999; **24**: 178–205.
6. Scott SL, James GM and Sugar CA. Hidden Markov models for longitudinal comparisons. *J Am Stat Assoc* 2005; **100**: 359–369.
7. Schmittmann VD, Dolan CV, van der Maas HL, et al. Discrete latent Markov models for normally distributed response data. *Multivariate Behav Res* 2005; **40**: 461–488.
8. Bartolucci F and Farcomeni A. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *J Am Stat Assoc* 2009; **104**: 816–831.
9. Chow SM, Grimm KJ, Filteau G, et al. Regime-switching bivariate dual change score model. *Multivariate Behav Res* 2013; **48**: 463–502.
10. Albert PS, McFarland HF, Smith ME, et al. Time series for modelling counts from a relapsingremitting disease: application to modelling disease activity in multiple sclerosis. *Stat Med* 1994; **13**: 453–466.
11. Altman RM and Petkau AJ. Application of hidden Markov models to multiple sclerosis lesion count data. *Stat Med* 2005; **24**: 2335–2344.
12. Altman RM. Mixed hidden Markov models. *J Am Stat Assoc* 2007; **102**: 201–210.
13. Ip E, Zhang Q, Rejeski J, et al. Partially ordered mixed hidden Markov model for the disablement process of older adults. *J Am Stat Assoc* 2013; **108**: 370–384.
14. Song XY, Xia YM and Zhu HT. Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics* 2017; **73**: 313–323.
15. Yau C, Papaspiliopoulos O, Roberts GO, et al. Bayesian nonparametric hidden Markov models with application to the analysis of copy-number-variation in mammalian genomes. *J R Stat Soc Ser B* 2011; **73**: 37–57.
16. Wulfsohn MS and Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–339.
17. Chi YY and Ibrahim JG. Regime-switching bivariate dual change score model. *Biometrics* 2006; **62**: 432–445.
18. Scott SL. Bayesian methods for hidden Markov models: recursive computing in the 21st century. *J Am Stat Assoc* 2002; **97**: 337–351.
19. Johnson MJ and Willsky AS. Bayesian nonparametric hidden semi-Markov Models. *J Mach Learn Res* 2013; **14**: 673–701.
20. Agresti A. *Categorical data analysis*. New York: John Wiley & Sons, 2002.
21. Panagiotelis A and Smith M. Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *J Econom* 2008; **143**: 291–316.
22. Song XY and Lu ZH. Semiparametric latent variable models with Bayesian P-splines. *J Comput Graph Stat* 2010; **19**: 590–608.
23. Congdon P. *Bayesian models for categorical data*. Chichester: Wiley, 2005.
24. Hedeker D and Gibbons RD. *Longitudinal data analysis*. New Jersey: Wiley, 2006.
25. Song XY and Lee SY. Model comparison of generalized linear mixed models. *Stat Med* 2006; **25**: 1685–1698.
26. Frühwirth-Schnatter S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J Am Stat Assoc* 2001; **96**: 194–209.
27. Berry SM, Carroll RJ and Ruppert D. Bayesian smoothing and regression splines for measurement error problems. *J Am Stat Assoc* 2002; **97**: 160–169.
28. Lang S and Brezger A. Bayesian P-splines. *J Comput Graph Stat* 2004; **13**: 183–212.
29. Fahrmeir L and Raach A. A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika* 2007; **72**: 327–346.
30. De Boor C. *A practical guide to splines*, Rev. ed. New York: Springer-Verlag, 2001.
31. Eilers PHC and Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci* 1996; **11**: 89–121.
32. Metropolis N, Rosenbluth AW, Rosenbluth MN, et al. Equations of state calculations by fast computing machine. *J Chem Phys* 1953; **21**: 1087–1091.
33. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970; **57**: 97–109.
34. Celeux G, Forbes F, Robert CP, et al. Deviance information criteria for missing data models. *Bayesian Anal* 2006; **1**: 651–674.
35. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Ser B* 2002; **64**: 583–639.

36. Okuizumi K, Onodera O, Tanaka H, et al. ApoE-ε4 and early onset Alzheimer's. *Nat Genet* 1994; **7**: 10–11.
37. Kesslak JP, Nalcioglu O and Cotman CW. Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. *Neurology* 1991; **41**: 51–54.
38. Jack CR, Petersen RC, O'Brien PC, et al. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology* 1992; **42**: 183–188.
39. Dickerson BC and Wolk D. Biomarker-based prediction of progression in MCI: comparison of AD-signature and hippocampal volume with spinal fluid amyloid-β and tau. *Front Aging Neurosci* 2013; **5**: 55.
40. Kantarci K, Gunter JL, Tosakulwong N, et al. Alzheimer's Disease Neuroimaging Initiative. Focal hemosiderin deposits and $\hat{I}^2$-amyloid load in the ADNI cohort. *Alzheimers Dement* 2013; **9**: S116–S123.
41. Gao S, Hendrie HC, Hall KS, et al. The relationships between age, sex, and the incidence of dementia and Alzheimer disease: a meta-analysis. *Arch Gen Psychiatry* 1998; **55**: 809–815.
42. Lindsay J, Laurin D, Verreault R, et al. Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. *Am J Epidemiol* 2002; **156**: 445–453.
43. Lee E, Zhu H, Kong D, et al. BFLCRM: a Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *Ann Appl Stat* 2015; **9**: 2153–2178.
44. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004; **256**: 183–194.