*Statistical Modelling* xxxx; **xx(x)**: 1–30

# A Bayesian transition model for missing longitudinal binary outcomes and an application to a smoking cessation study

**Li Li[1], Ji-Hyun Lee[2], Steven K Sutton[3], Vani N Simmons[4] and Thomas H Brandon[4]**
[1]Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, USA.
[2]Division of Quantitative Sciences, University of Florida Health Cancer Center; Department of Biostatistics, University of Florida, Gainesville, FL, USA.
[3]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA.
[4]Department of Health Outcomes and Behaviour, Moffitt Cancer Center, Tampa, FL, USA.

**Abstract:** Smoking cessation intervention studies often produce data on smoking status at discrete follow-up assessments, often with missing data in different amounts at each assessment. Smoking status in these studies is a dynamic process with individuals transitioning from smoking to abstinent, as well as abstinent to smoking, at different times during the intervention. Directly assessing transitions provides an opportunity to answer important questions like 'Does the proposed intervention help smokers remain abstinent or quit smoking more effectively than other interventions?' In this article, we model changes in smoking status and examine how interventions and other covariates affect the transitions. We propose a Bayesian approach for fitting the transition model to the observed data and impute missing outcomes based on a logistic model, which accounts for both missing at random (MAR) and missing not at random (MNAR) mechanisms. The proposed Bayesian approach treats missing data as additional unknown quantities and samples them from their posterior distributions. The performance of the proposed method is investigated through simulation studies and illustrated by data from a randomized controlled trial of smoking cessation interventions. Finally, posterior predictive checking and log pseudo marginal likelihood (LPML) are used to assess model assumptions and perform model comparisons, respectively.

**Key words:** Transition model, Bayesian method, generalized linear mixed model, missing values, smoking cessation

## 1 Introduction

Tobacco smoking continues to be a pressing healthcare problem because it significantly shortens the lifespan of smokers (Jha and Peto, 2014). According to Papadakis et al. (2010), there is still a need for low-cost, easily implemented and effective interventions. Smoking cessation studies are commonly used to investigate

Address for correspondence: Li Li, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA.

 10.1177/1471082X18821489

interventions which may help smokers to quit smoking, for example, see the meta-analyses on self-help interventions by Fiore et al. (2008) and Hartmann-Boyce (2014). In a recent randomized controlled trial by Brandon et al. (2016) (see Section 4 for more details), smoking status ('smoker' versus 'abstinent' using 7-day-point-prevalence) was assessed every 6 months for 2 years for participants randomly assigned to one of three treatment conditions. Longitudinal data analysis of binary smoking status at multiple times needs to adequately account for the correlations of the responses from each, and deal with missing responses that often arise in such studies.

There is a vast literature for analysing repeated binary data with missing values. We refer to Daniels and Hogan (2008) for a comprehensive review. Most modelling approaches focus on understanding covariates' effects on the binary response, where they typically equate the expected value of the response variable to a systematic component, which specifies covariates in a linear predictor function. Developed methodologies of this category include weighted generalized estimation equation (GEE), multiple imputation combined with GEE (Robins et al., 1995; Carpenter et al., 2006; Beunckens et al., 2008), multiple imputation coupled with generalized linear mixed effects model (GLMM; Diggle et al., 2002; Jansen et al., 2006), and marginalized transition models (Azzalini, 1994; Heagerty, 2002; Kurland and Heagerty, 2004). The marginalized transition approach combines a mean model for the response with a conditional mean model that describes a serial dependence and treats the serial dependence as a nuisance quantity.

Another class of transition models (Steele et al., 2004; Steele, 2011; Yeh et al., 2012) focus on quantifying the variations in the expected value of the response variable in response to covariates as well as past responses. We refer to this class of models as multilevel transition models and summarize it in the following:

$$\begin{aligned}\text{logit}(P(Y_{ti}=1|Y_{t-1,i}=0)) &= \mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01}+u_{0i},\\ \text{logit}(P(Y_{ti}=0|Y_{t-1,i}=1)) &= \mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10}+u_{1i},\end{aligned}$$

where $\text{logit}(\cdot)=\exp(\cdot)/(1+\exp(\cdot))$, $Y_{ti}$ is a binary variable indicating the state for individual $i$ at time $t$, $\mathbf{x}_{0ti}$ and $\mathbf{x}_{1ti}$ are covariates, $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$ are regression coefficients, and $u_{0i}$ and $u_{1i}$ are time-invariant random effects. Probabilities $P(Y_{ti}=1|Y_{t-1,i}=0)$ and $P(Y_{ti}=0|Y_{t-1,i}=1)$ are referred as transition probabilities. This class of models has potential to better understand the dynamics of smoking status using transition probabilities rather than only the means of responses. Figure 1 presents an example of smoking status transitions with all participants smoking at the start of the two-group comparison study. In both groups, 80% continue to smoke and 20% transition to abstinent at the 6-month assessment. The two groups diverge between the 6-month and 12-month assessments. In Group 1, 25% of smokers at 6 months transition to abstinent at 12 months and 20% of those abstinent at 6 months transition to smokers at 12 months. In Group 2, 40% of smokers at 6 months transition to abstinent at 12 months and 80% of those abstinent at 6 months

transition to smokers at 12 months. At 12 months, both groups will have 64% smokers and 36% abstinent. However, there are group differences in the composition of the smokers. In Group 1, 60% of smokers at 12 months were continuous smokers (i.e., also smoked at 6 months). In Group 2, only 48% of smokers at 12 were continuous smokers. In terms of those abstinent at 12 months, Group 2 has a larger proportion who quit smoking later in the study. Without missing responses, the joint likelihood of observing $\{\mathbf{y}_i, i = 1, \ldots, n\}$ under a multilevel transition model is equivalent to that under a GLMM where

$$\text{logit}(P(Y_{ti} = 1)) = (\mathbf{x}_{0ti}^T - y_{t-1,i}\mathbf{x}_{0ti}^T)\boldsymbol{\beta}^{01} - y_{t-1,i}\mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} + (1 - y_{t-1,i})u_{0i} - y_{t-1,i}u_{1i}.$$

However, when responses are partially missing, the likelihood marginalized over missing responses under a multilevel transition event model can no longer find an equivalence under the GLMM framework. Hence multiple imputations coupled with GLMM can no longer apply here. To our knowledge, extensions of the multilevel transition models that accommodate missing responses are limited.

In this article, we propose a Bayesian approach for extending the multilevel transition model to account for missing responses. Typically, Bayesian methods for missing longitudinal responses require a specification of the joint distribution of the responses and the missing indicators. Once the missing model is specified, missing responses can be augmented as additional unknown parameters and be sampled together with other parameters. We refer to Daniels and Hogan (2008) for a review of Bayesian missing imputation which has become a highly useful paradigm for handling missing values in many settings as it accounts for the uncertainty in imputation naturally. We further refer to Ibrahim and Molenberghs (2009) for a comprehensive review of missing data modelling in longitudinal studies. Notably, there are three important missing data mechanisms: missing completely at random (MCAR) when missingness is unrelated to the data, missing at random (MAR) when missingness depends on the observed data only, and missing not at random (MNAR) when missingness depends on the unobserved data, perhaps in addition to the observed data. A favourite type of model that accommodates the aforementioned three missing mechanisms is the selection model (Diggle and Kenward, 1994; Ibrahim et al., 2001), which specifies a conditional distribution, for example, logistic distribution, for the indicators of missing responses, given hypothetical complete data. In this work, we combine the multilevel transition models for the responses with logistic regressions for the missing indicators.

For the proposed joint modelling of transition events and missing responses, we developed a new Bayesian approach for fitting the models while employing Bayesian multiple imputations for missing values. In our proposed Bayesian approach, we utilize an approximation of a particular class of *t*-distributions to a class of logistic distributions (Kinney and Dunson, 2007), which devises conjugate conditional distributions for most of the parameters. The resulting MCMC mixes quickly after a short burn-in, which is appealing for imputing a large number of missing responses. The posterior samples allow us to compute the posterior mean probabilities of each smoking patterns, together with their credible intervals. Using

the posterior samples, we are also able to assess models' goodness of fit based on a posterior predictive checking method (Gelman et al., 2014; Xu et al., 2016) and perform model comparisons by log pseudo-marginal likelihood (LPML) (Geisser and Eddy, 1979; Linero and Daniels, 2015) and Watanabe-Akaike information criterion (WAIC) (Watanabe, 2010). We present simulation results to demonstrate the performance of the proposed algorithm and model selection criteria. We also demonstrate a real data analysis using data from the randomized controlled trial by Brandon et al. (2016).

The rest of the article is organized as follows: Section 2 presents the details of the model development, which include model base, prior construction, posterior inference and model assessment; Section 3 evaluates the proposed method using simulation studies; Section 4 illustrates the data analyses for a recent randomized controlled intervention trial for smoking cessation; we conclude the article with a summary in Section 5.

## 2 Model development

### 2.1 Model base

Two states of smoking behaviour are considered in our study: abstinent (A) and smoking (S). Let $Y_{ti}$ be a binary variable indicating the state for individual $i$ at time $t$; $Y_{ti} = 0$ for abstinent and $Y_{ti} = 1$ for smoking. The probabilities of transitioning from a state at time $t-1$ to either smoking or abstinent are:

$$p_{ti}^{01} = P(Y_{ti} = 1|Y_{t-1,i} = 0),\ p_{ti}^{00} = P(Y_{ti} = 0|Y_{t-1,i} = 0), \tag{2.1}$$

$$p_{ti}^{10} = P(Y_{ti} = 0|Y_{t-1,i} = 1),\ p_{ti}^{11} = P(Y_{ti} = 1|Y_{t-1,i} = 1), \tag{2.2}$$

where $p_{ti}^{00} = 1 - p_{ti}^{01}, p_{ti}^{11} = 1 - p_{ti}^{10}, t = 1, \ldots, m$ and $i = 1, \ldots, n$. Suppose all initial states $y_{0i}$ are known. If all initial states are 'smoking', using the definition of transition probabilities in equations (2.1) and (2.2), the probability of the pattern $S \to A \to A \to A \to A$ for the first four states is $p_{1i}^{10} p_{2i}^{00} p_{3i}^{00} p_{4i}^{00}$ and the probability of the pattern $S \to S \to S \to S \to S$ is $p_{1i}^{11} p_{2i}^{11} p_{3i}^{11} p_{4i}^{11}$. All other types of patterns are feasible following these notations. The class of multilevel transition models we consider here assumes logistic regression models for the transition probabilities,

$$\log\left(\frac{p_{ti}^{01}}{1 - p_{ti}^{01}}\right) = \mathbf{x}_{0ti}^T \boldsymbol{\beta}^{01} + u_{0i} \tag{2.3}$$

$$\log\left(\frac{p_{ti}^{10}}{1 - p_{ti}^{10}}\right) = \mathbf{x}_{1ti}^T \boldsymbol{\beta}^{10} + u_{1i} \tag{2.4}$$

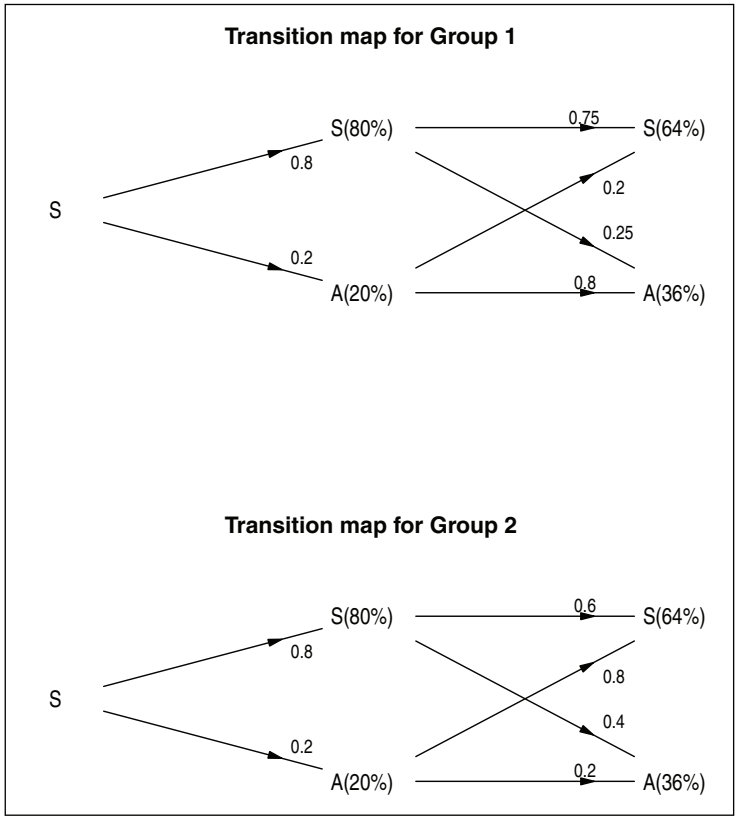


**Figure 1** Smoking status transition map for Group 1 and 2

**Notes:** S–smoking; A—abstinent; arrows are the transitions; the numeric values above the arrows are transition percentages; the numeric values after smoking statues are the overall percentages of smoking and abstinent.

where $\mathbf{x}_{0ti}$ and $\mathbf{x}_{1ti}$ (including intercepts) are covariates that may or may not differ from each other, $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$ are regression coefficients, $u_{0i}$ and $u_{1i}$ are time-invariant random effects, allowing unobserved heterogeneities among individuals in their probabilities of transitioning from state 0 to 1 and 1 to 0, respectively. Random vectors $\mathbf{u}_i = (u_{0i}, u_{1i}), i = 1, \ldots, n$ are assumed to follow a bivariate normal distribution independently, that is, $\mathbf{u}_i \overset{ind.}{\sim} N(\mathbf{0}, \Sigma)$, where $\Sigma$ is a two-dimensional covariance matrix.

As we mentioned in the introduction, the observed dataset has a large portion of missing values of the response variable, among which most are non-monotonic. Let $\delta_{ti}$ be a binary indicator which equals 1 if the state variable $y_{ti}$ is observed and equals 0 if $y_{ti}$ is missing. Define $\pi_{ti} = P(\delta_{ti} = 1)$. We consider a logistic regression model for

the missing indicators,

$$\text{logit}(\pi_{ti}) = \boldsymbol{\alpha}_1^T \mathbf{z}_{ti} + \boldsymbol{\alpha}_2^T \mathbf{y}_i^*, \tag{2.5}$$

where $\mathbf{z}_{ti} = (z_{t,i,1}, \ldots, z_{t,i,p})$ is a covariate vector and $\mathbf{y}_i^* = (y_{ti}, y_{t-1,i})$. The aforementioned model implies MCAR when all components in $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ except the intercept equal zero, MAR when $\boldsymbol{\alpha}_2$ equals zero and MNAR otherwise. We refer to Ibrahim et al. (2001) for other flexible missing data models and our method can be extended to consider those missing models. An important issue related to complicated missing data mechanism is model assumption validation. In Section 2.4, we propose a posterior predictive checking for validating the transition model and the missing data model. Let $\boldsymbol{\beta} = \{\boldsymbol{\beta}^{01}, \boldsymbol{\beta}^{10}\}$ and $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$. Denote $\mathbf{y}_{obs,i}$ and $\mathbf{y}_{mis,i}$ as the observed and missing responses for individual $i$, respectively. Denote $\boldsymbol{\delta}_i = (\delta_{1i}, \ldots, \delta_{mi})$ as missing indicators for individual $i$ at $m$ time points, $\mathbf{x}_i = (\mathbf{x}_{1i,} \ldots, \mathbf{x}_{mi})$ and $\mathbf{z}_i = (\mathbf{z}_{1i}, \ldots, \mathbf{z}_{mi})$ as the corresponding set of covariates. We write the observed dataset as $\mathcal{D} = \{(\mathbf{y}_{obs,i}, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\delta}_i); i = 1, \ldots, n\}$.

## 2.2 Prior specifications

Fully Bayesian methods require specifying priors for all the model parameters, which include $\boldsymbol{\beta} = (\boldsymbol{\beta}^{01}, \boldsymbol{\beta}^{10})$, $\boldsymbol{\alpha}$ and $\Sigma$. We assume independent multivariate Normal priors for $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$, and $\boldsymbol{\alpha}$, that is, $\boldsymbol{\beta}^{01} \sim N(\mathbf{0}, c_{\boldsymbol{\beta}}^2\mathbf{I})$, $\boldsymbol{\beta}^{10} \sim N(\mathbf{0}, c_{\boldsymbol{\beta}}^2\mathbf{I})$ and $\boldsymbol{\alpha} \sim N(\mathbf{0}, c_{\boldsymbol{\alpha}}^2\mathbf{I})$), where $c_{\boldsymbol{\beta}}^2$ and $c_{\boldsymbol{\alpha}}^2$ are scalars. Larger values of $c_{\boldsymbol{\beta}}^2$ and $c_{\boldsymbol{\alpha}}^2$ lead to less informative priors on the corresponding parameters. Our simulations in Section 3 show that setting $c_{\boldsymbol{\beta}}^2 = c_{\boldsymbol{\alpha}}^2 = 5^2$ allows accurate estimations of the parameters for all scenarios.

For the covariance matrix $\Sigma$, we adopt the hierarchical half-t prior in Huang and Wand (2013), which is a multivariate extension of the half-t prior in Gelman (2006). An important advantage of the hierarchical prior over the commonly used inverse Wishart prior is that it allows weakly informative priors on the standard deviation terms and imposes less restrictions on the relationships between standard deviations and correlations. A hierarchical half-t prior for two-dimensional $\Sigma$ is written as follows:

$$\begin{aligned}
&\Sigma \sim IW(a+1, 2a\Lambda),\\
&\Lambda = \text{diag}(\lambda_1, \lambda_2),\\
&\lambda_j \overset{ind}{\sim} \Gamma(\frac{1}{2}, \frac{1}{\varsigma_j^2}),
\end{aligned}$$

where $a$ is the degree of freedom, $1/\varsigma_j^2$ is the rate parameter of the Gamma distribution $\Gamma(\cdot,\cdot)$, and $IW(\cdot,\cdot)$ represents inverse Wishart distribution. Denote the standard deviation terms as $\sigma_j = \sqrt{\Sigma_{jj}}$, $j = 1, 2$ and the correlation term as $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$.

According to Huang and Wand (2013), the joint distribution of $\sigma_1, \sigma_2$, and $\rho$ is

$$p(\sigma_1, \sigma_2, \rho) \propto (1-\rho^2)^{-\frac{a}{2}-2}\sigma_1^{-a-2}\sigma_2^{-a-2}\left[-\frac{a}{(1-\rho)^2\sigma_1^2}+\frac{1}{\varsigma_1^2}\right]^{-(a+2)/2} \times \left[-\frac{a}{(1-\rho)^2\sigma_2^2}+\frac{1}{\varsigma_2^2}\right]^{-(a+2)/2}, \sigma_1, \sigma_2 > 0, -1 < \rho < 1.$$

The marginal distribution of $\sigma_j$ is Half-$t(a, \varsigma_j)$, which has density

$$p(\sigma_j) \propto \{1 + (\sigma_j/\varsigma_j)^2/a\}^{-(a+1)/2},$$

and

$$\mathrm{E}(\sigma_{jj}) = 2\varsigma_j\sqrt{\frac{a}{\pi}}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)(a-1)}, \mathrm{Var}(\sigma_{jj}) = \varsigma_j^2\frac{a}{a-2}\frac{4a}{\pi(a-1)^2}\left(\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)}\right)^2.$$

Therefore, larger values of $\varsigma_j^2$ lead to less informative priors on the corresponding standard deviation term. Huang and Wand (2013) recommend setting $\varsigma_j^2 = 10^5$, which leads to a weakly informative prior on the standard deviation term. The marginal distribution of the correlation parameter $\rho$ has density

$$p(\rho) \propto (1-\rho^2)^{\frac{a}{2}-1}.$$

When $a = 2$, the marginal distributions for the correlations are uniform distributions over $[-1, 1]$. Since the conditional distribution of $\rho$ given $\sigma_1$ and $\sigma_2$ is symmetric around 0, $E(\rho|\sigma_1, \sigma_2) = 0$. Therefore, $\Sigma_{12} = \rho\sigma_1\sigma_2$ also has prior mean at 0. We set $a = 2$ throughout simulations and data analyses.

In summary, the joint prior of the parameters in $\Theta = \{\boldsymbol{\beta}, \Sigma, \boldsymbol{\alpha}\}$ is

$$N(\boldsymbol{\beta}^{01}; 0, c_{\boldsymbol{\beta}}^2\mathbf{I})N(\boldsymbol{\beta}^{10}; 0, c_{\boldsymbol{\beta}}^2\mathbf{I})N(\boldsymbol{\alpha}; 0, c_{\boldsymbol{\alpha}}^2\mathbf{I}) \times IW(\Sigma; a+1, 2a\Lambda)\prod_{j=1}^{2}\Gamma(\lambda_j; \frac{1}{2}, \frac{1}{\zeta_j^2}). \quad (2.6)$$

## 2.3 Bayesian multiple imputations and posterior inference

Denote $\mathcal{M}$ as the set of indices $ti$ for all the missing responses. Let $\mathbf{u} = \{\mathbf{u}_i, i = 1, \ldots, n\}$ be the set of random effects. The joint likelihood $L(\mathbf{u}, y_{ti}, (t,i) \in \mathcal{M}, \Theta|\mathcal{D})$ is

$$\prod_{i=1}^{n}\prod_{t=1}^{m}\left[\left(p_{ti}^{01}\right)^{y_{ti}(1-y_{t-1,i})}\left(p_{ti}^{00}\right)^{(1-y_{ti})(1-y_{t-1,i})}\left(p_{ti}^{11}\right)^{y_{ti}y_{t-1,i}}\left(p_{ti}^{10}\right)^{(1-y_{ti})y_{t-1,i}}\right] \times \pi_{ti}^{\delta_{ti}}(1-\pi_{ti})^{1-\delta_{ti}}N(\mathbf{u}_i;\mathbf{0},\Sigma), \tag{2.7}$$

where $\pi_{ti} = P(\delta_{ti} = 1)$. To further exploit conjugacy in the exponential families for MCMC, we introduce auxiliary variables $\omega_{ti}, t = 1, \ldots, m; i = 1, \ldots, n$, which follow logistic distributions with location parameters as $[\mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01} + u_{0i}]I(y_{t-1,i} = 0) + [\mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} + u_{1i}]I(y_{t-1,i} = 1)$, where $I(\cdot)$ is an indicator function taking values 0 and 1. The distribution of $\omega_{ti}$ can be approximated by $N(\mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01} + u_{0i}, \tilde{\sigma}^2/\phi_{ti})I(y_{t-1,i} = 0) + N(\mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} + u_{1i}, \tilde{\sigma}^2/\phi_{ti})I(y_{t-1,i} = 1)$, $\tilde{\sigma}^2 = \pi^2(c-2)/3c$, $\phi_{ti} \sim \Gamma(c/2, c/2)$, $\pi = 3.1416$, and $c = 7.3$, for which the approximation error is negligible and can be corrected by importance sampling (Kinney and Dunson, 2007). Probabilities $p_{ti}^{01}$ and $p_{ti}^{10}$ are then approximated by $\Phi(\mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01} + u_{0i})$ and $\Phi(\mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} + u_{1i})$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Utilizing the auxiliary variables $\boldsymbol{\phi} = \{\phi_{ti}, t = 1, \ldots, m; i = 1, \ldots, n\}$ and $\boldsymbol{\omega} = \{\omega_{ti}, t = 1, \ldots, m; i = 1, \ldots, n\}$, the augmented likelihood $L(\boldsymbol{\omega}, \mathbf{u}, \boldsymbol{\phi}, y_{ti}, (t,i) \in \mathcal{M}, \Theta|\mathcal{D})$ is

$$\prod_{i=1}^{n}\prod_{t=1}^{m}\left\{N\left(\omega_{ti};\mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01} + u_{0i}, \frac{\tilde{\sigma}^2}{\phi_{ti}}\right)(1-y_{t-1,i})\left[I(\omega_{ti} > 0)y_{ti} + I(\omega_{ti} \le 0)(1-y_{ti})\right] + N\left(\omega_{ti};\mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} + u_{1i}, \frac{\tilde{\sigma}^2}{\phi_{ti}}\right)y_{t-1,i}\left[I(\omega_{ti} \le 0)y_{ti} + I(\omega_{ti} > 0)(1-y_{ti})\right]\right\} \times \Gamma\left(\phi_{ti};\frac{c}{2},\frac{c}{2}\right)\pi_{ti}^{\delta_{ti}}(1-\pi_{ti})^{1-\delta_{ti}}\prod_{i=1}^{n}N(\mathbf{u}_i;\mathbf{0},\Sigma). \tag{2.8}$$

We draw posterior samples of each variable in the set $\{\boldsymbol{\omega}, \mathbf{u}, \boldsymbol{\phi}, y_{ti}, (t,i) \in \mathcal{M}, \Theta, \Lambda\}$ from its corresponding conditional posterior distribution, that is, conditional distribution given observed data and the rest of the variables in the set. The following steps are repeated to obtain a desired number of samples.

**Sample missing $y_{li}$s:** For $(l,i) \in \mathcal{M}$, We sample $y_{li}$ from a Bernoulli distribution with $P(y_{li} = 1)$ and $P(y_{li} = 0)$ proportional to

$$\prod_{t=1}^{m}\left\{\left[\left(p_{ti}^{01}\right)^{y_{ti}}\left(p_{ti}^{00}\right)^{1-y_{ti}}\right]^{(1-y_{t-1,i})}\left[\left(p_{ti}^{10}\right)^{1-y_{ti}}\left(p_{ti}^{00}\right)^{y_{ti}}\right]^{y_{t-1,i}}\pi_{ti}^{\delta_{ti}}(1-\pi_{ti})^{1-\delta_{ti}}\right\} \tag{2.9}$$

with $y_{li}$ evaluated at 1 and 0 respectively.

**Sample $\omega_{ti}$s:** We sample $\omega_{ti}$ from its conditional posterior distribution that is proportional to

$$\begin{cases} N\left(\omega_{ti}; \mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01} + u_{0i}, \frac{\tilde{\sigma}^2}{\phi_{ti}}\right)[I(\omega_{ti} > 0)y_{ti} + I(\omega_{ti} < 0)(1 - y_{ti})]\,; \text{ if } y_{t-1,i} = 0, \\ N\left(\omega_{ti}; \mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} + u_{1i}, \frac{\tilde{\sigma}^2}{\phi_{ti}}\right)[I(\omega_{ti} \leq 0)y_{ti} + I(\omega_{ti} > 0)(1 - y_{ti})]\,; \text{ if } y_{t-1,i} = 1. \end{cases}$$

**Sample $\mathbf{u}_i$s:** Define $\zeta_{0i} = \sum_{t=1}^m \frac{\phi_{ti}}{\tilde{\sigma}^2}(1 - y_{t-1,i})(\omega_{ti} - \mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01})$, $\zeta_{1i} = \sum_{t=1}^m \frac{\phi_{ti}}{\tilde{\sigma}^2}y_{t-1,i}(\omega_{ti} - \mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10})$, $\kappa_{0i} = \sum_{t=1}^m \frac{\phi_{ti}}{\tilde{\sigma}^2}(1 - y_{t-1,i})$, and $\kappa_{1i} = \sum_{t=1}^m \frac{\phi_{ti}}{\tilde{\sigma}^2}y_{t-1,i}$. Sample $\mathbf{u}_i$ from a bivariate normal distribution $N(\boldsymbol{\mu}_i, \Omega_i)$ where $\boldsymbol{\mu}_i = \Omega_i(\zeta_{0i}, \zeta_{1i})^T$ and

$$\Omega_i = \left[\text{diag}(\kappa_{0i}, \kappa_{1i}) + \Sigma^{-1}\right]^{-1}.$$

**Sample $\phi_{ti}$s:** Sample $\phi_{ti}$ from $\Gamma(\frac{c}{2} + 0.5, \frac{A_{ti}}{2\tilde{\sigma}^2} + \frac{c}{2})$ where $A_{ti} = (\omega_{ti} - \mathbf{x}_{0ti}^T\boldsymbol{\beta}^{01} - u_{0i})^2 I(y_{t-1,i} = 0) + (\omega_{ti} - \mathbf{x}_{1ti}^T\boldsymbol{\beta}^{10} - u_{1i})^2 I(y_{t-1,i} = 1)$ and $c = 7.3$.
**Sample $\Sigma$:** Sample $\Sigma$ from

$$IW(n + a + 1, \sum_{i=1}^n \mathbf{u}_i\mathbf{u}_i^T + 2a\Lambda),$$

where $a$ is the degree of freedom in the hierarchical half-t prior.
**Sample $\Lambda$:** Sample diagonal value $\lambda_j$ from $\Gamma\left((a + 2)/2, \frac{1}{\varsigma_j^2} + a\Sigma_{jj}^{-1}\right)$, where $\Sigma_{jj}^{-1}$ is the $j$th diagonal value of $\Sigma^{-1}$.
**Sample $\boldsymbol{\beta}$:** Sample $\boldsymbol{\beta}^{01}$ from $N(\Psi_0 B_0, \Psi_0)$ where

$$\Psi_0^{-1} = \sum_{t=1}^m \sum_{i \in S_{0t}} \frac{\phi_{ti}}{\tilde{\sigma}^2}\mathbf{x}_{0ti}\mathbf{x}_{0ti}^T + \frac{1}{c_{\boldsymbol{\beta}}^2}\mathbf{I},\ B_0 = \sum_{t=1}^m \sum_{i \in S_{0t}} \frac{\phi_{ti}}{\tilde{\sigma}^2}(\omega_{ti} - u_{0i})\mathbf{x}_{0ti},$$

and $S_{0t}$ is the set of indices for individuals whose states are 0 at time $t - 1$. Similarly, define $S_{1t}$ is the set of indices for individuals whose states are 1 at time $t - 1$. We sample $\boldsymbol{\beta}^{10}$ from $N(\Psi_1 B_1, \Psi_1)$ where

$$\Psi_1^{-1} = \sum_{t=1}^m \sum_{i \in S_{1t}} \frac{\phi_{ti}}{\tilde{\sigma}^2}\mathbf{x}_{1ti}\mathbf{x}_{1ti}^T + \frac{1}{c_{\boldsymbol{\beta}}^2}\mathbf{I},\ B_1 = \sum_{t=1}^m \sum_{i \in S_{1t}} \frac{\phi_{ti}}{\tilde{\sigma}^2}(\omega_{ti} - u_{1i})\mathbf{x}_{1ti}.$$

**Sample $\boldsymbol{\alpha}$:** We propose a Metropolis–Hasting step for updating $\boldsymbol{\alpha}$. The conditional distribution of $\boldsymbol{\alpha}$ given $\mathbf{y}$, $\mathbf{u}$, $\Theta$, and $\mathcal{D}$, denoted as $p(\boldsymbol{\alpha}|\cdot)$, is proportional to

$$\prod_{i=1}^n \prod_{t=1}^m \pi_{ti}^{\delta_{ti}}(1 - \pi_{ti})^{1-\delta_{ti}} N(\boldsymbol{\alpha}; 0, c_{\boldsymbol{\alpha}}^2\mathbf{I}).$$

Let $\boldsymbol{\alpha}^*$ be the latest accepted value for $\boldsymbol{\alpha}$. Sample $\boldsymbol{\alpha}'$ from proposal distribution $N(\boldsymbol{\alpha}^*, \mathbf{V})$ and accept $\boldsymbol{\alpha}'$ with probability $\min\{p(\boldsymbol{\alpha}'|.)/p(\boldsymbol{\alpha}^*|.), 1\}$. The automatic tuning of $\mathbf{V}$ in Haario et al. (2005) is used in this article to get acceptance rate in the 20%–50% range. Specifically, let the sequence $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots$ be the states of the Markov chain for $\boldsymbol{\alpha}$. When deciding the $k$-th state $\boldsymbol{\alpha}$, we sample $\boldsymbol{\alpha}' \sim N(\boldsymbol{\alpha}^{(k-1)}, \mathbf{V}^{(k)})$ with

$$\mathbf{V}^{(k)} = \begin{cases} \mathbf{V}^{(0)}, & k < k_0, \\ s\mathrm{Var}\left\{\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(k-1)}\right\} + s_0\mathbf{I}_q, & k > k_0, \end{cases}$$

where $q$ is the dimension of $\alpha$, $s$ is recommended to be $2.4^2/q$, $k_0$ is the number of burn-in iterations, $s_0$ is a small constant, $\mathbf{V}^{(0)}$ is the initial variance of the proposal distribution and $\mathbf{I}_q$ is an identity matrix.

We point out that a recently developed Polya–Gamma data augmentation approach (Polson et al., 2013) can also yield a Gibbs sampler for the Bayesian logistic-regression type model. It is our future interest to compare the two augmentation approaches in extensions of the transition event model.

## 2.4 Posterior predictive checking and model comparison

**Posterior predictive checking:** To assess the goodness of fit of our models, we adopt the posterior predictive checking approach in Gelman et al. (2014, Chapter 6) and Xu et al. (2016). The idea is that the replicated data generated under the model should look similar to observed data if the model fits. In our scenario, observed data include $\mathbf{y}_{obs}$ and $\boldsymbol{\delta}$. Define $\mathbf{y}_{obs}^{rep}$ and $\boldsymbol{\delta}^{rep}$ as the replicated data that could have been observed, or we would see if the data generating process for $\mathbf{y}_{obs}$ and $\boldsymbol{\delta}$ were replicated with the same model (including model covariates and examination times) and the same value of model parameters. As an illustration, we define a test quantity $G(\mathbf{y}_{t,obs}) = \sum_{i\in\mathcal{O}_t} y_{ti}/n$ as the percentage of individuals smoking at time $t$ among the all responses, where $\mathcal{O}_t$ as the set of indices $(ti)$ for all the responses that are not missing at time $t$. Posterior predictive distribution of the test quantity can be approximated by MCMC samples $G(\mathbf{y}_{t,obs}^{rep,l}) = \sum_{i\in\mathcal{O}_t^{rep,l}} y_{ti}^{rep,l}/n, l = 1, \ldots, L$. We summarize the posterior predictive distribution by the sample mean, 2.5% quantile and 97.5% quantile of the samples. Since we are interested in smoking status transitions and missing model estimation, we also compare observed transition percentages to simulated transition percentages and observed missing percentages to simulated missing percentages.

**Model comparison:** We compare models using LPML (Geisser and Eddy, 1979) and WAIC (Watanabe, 2010). LPML is a measure of a model's predictive ability and WAIC is a fully Bayesian approach for estimating point wise out-of-sample prediction accuracy from a fitted Bayesian model. More recently, Linero and Daniels (2015) used LPML to compare models for longitudinal studies with missing data. Denote $\mathbf{y}_{obs}$ and

$\mathbf{y}_{mis}$ as the observed and missing responses respectively, $\mathbf{y}_{obs,-i}$ and $\boldsymbol{\delta}_{-i}$ as the observed responses and missing indicators for those with individual $i$ removed.

By definition,

$$\text{LPML} = \sum_{i=1}^{n} \log(p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{y}_{obs,-i}, \boldsymbol{\delta}_{-i})),$$

where $p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{y}_{obs,-i}, \boldsymbol{\delta}_{-i})$ is the predictive probability of $\mathbf{y}_{obs,i}$ and $\boldsymbol{\delta}_i$ given remaining observed data $\mathbf{y}_{obs,-i}$ and $\boldsymbol{\delta}_{-i}$. Using the representation in Gelfand and Dey (1994),

$$p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{y}_{obs,-i}, \boldsymbol{\delta}_{-i})^{-1} = \int \frac{1}{p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i, \Theta)} p(\Theta, \mathbf{u} | \mathbf{y}_{obs}, \boldsymbol{\delta}) d\mathbf{u} d\Theta.$$

Therefore, LPML can be approximated by

$$\frac{1}{L} \sum_{l=1}^{L} \frac{1}{p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i^{(l)}, \Theta^{(l)})} \tag{2.10}$$

where $\{\mathbf{u}_i^{(l)}, \Theta^{(l)}, l = 1, \ldots, L\}$ are MCMC iterates.

By definition,

$$\text{WAIC} = \sum_{i=1}^{n} \log \int p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i, \Theta) p(\Theta, \mathbf{u} | \mathbf{y}_{obs}, \boldsymbol{\delta}) d\mathbf{u} d\Theta - p_{\text{waic}}$$

where $p_{\text{waic}}$ is the effective number of parameters. Two approaches of formulating $p_{\text{waic}}$ have been proposed in literature (Gelman et al., 2014). In this work, we use the variance of individual terms in the log predictive density summed over the $n$ data points, that is, $p_{\text{waic}} = \sum_{i=1}^{n} \text{Var}_{\text{post}}(\log(p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i, \Theta)))$. The WAIC value can be approximated by

$$\sum_{i=1}^{n} \log \left( \frac{1}{L} \sum_{l=1}^{L} p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i^{(l)}, \Theta^{(l)}) \right) - \sum_{i=1}^{n} V_{l=1}^{L} \left( \log \left( p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i^{(l)}, \Theta^{(l)} \right) \right) \tag{2.11}$$

where $V_{l=1}^{L} a_l = \frac{1}{L-1} \sum_{l=1}^{L} (a_l - \bar{a})^2$.

Note that for the missing model we propose,

$$p(\mathbf{y}_{obs,i}, \boldsymbol{\delta}_i | \mathbf{u}_i, \Theta) = \sum_{\mathbf{y}_{mis,i}} p(\mathbf{y}_{obs,i}, \mathbf{y}_{mis,i}, \boldsymbol{\delta}_i | \mathbf{u}_i, \Theta)$$

and

$$p(\mathbf{y}_i, \boldsymbol{\delta}_i|\mathbf{u}_i, \Theta)=\prod_{t=1}^{m}\left\{\left[\left(p_{ti}^{01}\right)^{y_{ti}}\left(p_{ti}^{00}\right)^{1-y_{ti}}\right]^{(1-y_{t-1,i})}\left[\left(p_{ti}^{10}\right)^{1-y_{ti}}\left(p_{ti}^{00}\right)^{y_{ti}}\right]^{y_{t-1,i}}\pi_{ti}^{\delta_{ti}}(1-\pi_{ti})^{1-\delta_{ti}}\right\}.$$

When comparing two models, a greater LPML or WAIC indicates a better fit of the data.

## 3 Simulations

In this section, we aim to examine parameter estimation under a correctly specified or partially mis-specified joint model of multilevel transitions and missing imputations. Specifically, we consider (a) both the multilevel transition model and the missing model are correctly specified; (b) the missing model is mis-specified and (c) the random effects distribution is mis-specified. Also, we perform sensitivity analyses on the hyper-parameters and conduct evaluations on the performance of model selection criteria LPML and WAIC for the proposed models.

**Simulation I:** In this simulation, we examine parameter estimation under a correctly specified multilevel transition model and missing model. For each simulated dataset, consider $n$ individuals and $m$ follow-up times. Covariate vectors $\mathbf{x}_{sti}; s = 0, 1; t = 1, \ldots, m; i = 1, \ldots, n$ were simulated independently, with their first elements being one, their second elements sampled from $N(0, 1)$ and their third elements sampled from Bernoulli (0.5). Initial states $\{y_{0i}, i = 1, \ldots, n\}$ were simulated from Bernoulli(0.5). We set true regression vectors $\boldsymbol{\beta}^{01} = (1.0, -0.5, 0.5)$ and $\boldsymbol{\beta}^{10} = (-1.0, 0.5, -0.5)$. Random effects $\mathbf{u}_i, i = 1, \ldots, n$ were sampled independently from a bivariate normal distribution $N_2(\mathbf{0}, \Sigma)$, where the diagonal elements of $\Sigma$ are 0.6 and 0.4, and the off-diagonal elements are $-0.3$. Conditional on the simulated $\mathbf{u}_i$ and $\mathbf{x}_{sti}$, we simulated responses from the transition event model:

$$\log\left(\frac{p_{ti}^{01}}{1-p_{ti}^{01}}\right) = \mathbf{x}_{0ti}^{T}\boldsymbol{\beta}^{01} + u_{0i}$$
$$\log\left(\frac{p_{ti}^{10}}{1-p_{ti}^{10}}\right) = \mathbf{x}_{1ti}^{T}\boldsymbol{\beta}^{10} + u_{1i}$$

where $P(Y_{ti} = 1|Y_{t-1,i} = 0) = p_{ti}^{01}$ and $P(Y_{ti} = 1|Y_{t-1,i} = 1) = 1 - p_{ti}^{10}$. Conditional on $\boldsymbol{\alpha}$ and the sampled responses, we simulated the missing indicators from model (2.5),

$$\text{logit}(\pi_{ti} = 1) = \alpha_1 + \alpha_2 z_{ti} + \alpha_3 y_{ti} + \alpha_4 y_{t-1,i},$$

where $z_{ti}$ was sampled from $N(0, 0.5^2)$ and $\boldsymbol{\alpha} = (2.0, 0.5, -1.0, -1.0)$. We consider two follow-up times for $m$—6 and 10, and two sample sizes for $n$—200 and 300. The simulated datasets have approximately 30% missing values. The starting values are zeros for $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$, and random effects, (1, 0, 0) for $\boldsymbol{\alpha}$ and 0.2, 0, 0.2 for $\Sigma_{11}$, $\Sigma_{12}$, and $\Sigma_{22}$, respectively. Hyper-parameters $c^2_{\boldsymbol{\beta}}$, $c^2_{\boldsymbol{\alpha}}$ and $\varsigma^2_j$ were fixed at $5^2$, $5^2$ and $10^5$, respectively. We obtained a chain of 6 000 iterates after a burn-in of 20 000 and thinning of every other 5 iterates. Each simulation scenario was repeated 500 times using the MCMC steps outlined in Section 2.3. For each parameter in $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$, $\boldsymbol{\alpha}$ and $\Sigma$, we present the numerical summaries of posterior mean, bias, mean squared error (MSE) and 95% coverage probability.

Under 'missing model correctly specified' in Table 1, we represent the numerical summaries of Simulation I for $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$, $\Sigma$ and $\boldsymbol{\alpha}$. We observe significant reductions in biases and MSEs when sample size or follow-up times increase. We also see the same patterns under a different set-up of $\boldsymbol{\alpha}$ (simulations not shown here). The results suggest that the posterior means under correct model specifications provide a consistent estimation of all parameters. Under the scenario $n = 300, m = 10$, all parameters except $\Sigma_{12}$ have estimated biases ranging from 0.01% to 5% of their true values. For $\Sigma_{12}$, the bias is 29% of its true value. In fact, when $n = 200$, the percentages of bias for $\Sigma_{12}$ are 58% and 42% for $m = 6$ and 10, respectively; when $n = 300$, they are 55% and 29% for $m = 6$ and 10, respectively. As we increased $m$ to 20 and kept $n$ at 200, the percentage of bias is 26% (based on simulations not shown in the table). Random effects distributions are generally hard to estimate. The correlation parameter of $u_{i1}$ and $u_{i2}$ is even more difficult to estimate as it relies on the correct estimations of $u_{i1}$ and $u_{i2}$ simultaneously for each individual. Without much information about the correlation, the posterior mean of $\Sigma_{12}$ is skewed towards its prior mean, which is zero under the hierarchical half-t prior with $a = 2$. Using a simulated dataset under the setting $n = 200, m = 6$, we observe that the empirical posterior distributions of $\Sigma_{11}$ and $\Sigma_{22}$ are skewed toward positive infinity, but the empirical posterior distribution of $\Sigma_{12}$ is skewed towards 0. Finally, the coverage probabilities of all parameters except $\Sigma_{12}$ are close to 95% and above 93%. Both the bias and skewness of the posterior distribution for $\Sigma_{12}$ have impacts on its 95% credible interval and results in their comparatively lower coverage probability. Greater sample sizes or follow-up times ($n$ and $m$) draw the coverage probabilities of all parameters closer to 95% slightly. In addition to the effect of sample size and follow-up times, prior distributions with less prior variances may lead to greater biases and changes in coverage probabilities, which we will discuss in the section of Simulation III.

**Simulation II:** In this simulation, we fitted the simulated datasets from Simulation I again but assumes the MAR assumption for the missing model:

$$\text{logit}(\pi_{ti} = 1) = \alpha_1 + \alpha_2 z_{ti}.$$

Comparing the estimated biases under 'missing model correctly specified' (assuming MNAR) and 'missing model mis-specified' (assuming MAR) in Table 1, mis-specifying the missing model leads to greater estimated biases for all parameters and

the differences are especially significant for the intercept terms of $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$, and $\Sigma_{11}$. Under the scenario of $n = 300$, $m = 10$, estimated biases for the intercept terms of $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$, and $\Sigma_{11}$ are 17%, 18%, and 40% of the true values respectively, which are much higher than the range of 0.01% to 5% from Simulation I. Therefore, we conclude that mis-specifying the missing model can lead to substantial biases in estimation.

**Simulation III:** In this simulation, we fitted the simulated datasets from Simulation I to three additional set-ups of hyper-parameters: (a) $\varsigma_j^2 = 10$, $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2 = 5^2$, (b) $\varsigma_j^2 = 10^2$, $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2 = 5^2$ and (c) $\varsigma_j^2 = 10^5$, $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2 = 3^2$. Smaller $\varsigma_j^2$, $c_{\boldsymbol{\alpha}}^2$ and $c_{\boldsymbol{\beta}}^2$ correspond to priors with smaller variances. The results are presented in Table 2. Comparing the results in Table 1 under 'missing model correctly specified' ($\varsigma_j^2 = 10^5$, $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2 = 5^2$), prior settings (a), (b) and (c) have smaller MSEs for all parameters but have greater biases and changes in coverage probabilities for some parameters. For example, under $\zeta_j^2 = 10$ and $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2 = 5^2$, $\beta^{01}$, $\beta^{10}$ and $\boldsymbol{\alpha}$ have greater biases in general, while $\Sigma_{11}$ and $\Sigma_{12}$ have slightly smaller biases. Coverage probabilities for $\beta^{01}$, $\beta^{10}$ and $\boldsymbol{\alpha}$ are less close to 95%, compared to Table 1. Similar comparisons are observed for $\varsigma_j^2 = 10$ and $10^2$. Changing $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2$ from $5^2$ to $3^2$ have small impacts in biases and coverage probabilities. We therefore recommend setting $\varsigma_j^2 = 10^5$, $c_{\boldsymbol{\alpha}}^2 = c_{\boldsymbol{\beta}}^2 = 5^2$ and performing sensitivity analysis on hyper-parameters.'

**Simulation IV:** In this simulation, true random effects were simulated from a normal mixture distribution $0.5\mathrm{N}(\boldsymbol{\mu}_1, \Sigma) + 0.5\mathrm{N}(\boldsymbol{\mu}_2, \Sigma)$ where $\boldsymbol{\mu}_1 = (-0.5, 0.5)$, $\boldsymbol{\mu}_2 = (0.5, -0.5)$, and $\Sigma$ has diagonal elements 0.6 and 0.4 and off-diagonal elements $-0.3$. The marginal distribution of each random effect term has zero mean and two modes. Results in Table 4 show small estimated biases for $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$ but large estimated biases for the intercept term and the coefficient of $y_{ti}$ in $\boldsymbol{\alpha}$. Under the scenario of $n = 300$ and $m = 10$, estimated biases range from 0.4% to 6% for $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$ and 1.4% to 18% for $\boldsymbol{\alpha}$. The small estimated biases for $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$ suggest a degree of robustness of the proposed model to random effect distribution mis-specification. It agrees with McCulloch and Neuhaus (2011) who demonstrate a large degree of robustness of estimated covariates effects in generalized linear mixed models. Nevertheless, the missing model parameters are sensitive to random effect distribution mis-specification. A Bayesian non-parametric modelling of the random effects distribution (Antonelli et al., 2016) may reduce bias in parameter estimation and provide a framework to assess the Normality assumption. It is our interest to explore this extension in the future.

**Simulation V:** In this simulation, we evaluate the performance of model selection criteria LPML and WAIC. We simulated data from two additional set-ups of $\boldsymbol{\alpha}$ and computed the proportions of fits choosing MNAR over MAR for the missing model based on LPML and WAIC. Null set considers $\boldsymbol{\alpha}$=(1, 0.5, 0.0, 0.0), for which the true missing model assumes MAR; Sets 1 and 2 assume $\boldsymbol{\alpha}$=(1, 0.5, $-0.5$, $-0.5$) and (2, 0.5, $-1$, $-1$) respectively. Both Sets 1 and 2 assume MNAR for the true missing

model. Greater LPML (WAIC) indicates a better model fit. Based on a cutoff of 0 (Table 3), the proportions of selecting the true models increase for both LPML and WAIC as sample size $n$ or follow-up times $m$ increases. Under the scenario $n = 300$ and $m = 10$, the proportions of choosing the true models are all close to 1. WAIC has lower proportions of selecting MNAR than LPML does under the Null set and Set 1, suggesting a more conservative measure favouring the simpler model (MAR). Both criteria appear to be reasonable model selection tools for our proposed model.

## 4 Data analysis

We apply the proposed Bayesian transition method to the data from a randomized controlled trial evaluating extended self-help intervention for smoking cessation (Brandon et al., 2016). In this trial, participants (smokers who wanted to quit smoking) were randomly assigned to one of the three interventions to help them quit smoking: traditional self-help (TSH), standard repeated mailing (SRM) and intensive repeated mailing (IRM). TSH received an existing self-help booklet for quitting smoking. SRM received eight different cessation booklets mailed over a 12-month period. IRM received monthly mailings of 10 booklets and additional material designed to enhance social support over 18 months. The authors hypothesized that IRM would produce the greatest smoking abstinence over time, followed by SRM and then TSH, which was included as a 'usual care' comparison condition. The primary outcomes were the self-reported 7-day-point-prevalence (i.e., no tobacco cigarettes smoked in previous seven days) smoking status collected at 6, 12, 18 and 24 months, and were coded as '1' for smoking and '0' for abstinent. All participants were reported to be smoking at the baseline, that is, $y_{0t} = 1$ for all individuals. Their findings indicated that the self-help interventions with increased intensity and duration resulted in significantly improved abstinence rates, using a logistic regression model at each time point and the GEE approach over time. To manage missing data, they applied multiple imputations under the MAR assumption. A post hoc approach was then used to address the influence of MNAR on smoking status. For our analysis, baseline characteristics expected to be related to smoking status at follow-up assessments were age, sex, marital status and income level. After excluding observations with missing values of the baseline features and another 308 participants who missed all of the follow-up measurements, we have $n = 1\,530$ participants as the analytical data. Figure 2 displays the smoking status transitions, and the numeric values (outside the parentheses) are the observed transition percentages.

In the following analyses below, age was standardized; sex='1' for male and '0' for female; marital status ='1' for married, and '0' otherwise; income level was standardized. Reference group refers to female participants at the sample mean age (48 years old), not married, with income US \$30 000–40 000 annually, and treated with IRM intervention. More detailed information about the study design and covariates can be found in Brandon et al. (2016). Hyper-parameters for the priors are initially fixed at $\varsigma_j^2 = 10^3$, $c_{\boldsymbol{\beta}} = 5^2$ and $c_{\boldsymbol{\alpha}} = 4^2$. Starting values for the

**Table 1** Simulated data: True value and Monte Carlo mean, bias, and mean square error (MSE) and 95% coverage probability of the transition event model parameters where $\boldsymbol{\alpha}$=(2.0, 0.5, −1.0, −1.0)

| | | | | Missing model correctly specified | | | | Missing model mis-specified | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| n | m | Parameter | True Value | Mean | Bias | MSE | 95% CP | Mean | Bias | MSE | 95% CP |
| 200 | 6 | $\beta_1^{01}$ | 1.0 | 0.999 | 0.001 | 0.120 | 0.936 | 0.867 | 0.133 | 0.147 | 0.912 |
| | | $\beta_2^{01}$ | −0.5 | −0.532 | 0.032 | 0.066 | 0.940 | −0.544 | 0.044 | 0.073 | 0.932 |
| | | $\beta_3^{01}$ | 0.5 | 0.532 | 0.032 | 0.216 | 0.942 | 0.534 | 0.034 | 0.234 | 0.940 |
| | | $\beta_1^{10}$ | −1.0 | −1.044 | 0.044 | 0.064 | 0.938 | −0.641 | 0.359 | 0.196 | 0.680 |
| | | $\beta_2^{10}$ | 0.5 | 0.529 | 0.029 | 0.034 | 0.934 | 0.567 | 0.067 | 0.042 | 0.924 |
| | | $\beta_3^{10}$ | −0.5 | −0.497 | 0.003 | 0.097 | 0.944 | −0.528 | 0.028 | 0.121 | 0.944 |
| | | $\Sigma_{11}$ | 0.6 | 0.752 | 0.152 | 0.207 | 0.942 | 1.051 | 0.451 | 0.499 | 0.864 |
| | | $\Sigma_{12}$ | −0.3 | −0.126 | 0.174 | 0.051 | 0.860 | −0.117 | 0.183 | 0.062 | 0.872 |
| | | $\Sigma_{22}$ | 0.4 | 0.503 | 0.103 | 0.098 | 0.960 | 0.496 | 0.096 | 0.145 | 0.964 |
| | | $\alpha_1$ | 2.0 | 2.038 | 0.038 | 0.097 | 0.944 | 0.689 | 1.311 | 1.723 | 0.000 |
| | | $\alpha_2$ | 0.5 | 0.500 | 0.000 | 0.017 | 0.952 | 0.452 | 0.048 | 0.017 | 0.932 |
| | | $\alpha_3$ | −1.0 | −1.014 | 0.014 | 0.139 | 0.934 | * | * | * | * |
| | | $\alpha_4$ | −1.0 | −1.017 | 0.017 | 0.028 | 0.948 | * | * | * | * |
| 200 | 10 | $\beta_1^{01}$ | 1.0 | 1.007 | 0.007 | 0.084 | 0.938 | 0.856 | 0.144 | 0.103 | 0.894 |
| | | $\beta_2^{01}$ | −0.5 | −0.528 | 0.028 | 0.038 | 0.944 | −0.531 | 0.031 | 0.040 | 0.944 |
| | | $\beta_3^{01}$ | 0.5 | 0.528 | 0.028 | 0.140 | 0.934 | 0.527 | 0.027 | 0.139 | 0.944 |
| | | $\beta_1^{10}$ | −1.0 | −1.042 | 0.042 | 0.042 | 0.934 | −0.627 | 0.373 | 0.181 | 0.526 |
| | | $\beta_2^{10}$ | 0.5 | 0.516 | 0.016 | 0.018 | 0.942 | 0.556 | 0.056 | 0.024 | 0.926 |
| | | $\beta_3^{10}$ | −0.5 | −0.492 | 0.008 | 0.063 | 0.942 | −0.526 | 0.026 | 0.077 | 0.950 |
| | | $\Sigma_{11}$ | 0.6 | 0.667 | 0.067 | 0.075 | 0.952 | 0.868 | 0.268 | 0.161 | 0.864 |
| | | $\Sigma_{12}$ | −0.3 | −0.175 | 0.125 | 0.028 | 0.892 | −0.17 | 0.13 | 0.035 | 0.904 |
| | | $\Sigma_{22}$ | 0.4 | 0.425 | 0.025 | 0.036 | 0.956 | 0.45 | 0.05 | 0.072 | 0.950 |
| | | $\alpha_1$ | 2.0 | 2.020 | 0.020 | 0.046 | 0.956 | 0.679 | 1.321 | 1.748 | 0.000 |
| | | $\alpha_2$ | 0.5 | 0.502 | 0.002 | 0.010 | 0.952 | 0.451 | 0.049 | 0.011 | 0.926 |
| | | $\alpha_3$ | −1.0 | −1.005 | 0.005 | 0.067 | 0.948 | * | * | * | * |
| | | $\alpha_4$ | −1.0 | −1.012 | 0.012 | 0.016 | 0.940 | * | * | * | * |
| 300 | 6 | $\beta_1^{01}$ | 1.0 | 0.993 | 0.007 | 0.078 | 0.942 | 0.853 | 0.147 | 0.102 | 0.904 |
| | | $\beta_2^{01}$ | −0.5 | −0.540 | 0.040 | 0.039 | 0.958 | −0.553 | 0.053 | 0.044 | 0.952 |
| | | $\beta_3^{01}$ | 0.5 | 0.546 | 0.046 | 0.144 | 0.952 | 0.561 | 0.061 | 0.157 | 0.948 |
| | | $\beta_1^{10}$ | −1.0 | −1.025 | 0.025 | 0.039 | 0.940 | −0.614 | 0.386 | 0.191 | 0.498 |
| | | $\beta_2^{10}$ | 0.5 | 0.507 | 0.007 | 0.019 | 0.942 | 0.535 | 0.035 | 0.024 | 0.938 |
| | | $\beta_3^{10}$ | −0.5 | −0.515 | 0.015 | 0.071 | 0.948 | −0.546 | 0.046 | 0.086 | 0.936 |
| | | $\Sigma_{11}$ | 0.6 | 0.668 | 0.068 | 0.100 | 0.954 | 0.961 | 0.361 | 0.269 | 0.856 |
| | | $\Sigma_{12}$ | −0.3 | −0.134 | 0.166 | 0.045 | 0.850 | −0.117 | 0.183 | 0.057 | 0.840 |
| | | $\Sigma_{22}$ | 0.4 | 0.457 | 0.057 | 0.061 | 0.976 | 0.403 | 0.003 | 0.080 | 0.948 |
| | | $\alpha_1$ | 2.0 | 2.017 | 0.017 | 0.051 | 0.954 | 0.688 | 1.312 | 1.725 | 0.000 |
| | | $\alpha_2$ | 0.5 | 0.505 | 0.005 | 0.011 | 0.960 | 0.459 | 0.041 | 0.011 | 0.932 |
| | | $\alpha_3$ | −1.0 | −1.012 | 0.012 | 0.073 | 0.938 | * | * | * | * |
| | | $\alpha_4$ | −1.0 | −0.999 | 0.001 | 0.018 | 0.948 | * | * | * | * |
| 300 | 10 | $\beta_1^{01}$ | 1.0 | 1.011 | 0.011 | 0.046 | 0.947 | 0.849 | 0.151 | 0.065 | 0.923 |
| | | $\beta_2^{01}$ | −0.5 | −0.521 | 0.021 | 0.024 | 0.937 | −0.526 | 0.026 | 0.024 | 0.943 |
| | | $\beta_3^{01}$ | 0.5 | 0.477 | 0.023 | 0.079 | 0.953 | 0.479 | 0.021 | 0.081 | 0.957 |
| | | $\beta_1^{10}$ | −1.0 | −1.021 | 0.021 | 0.024 | 0.950 | −0.603 | 0.397 | 0.182 | 0.283 |
| | | $\beta_2^{10}$ | 0.5 | 0.503 | 0.003 | 0.011 | 0.947 | 0.538 | 0.038 | 0.015 | 0.920 |
| | | $\beta_3^{10}$ | −0.5 | −0.511 | 0.011 | 0.038 | 0.933 | −0.546 | 0.046 | 0.050 | 0.947 |
| | | $\Sigma_{11}$ | 0.6 | 0.646 | 0.046 | 0.045 | 0.960 | 0.838 | 0.238 | 0.111 | 0.853 |
| | | $\Sigma_{12}$ | −0.3 | −0.214 | 0.086 | 0.018 | 0.897 | −0.209 | 0.091 | 0.025 | 0.890 |
| | | $\Sigma_{22}$ | 0.4 | 0.402 | 0.002 | 0.023 | 0.973 | 0.407 | 0.007 | 0.041 | 0.947 |
| | | $\alpha_1$ | 2.0 | 2.024 | 0.024 | 0.031 | 0.957 | 0.686 | 1.314 | 1.730 | 0.000 |
| | | $\alpha_2$ | 0.5 | 0.505 | 0.005 | 0.007 | 0.947 | 0.458 | 0.042 | 0.008 | 0.907 |
| | | $\alpha_3$ | −1.0 | −1.014 | 0.014 | 0.045 | 0.960 | * | * | * | * |
| | | $\alpha_4$ | −1.0 | −1.006 | 0.006 | 0.011 | 0.943 | * | * | * | * |

**Note:** * indicates that the corresponding parameter is not estimated.

**Table 2** Simulated data: True value and Monte Carlo mean, bias and mean square error (MSE) and 95% coverage probability of the transition event model parameters where $\boldsymbol{\alpha}$=(2.0, 0.5, −1.0, −1.0)

| | | | | $\varsigma_j^2 = 10, c_\alpha^2 = c_\beta^2 = 5^2$ | | | | $\varsigma_j^2 = 10^2, c_\alpha^2 = c_\beta^2 = 5^2$ | | | | $\varsigma_j^2 = 10^5, c_\alpha^2 = c_\beta^2 = 3^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| n | m | Parameter | True Value | Mean | Bias | MSE | 95% CP | Mean | Bias | MSE | 95% CP | Mean | Bias | MSE | 95% CP |
| 200 | 6 | $\beta_1^{01}$ | 1.0 | 1.015 | 0.015 | 0.120 | 0.952 | 1.006 | 0.006 | 0.121 | 0.942 | 0.987 | 0.013 | 0.114 | 0.942 |
| | | $\beta_2^{01}$ | −0.5 | −0.530 | 0.030 | 0.065 | 0.932 | −0.545 | 0.045 | 0.066 | 0.950 | −0.530 | 0.030 | 0.063 | 0.946 |
| | | $\beta_3^{01}$ | 0.5 | 0.528 | 0.028 | 0.189 | 0.962 | 0.526 | 0.026 | 0.206 | 0.952 | 0.540 | 0.040 | 0.176 | 0.960 |
| | | $\beta_1^{10}$ | −1.0 | −1.035 | 0.035 | 0.062 | 0.952 | −1.015 | 0.015 | 0.070 | 0.938 | −1.043 | 0.043 | 0.070 | 0.920 |
| | | $\beta_2^{10}$ | 0.5 | 0.527 | 0.027 | 0.031 | 0.952 | 0.534 | 0.034 | 0.032 | 0.938 | 0.512 | 0.012 | 0.030 | 0.944 |
| | | $\beta_3^{10}$ | −0.5 | −0.536 | 0.036 | 0.114 | 0.938 | −0.550 | 0.050 | 0.114 | 0.926 | −0.508 | 0.008 | 0.104 | 0.946 |
| | | $\Sigma_{11}$ | 0.6 | 0.714 | 0.114 | 0.165 | 0.952 | 0.755 | 0.155 | 0.187 | 0.952 | 0.704 | 0.104 | 0.175 | 0.946 |
| | | $\Sigma_{12}$ | −0.3 | −0.120 | 0.180 | 0.053 | 0.856 | −0.127 | 0.173 | 0.052 | 0.870 | −0.112 | 0.188 | 0.056 | 0.820 |
| | | $\Sigma_{22}$ | 0.4 | 0.489 | 0.089 | 0.094 | 0.964 | 0.526 | 0.126 | 0.117 | 0.962 | 0.509 | 0.109 | 0.106 | 0.968 |
| | | $\alpha_1$ | 2.0 | 2.054 | 0.054 | 0.080 | 0.954 | 2.039 | 0.039 | 0.095 | 0.930 | 2.021 | 0.021 | 0.080 | 0.948 |
| | | $\alpha_2$ | 0.5 | 0.502 | 0.002 | 0.017 | 0.960 | 0.511 | 0.011 | 0.021 | 0.942 | 0.517 | 0.017 | 0.018 | 0.950 |
| | | $\alpha_3$ | −1.0 | −1.036 | 0.036 | 0.108 | 0.958 | −1.012 | 0.012 | 0.131 | 0.938 | −0.999 | 0.001 | 0.109 | 0.946 |
| | | $\alpha_4$ | −1.0 | −1.017 | 0.017 | 0.024 | 0.950 | −1.014 | 0.014 | 0.024 | 0.952 | −1.010 | 0.010 | 0.026 | 0.962 |
| 200 | 10 | $\beta_1^{01}$ | 1.0 | 0.974 | 0.026 | 0.092 | 0.910 | 1.000 | 0.000 | 0.082 | 0.946 | 1.002 | 0.002 | 0.069 | 0.958 |
| | | $\beta_2^{01}$ | −0.5 | −0.540 | 0.040 | 0.041 | 0.910 | −0.534 | 0.034 | 0.040 | 0.934 | −0.517 | 0.017 | 0.034 | 0.946 |
| | | $\beta_3^{01}$ | 0.5 | 0.519 | 0.019 | 0.142 | 0.960 | 0.539 | 0.039 | 0.143 | 0.938 | 0.524 | 0.024 | 0.129 | 0.944 |
| | | $\beta_1^{10}$ | −1.0 | −1.024 | 0.024 | 0.033 | 0.980 | −1.034 | 0.034 | 0.040 | 0.938 | −1.038 | 0.038 | 0.036 | 0.950 |
| | | $\beta_2^{10}$ | 0.5 | 0.505 | 0.005 | 0.018 | 0.950 | 0.514 | 0.014 | 0.016 | 0.956 | 0.507 | 0.007 | 0.016 | 0.950 |
| | | $\beta_3^{10}$ | −0.5 | −0.520 | 0.020 | 0.061 | 0.950 | −0.503 | 0.003 | 0.061 | 0.942 | −0.487 | 0.013 | 0.061 | 0.946 |
| | | $\Sigma_{11}$ | 0.6 | 0.701 | 0.101 | 0.081 | 0.970 | 0.668 | 0.068 | 0.081 | 0.948 | 0.655 | 0.055 | 0.071 | 0.938 |
| | | $\Sigma_{12}$ | −0.3 | −0.174 | 0.126 | 0.030 | 0.880 | −0.171 | 0.129 | 0.030 | 0.872 | −0.181 | 0.119 | 0.030 | 0.868 |
| | | $\Sigma_{22}$ | 0.4 | 0.407 | 0.007 | 0.028 | 0.990 | 0.414 | 0.014 | 0.034 | 0.958 | 0.428 | 0.028 | 0.038 | 0.962 |
| | | $\alpha_1$ | 2.0 | 2.018 | 0.018 | 0.037 | 0.980 | 2.020 | 0.020 | 0.053 | 0.944 | 2.011 | 0.011 | 0.050 | 0.942 |
| | | $\alpha_2$ | 0.5 | 0.501 | 0.001 | 0.013 | 0.910 | 0.511 | 0.011 | 0.010 | 0.954 | 0.507 | 0.007 | 0.010 | 0.962 |
| | | $\alpha_3$ | −1.0 | −0.986 | 0.014 | 0.057 | 0.950 | −1.005 | 0.005 | 0.074 | 0.944 | −1.002 | 0.002 | 0.070 | 0.942 |
| | | $\alpha_4$ | −1.0 | −1.022 | 0.022 | 0.013 | 0.970 | −1.008 | 0.008 | 0.015 | 0.954 | −0.999 | 0.001 | 0.015 | 0.948 |
| 300 | 6 | $\beta_1^{01}$ | 1.0 | 0.992 | 0.008 | 0.081 | 0.934 | 0.985 | 0.015 | 0.070 | 0.954 | 0.973 | 0.027 | 0.074 | 0.942 |
| | | $\beta_2^{01}$ | −0.5 | −0.526 | 0.026 | 0.037 | 0.940 | −0.521 | 0.021 | 0.041 | 0.946 | −0.507 | 0.007 | 0.038 | 0.942 |
| | | $\beta_3^{01}$ | 0.5 | 0.517 | 0.017 | 0.142 | 0.936 | 0.531 | 0.031 | 0.134 | 0.956 | 0.506 | 0.006 | 0.122 | 0.956 |
| | | $\beta_1^{10}$ | −1.0 | −1.042 | 0.042 | 0.044 | 0.934 | −1.027 | 0.027 | 0.037 | 0.960 | −1.014 | 0.014 | 0.038 | 0.954 |
| | | $\beta_2^{10}$ | 0.5 | 0.513 | 0.013 | 0.021 | 0.940 | 0.518 | 0.018 | 0.019 | 0.956 | 0.518 | 0.018 | 0.020 | 0.954 |
| | | $\beta_3^{10}$ | −0.5 | −0.500 | 0.000 | 0.068 | 0.958 | −0.513 | 0.013 | 0.065 | 0.958 | −0.538 | 0.038 | 0.078 | 0.934 |
| | | $\Sigma_{11}$ | 0.6 | 0.660 | 0.060 | 0.101 | 0.950 | 0.655 | 0.055 | 0.105 | 0.942 | 0.675 | 0.075 | 0.104 | 0.964 |
| | | $\Sigma_{12}$ | −0.3 | −0.134 | 0.166 | 0.046 | 0.818 | −0.136 | 0.164 | 0.044 | 0.826 | −0.140 | 0.160 | 0.042 | 0.832 |
| | | $\Sigma_{22}$ | 0.4 | 0.452 | 0.052 | 0.070 | 0.940 | 0.456 | 0.056 | 0.063 | 0.972 | 0.456 | 0.056 | 0.063 | 0.960 |
| | | $\alpha_1$ | 2.0 | 2.040 | 0.040 | 0.062 | 0.942 | 2.012 | 0.012 | 0.049 | 0.964 | 2.009 | 0.009 | 0.046 | 0.958 |
| | | $\alpha_2$ | 0.5 | 0.496 | 0.004 | 0.012 | 0.942 | 0.500 | 0.000 | 0.012 | 0.952 | 0.512 | 0.012 | 0.011 | 0.940 |
| | | $\alpha_3$ | −1.0 | −1.028 | 0.028 | 0.090 | 0.920 | −0.997 | 0.003 | 0.071 | 0.954 | −0.996 | 0.004 | 0.068 | 0.946 |
| | | $\alpha_4$ | −1.0 | −1.006 | 0.006 | 0.016 | 0.952 | −1.011 | 0.011 | 0.018 | 0.936 | −1.003 | 0.003 | 0.017 | 0.940 |
| 300 | 10 | $\beta_1^{01}$ | 1.0 | 0.999 | 0.001 | 0.050 | 0.953 | 0.967 | 0.033 | 0.051 | 0.940 | 0.981 | 0.019 | 0.049 | 0.943 |
| | | $\beta_2^{01}$ | −0.5 | −0.531 | 0.031 | 0.024 | 0.947 | −0.524 | 0.024 | 0.024 | 0.947 | −0.512 | 0.012 | 0.022 | 0.953 |
| | | $\beta_3^{01}$ | 0.5 | 0.498 | 0.002 | 0.075 | 0.973 | 0.548 | 0.048 | 0.087 | 0.957 | 0.528 | 0.028 | 0.084 | 0.957 |
| | | $\beta_1^{10}$ | −1.0 | −1.036 | 0.036 | 0.028 | 0.933 | −1.033 | 0.033 | 0.024 | 0.933 | −1.005 | 0.005 | 0.026 | 0.930 |
| | | $\beta_2^{10}$ | 0.5 | 0.500 | 0.000 | 0.010 | 0.923 | 0.511 | 0.011 | 0.010 | 0.963 | 0.511 | 0.011 | 0.012 | 0.947 |
| | | $\beta_3^{10}$ | −0.5 | −0.496 | 0.004 | 0.040 | 0.950 | −0.500 | 0.000 | 0.038 | 0.950 | −0.531 | 0.031 | 0.043 | 0.940 |
| | | $\Sigma_{11}$ | 0.6 | 0.633 | 0.033 | 0.045 | 0.943 | 0.628 | 0.028 | 0.043 | 0.947 | 0.663 | 0.063 | 0.053 | 0.937 |
| | | $\Sigma_{12}$ | −0.3 | −0.206 | 0.094 | 0.023 | 0.877 | −0.197 | 0.103 | 0.022 | 0.850 | −0.204 | 0.096 | 0.019 | 0.913 |
| | | $\Sigma_{22}$ | 0.4 | 0.411 | 0.011 | 0.026 | 0.940 | 0.397 | 0.003 | 0.020 | 0.970 | 0.401 | 0.001 | 0.022 | 0.967 |
| | | $\alpha_1$ | 2.0 | 2.017 | 0.017 | 0.028 | 0.960 | 2.009 | 0.009 | 0.030 | 0.943 | 1.996 | 0.004 | 0.030 | 0.950 |
| | | $\alpha_2$ | 0.5 | 0.498 | 0.002 | 0.006 | 0.960 | 0.492 | 0.008 | 0.008 | 0.913 | 0.501 | 0.001 | 0.008 | 0.950 |
| | | $\alpha_3$ | −1.0 | −1.003 | 0.003 | 0.041 | 0.957 | −0.995 | 0.005 | 0.038 | 0.967 | −0.985 | 0.015 | 0.043 | 0.960 |
| | | $\alpha_4$ | −1.0 | −1.010 | 0.010 | 0.010 | 0.953 | −1.010 | 0.010 | 0.009 | 0.963 | −1.005 | 0.005 | 0.013 | 0.917 |

**Table 3** Simulated data: Proportions of greater LPML (WAIC) for models assuming MNAR than those assuming MAR. Null set: $\alpha$=(1, 0.5, 0.0, 0.0), for which the true model assumes MAR; Set 1: $\alpha$=(1, 0.5, −0.5, −0.5); Set 2: $\alpha$=(2, 0.5, −1, −1). Both Sets 1 and 2 assume MNAR for the true model

| | | LPML | | | WAIC | | |
|---|---|---|---|---|---|---|---|
| n | m | Null set | Set 1 | Set 2 | Null set | Set 1 | Set 2 |
| 100 | 6 | 0.16 | 0.87 | 1.00 | 0.10 | 0.75 | 1.00 |
| 100 | 10 | 0.17 | 0.93 | 1.00 | 0.11 | 0.87 | 1.00 |
| 200 | 6 | 0.13 | 0.99 | 1.00 | 0.09 | 0.95 | 1.00 |
| 200 | 10 | 0.13 | 1.00 | 1.00 | 0.04 | 0.98 | 1.00 |
| 300 | 6 | 0.11 | 1.00 | 1.00 | 0.05 | 0.98 | 1.00 |
| 300 | 10 | 0.07 | 1.00 | 1.00 | 0.01 | 1.00 | 1.00 |

parameters are same as those in the simulation studies. All summaries were based on 30 000 posterior samples for the parameters, which were obtained after a burn-in of 200 000 iterations and a thinning of every other 10 iterates.

We first considered a full MNAR model that assesses all prospective moderators (i.e., sex, age, marital status and income level) of the treatment effect via interaction terms. We found that marital status interacts with treatment significantly. Therefore, we refitted a reduced MNAR model which includes the main effects of all baseline covariates plus an interaction term: treatment × marital status. LPML and WAIC for the full model are −5 014 and −4 806, and for the reduced model are −5 020 and −4 827. The full model is slightly preferred. For the reduced model, Table 5 provides estimates of the posterior means and the associated 95% credible intervals for the transition parameters $\boldsymbol{\beta}^{01}$ and $\boldsymbol{\beta}^{10}$ (01: abstinent to smoking and 10: smoking to abstinent) and the missing model parameters $\boldsymbol{\alpha}$, respectively.

Based on the summaries for $\boldsymbol{\beta}^{01}$, $P(A \to S) \approx \exp(0.53)/(1 + \exp(0.53)) \approx 0.629$ for the reference group, which indicates that participants in the reference group have a probability of 62.9% to transition from abstinence to smoking. Age has a negative effect on $P(A \to S)$. Sex and income have no significant effect on resuming smoking or remaining abstinent. Marital status interacts with treatment. Married participants in the SRM group are more likely to resume smoking than those in the IRM group. Based on the 95% credible intervals of the contrasts in Table 6, $P(A \to S)$ of the *SRM* group and the *IRM* group for married participants are also significantly different.

Based on the summaries for $\boldsymbol{\beta}^{10}$, $P(S \to A) \approx 0.093$ for the reference group. If we modify the treatment for the reference group to TSH or SRM, the probability of quitting smoking decreases by 5.0% or 3.3%, respectively, that is, ($P(S \to A) \approx$ 0.043 or 0.060). The differences are also significant according to the 95% credible intervals of the contrasts in Table 6. Age and marital status have no significant effect on cessation. Male participants are more likely to quit in any of the three groups than female participants, as the probability of the coefficient for sex being greater than zero is 97%. Finally, participants with higher income are more likely to quit as the probability of the coefficient of income being greater than zero is 92%.

Based on the summaries for $\boldsymbol{\alpha}$ in Table 5 under MNAR, the coefficient for current smoking status is significantly negative, indicating that participants who are currently

**Table 4** Simulated data: True value and Monte Carlo mean, bias and mean square error (MSE) and 95% coverage probability of the transition event model and missing model parameters where random effects were simulated from $0.5\mathrm{N}(\boldsymbol{\mu}_1, \Sigma) + 0.5\mathrm{N}(\boldsymbol{\mu}_2, \Sigma)$, $\boldsymbol{\mu}_1 = (-0.5, 0.5)$, $\boldsymbol{\mu}_2 = (0.5, -0.5)$ and $\Sigma$ has diagonal elements 0.6 and 0.4, and off-diagonal elements $-0.3$

| n | m | Parameter | True Value | Mean | Bias | MSE | 95% CP |
|---|---|---|---|---|---|---|---|
| 200 | 6 | $\beta_1^{01}$ | 1.0 | 0.963 | 0.037 | 0.090 | 0.958 |
| | | $\beta_2^{01}$ | −0.5 | −0.520 | 0.020 | 0.031 | 0.942 |
| | | $\beta_3^{01}$ | 0.5 | 0.559 | 0.059 | 0.109 | 0.954 |
| | | $\beta_1^{10}$ | −1.0 | −1.043 | 0.043 | 0.058 | 0.944 |
| | | $\beta_2^{10}$ | 0.5 | 0.501 | 0.001 | 0.016 | 0.950 |
| | | $\beta_3^{10}$ | −0.5 | −0.521 | 0.021 | 0.059 | 0.940 |
| | | $\alpha_1$ | 2.0 | 2.400 | 0.400 | 0.748 | 0.936 |
| | | $\alpha_2$ | 0.5 | 0.502 | 0.002 | 0.016 | 0.954 |
| | | $\alpha_3$ | −1.0 | −1.373 | 0.373 | 0.843 | 0.942 |
| | | $\alpha_4$ | −1.0 | −1.017 | 0.017 | 0.030 | 0.954 |
| 200 | 10 | $\beta_1^{01}$ | 1.0 | 0.986 | 0.014 | 0.059 | 0.944 |
| | | $\beta_2^{01}$ | −0.5 | −0.512 | 0.012 | 0.017 | 0.948 |
| | | $\beta_3^{01}$ | 0.5 | 0.537 | 0.037 | 0.069 | 0.936 |
| | | $\beta_1^{10}$ | −1.0 | −1.070 | 0.070 | 0.042 | 0.928 |
| | | $\beta_2^{10}$ | 0.5 | 0.496 | 0.004 | 0.008 | 0.960 |
| | | $\beta_3^{10}$ | −0.5 | −0.484 | 0.016 | 0.035 | 0.932 |
| | | $\alpha_1$ | 2.0 | 2.279 | 0.279 | 0.441 | 0.928 |
| | | $\alpha_2$ | 0.5 | 0.497 | 0.003 | 0.010 | 0.952 |
| | | $\alpha_3$ | −1.0 | −1.272 | 0.272 | 0.520 | 0.920 |
| | | $\alpha_4$ | −1.0 | −0.999 | 0.001 | 0.018 | 0.944 |
| 300 | 6 | $\beta_1^{01}$ | 1.0 | 0.994 | 0.006 | 0.066 | 0.936 |
| | | $\beta_2^{01}$ | −0.5 | −0.528 | 0.028 | 0.022 | 0.948 |
| | | $\beta_3^{01}$ | 0.5 | 0.521 | 0.021 | 0.078 | 0.944 |
| | | $\beta_1^{10}$ | −1.0 | −1.049 | 0.049 | 0.040 | 0.948 |
| | | $\beta_2^{10}$ | 0.5 | 0.509 | 0.009 | 0.010 | 0.96 |
| | | $\beta_3^{10}$ | −0.5 | −0.497 | 0.003 | 0.035 | 0.948 |
| | | $\alpha_1$ | 2.0 | 2.283 | 0.283 | 0.492 | 0.948 |
| | | $\alpha_2$ | 0.5 | 0.511 | 0.011 | 0.012 | 0.940 |
| | | $\alpha_3$ | −1.0 | −1.286 | 0.286 | 0.564 | 0.950 |
| | | $\alpha_4$ | −1.0 | −0.993 | 0.007 | 0.019 | 0.946 |
| 300 | 10 | $\beta_1^{01}$ | 1.0 | 1.023 | 0.023 | 0.040 | 0.942 |
| | | $\beta_2^{01}$ | −0.5 | −0.522 | 0.022 | 0.014 | 0.932 |
| | | $\beta_3^{01}$ | 0.5 | 0.512 | 0.012 | 0.040 | 0.954 |
| | | $\beta_1^{10}$ | −1.0 | −1.056 | 0.056 | 0.028 | 0.918 |
| | | $\beta_2^{10}$ | 0.5 | 0.498 | 0.002 | 0.005 | 0.962 |
| | | $\beta_3^{10}$ | −0.5 | −0.494 | 0.006 | 0.022 | 0.954 |
| | | $\alpha_1$ | 2.0 | 2.178 | 0.178 | 0.229 | 0.960 |
| | | $\alpha_2$ | 0.5 | 0.504 | 0.004 | 0.007 | 0.942 |
| | | $\alpha_3$ | −1.0 | −1.181 | 0.181 | 0.268 | 0.962 |
| | | $\alpha_4$ | −1.0 | −0.997 | 0.003 | 0.011 | 0.952 |

smoking are more likely to have missing responses. Sex, income and time $t$ (months 6, 12, 18, and 24 were coded as $t = -1, 0, 1, 2$) all have negative effects, indicating that male participants are more likely to have missing responses than women; higher income participants are more likely to have missing responses than lower income participants; more responses are missing as time increases.

Lastly, the correlation between $u_{0i}$ and $u_{1i}$ ($\Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$) has an estimated value $-0.536$ and 95% credible interval $(-0.827, -0.156)$. The correlation is significantly less than zero, which implies that entering abstinent status is negatively correlated with entering smoking status. The standard deviation of $u_{0i}$ ($\sqrt{\Sigma_{11}}$) has an estimated value 3.248 and 95% credible interval (2.241, 4.617). The standard deviation of $u_{1i}$ ($\sqrt{\Sigma_{22}}$) has an estimated value 1.652 and 95% credible interval (1.347, 1.966).

Using the MCMC samples for the reduced MNAR model, we can compute the population averaged probabilities of smoking patterns for every subgroup. To illustrate, we focus on patterns '$S \to A \to A \to A \to A$' and '$S \to S \to S \to S \to S$' for each treatment group, while other variables are held at reference values. Given $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$ and $\Sigma$, the probability of observing pattern $S \to A \to A \to A \to A$ is

$$\int \frac{\exp(\mathbf{x}^T\boldsymbol{\beta}^{10} + u_1)}{1 + \exp(\mathbf{x}^T\boldsymbol{\beta}^{10} + u_1)} \left[\frac{1}{1 + \exp(\mathbf{x}^T\boldsymbol{\beta}^{01} + u_0)}\right]^3 N(\mathbf{u}; \mathbf{0}, \Sigma) d\mathbf{u},$$

where $\mathbf{u} = (u_0, u_1)$. We approximate the integration above by

$$\frac{1}{L}\sum_{l=1}^{L}\left\{\frac{\exp(\mathbf{x}^T\boldsymbol{\beta}^{10} + u_1^{(l)})}{1 + \exp(\mathbf{x}^T\boldsymbol{\beta}^{10} + u_1^{(l)})}\left[\frac{1}{1 + \exp(\mathbf{x}^T\boldsymbol{\beta}^{01} + u_0^{(l)})}\right]^3\right\}$$

where $\mathbf{u}^1, \ldots, \mathbf{u}^{(L)}$ are independent samples drawn from $N(\mathbf{0}, \Sigma)$. Using the MCMC samples of $\boldsymbol{\beta}^{01}$, $\boldsymbol{\beta}^{10}$ and $\Sigma$, we obtain the posterior mean and the 95% credible interval of the probability for each pattern. Table 7 displays the results for each smoking cessation group. The IRM group has the highest estimated probability of '$S \to A \to A \to A \to A$' pattern and lowest estimated probability for '$S \to S \to S \to S \to S$'.

To assess the goodness of fit of the reduced MNAR model, we use the proposed posterior predictive checking method in Section 2.4 and 6 000 replicated datasets. Table 8 displays the summaries of smoking percentages and response rates (percentages of participants who responded smoking status) for both the observed and replicated data. We find a minimal discrepancy between the observed and replicated data. We also see no discrepancy between the observed and replicated data in response rates. In addition, we displayed in Figure 2 the summaries of transition probabilities among smoking, abstinent and missing. The transition probabilities between smoking and abstinent are well matched by the replicated data. The transition probabilities between missing and smoking statuses are not fully predicted by the replicated data.

For sensitivity analysis, we first compare the reduced MNAR model to a MAR version of the model. The LPML and WAIC of the MAR model are $-5\,056$ and $-4\,866$. Therefore the reduced MNAR model's predictive ability is better than that of reduced MAR model. The posterior predictive checking of the reduced MAR model in Table 8 and the transition map in Figure 3 also suggest that the MNAR model fits the observed data slightly better than the MAR model. We present the parameter estimation results of the reduced MAR model in Tables 5 and 6.

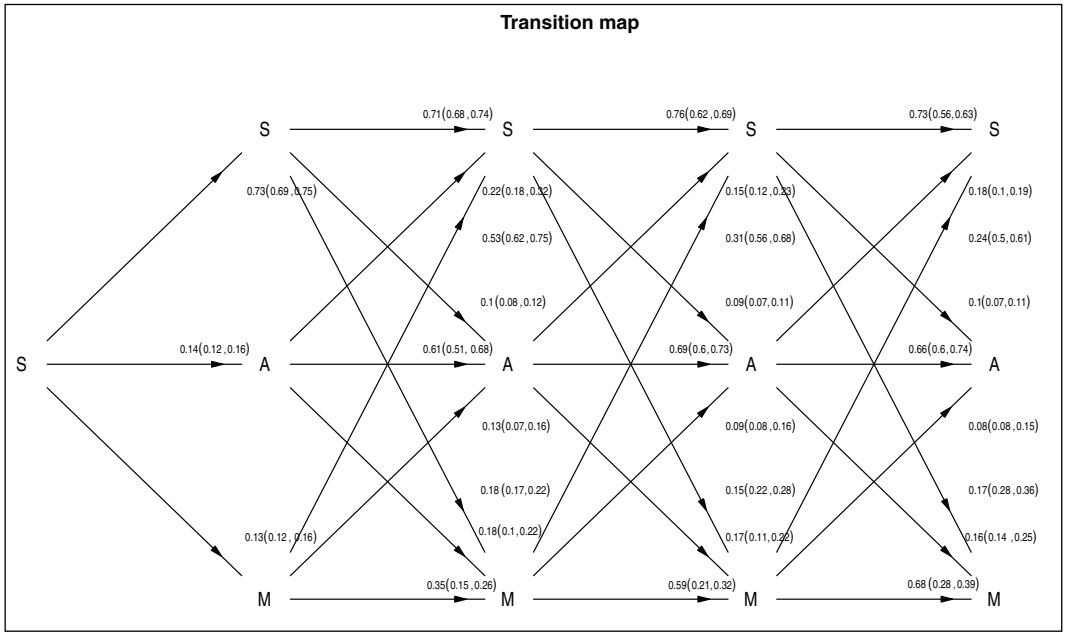


**Figure 2** Smoking status transition map based on MNAR assumption

**Notes:** S—smoking; A—abstinent; M—missing; arrows are the transitions; predictive summaries of the transitions are placed in the same vertical orders of the corresponding arrows; the numeric values outside the brackets are observed transition percentages; the numeric values inside the brackets are 95% sample intervals of transition percentages of the replicated data.

We also perform sensitivity analyses on hyper-parameters under the reduced MNAR model. The resulting LPML and WAIC of each set-up is represented in Table 9. We noticed very small changes in the LPML and WAIC values. Model parameter estimates also show no change under different set-ups of $c_{\boldsymbol{\beta}}^2$ and $\varsigma_j^2$. When we increase or decrease $c_{\boldsymbol{\alpha}}^2$, we noticed small changes in the estimates. Results of the reduced MNAR model under $c_{\boldsymbol{\alpha}}^2 = 3^2$ and $5^2$ are represented in Table 10.

Finally, we assess the MCMC convergence of the samples for the reduced MNAR model. Geweke' $z$-statistics (Geweke, 1992) for the parameters are displayed in Figure 4 which suggests convergence for all parameters. The $z$-statistic of $\boldsymbol{\beta}^{01}$ (marital status SRM) is a little larger than two but the MCMC chain shows no obvious poor mixing.

## 5 Conclusion

We proposed an extension of a two-state transition model to manage missing smoking status and applied the model to a smoking cessation intervention study. We further proposed an MCMC algorithm for obtaining accurate Bayesian inferences

**Table 5** Smoking cessation data: Posterior mean (95% credible interval) for model parameters with treatments interacting with marital status based on the reduced MNAR and MAR model

| | MNAR | | | MAR | | |
|---|---|---|---|---|---|---|
| | $\beta^{01}$ | $\beta^{10}$ | $\alpha$ | $\beta^{01}$ | $\beta^{10}$ | $\alpha$ |
| Intercept (IRM) | 0.53 (−0.90, 1.92) | −2.28 (−2.60, −1.97) | 4.18 ( 2.61, 7.10) | −1.42 (−3.21, 0.23) | −2.00 (−2.34, −1.68) | 1.52 ( 1.39, 1.66) |
| Intercept (TSH) | 1.53 (−0.08, 3.32) | −3.09 (−3.46, −2.74) | 4.37 ( 2.76, 7.30) | −0.20 (−2.32, 1.86) | −2.88 (−3.26, −2.52) | 1.58 ( 1.46, 1.72) |
| Intercept (SRM) | 0.28 (−1.42, 1.89) | −2.74 (−3.10, −2.40) | 4.41 ( 2.83, 7.34) | −1.96 (−4.23, 0.07) | −2.55 (−2.93, −2.21) | 1.69 ( 1.55, 1.82) |
| Age | −0.47 (−0.91, −0.06) | −0.07 (−0.06, 0.21) | 0.17 ( 0.11, 0.24) | −0.53 (−1.15, 0.02) | 0.04 (−0.11, 0.18) | 0.18 ( 0.12, 0.24) |
| Sex | −0.34 (−1.29, 0.50) | 0.27 (−0.02, 0.55) | −0.36 (−0.50, −0.23) | −0.95 (−2.19, 0.17) | 0.34 ( 0.05, 0.64) | −0.30 (−0.43, −0.17) |
| Marital status (IRM) | −0.01 (−1.33, 1.29) | −0.18 (−0.66, 0.30) | 0.04 (−0.12, 0.19) | −0.41 (−2.18, 1.30) | −0.24 (−0.75, 0.25) | 0.02 (−0.13, 0.17) |
| Marital status (TSH) | −0.00 (−1.60, 1.64) | 0.11 (−0.41, 0.63) | * | −0.66 (−2.81, 1.41) | 0.09 (−0.47, 0.64) | * |
| Marital status (SRM) | 2.07 ( 0.46, 3.94) | 0.23 (−0.30, 0.78) | * | 2.38 ( 0.35, 4.68) | 0.27 (−0.28, 0.82) | * |
| Income | 0.03 (−0.43, 0.50) | 0.11 (−0.04, 0.26) | −0.16 (−0.23, −0.09) | −0.01 (−0.63, 0.63) | 0.16 ( 0.00, 0.32) | −0.13 (−0.20, −0.07) |
| Time $t$ | * | * | −0.35 (−0.41, −0.30) | * | * | −0.32 (−0.37, −0.26) |
| $I(y_{ti} = 1)$ | * | * | −3.58 (−6.05, −1.77) | * | * | * |
| $I(y_{t-1,i} = 1)$ | * | * | 0.68 ( 0.38, 0.98) | * | * | * |

**Note:** * means that the corresponding parameter is not estimated. Blank indicates that values are not defined. $I(\cdot)$ is an indicator function.

**Table 6** Smoking cessation data: Posterior mean (95% credible interval) for contrasts of the coefficients for intercept and marital status between TSH and IRM, and SRM and IRM

| | MNAR | | MAR | |
|---|---|---|---|---|
| | $\beta^{01}$ | $\beta^{10}$ | $\beta^{01}$ | $\beta^{10}$ |
| Intercept (TSH − IRM) | 1.00 (−0.19, 2.32) | −0.82 (−1.24, −0.41) | 1.23 (−0.31, 2.90) | −0.88 (−1.31, −0.46) |
| Intercept (SRM −IRM) | −0.25 (−1.48, 0.94) | −0.46 (−0.88, −0.06) | −0.53 (−2.17, 1.03) | −0.56 (−0.98, −0.14) |
| Marital status (TSH − IRM) | −0.00 (−1.94, 1.89) | 0.29 (−0.39, 0.97) | −0.26 (−2.77, 2.21) | 0.33 (−0.37, 1.04) |
| Marital status (SRM − IRM) | 2.06 ( 0.16, 4.24) | 0.41 (−0.27, 1.10) | 2.79 ( 0.36, 5.50) | 0.51 (−0.19, 1.23) |

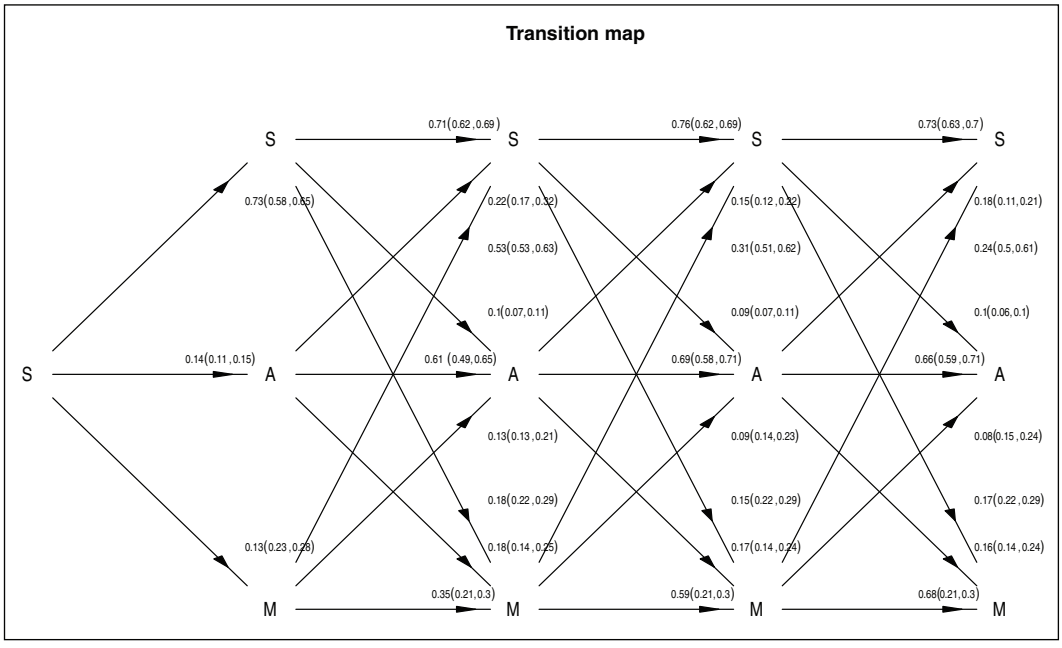


**Figure 3** Smoking status transition map based on MAR assumption

**Notes:** S—smoking; A—abstinent; M—missing; arrows are the transitions; predictive summaries of the transitions are placed in the same vertical orders of the corresponding arrows; the numeric values outside the brackets are observed transition percentages; the numeric values inside the brackets are 95% sample intervals of transition percentages of the replicated data.

under MAR and MNAR missing model assumption. We developed a posterior predictive checking to assess model assumption and evaluated the performances of LPML and WAIC in model comparison. The proposed model can be used to evaluate the effectiveness of smoking interventions by comparing the estimated transitioning probabilities of smoking behaviours and probabilities of longitudinal smoking behaviour patterns. The application of this method to a randomized controlled intervention trial for smoking cessation provides information on how the interventions influence the transition of smoking behaviours. The results are

**Table 7** Smoking cessation data: Posterior means and 95% credible intervals for the probability of observing patterns '$S \to A \to A \to A \to A$' and '$S \to S \to S \to S \to S$' based on the reduced MNAR model . The probabilities are for a female population with age 48 and annual income US $30 000–40 000

| Marital status | Smoking pattern | IRM | SRM | TSH | TSH − IRM | SRM − IRM |
|---|---|---|---|---|---|---|
| Not married | $S \to A \to A \to A \to A$ | 0.08 (0.05, 0.11) | 0.07 (0.04, 0.09) | 0.04 (0.02, 0.06) | −0.04 (−0.07, −0.02) | −0.01 (−0.04, 0.01) |
| | $S \to S \to S \to S \to S$ | 0.60 (0.53, 0.66) | 0.67 (0.62, 0.73) | 0.73 (0.68, 0.78) | 0.14 ( 0.07, 0.20) | 0.08 ( 0.01, 0.15) |
| Married | $S \to A \to A \to A \to A$ | 0.07 (0.05, 0.11) | 0.04 (0.02, 0.07) | 0.04 (0.02, 0.07) | −0.03 (−0.06, 0.00) | −0.03 (−0.07, 0.00) |
| | $S \to S \to S \to S \to S$ | 0.63 (0.55, 0.70) | 0.63 (0.55 , 0.71) | 0.71 (0.64 , 0.78) | 0.09 ( 0.00, 0.18) | 0.01 (−0.09, 0.10) |

**Table 8** Values in columns 'observed' are smoking percentages and response rates by questionnaire time of the smoking cessation data. Values in columns 'MNAR' and 'MAR' are sample means and 95% sample intervals of smoking percentages and response rates of the replicated data

| | Smoking percentage | | | Response percentage | | |
|---|---|---|---|---|---|---|
| time in month | Observed | MNAR | MAR | Observed | MNAR | MAR |
| 6 | 0.733 | 0.719 (0.690, 0.748) | 0.616 (0.584, 0.646) | 0.871 | 0.859 (0.837, 0.880) | 0.742 (0.715, 0.769) |
| 12 | 0.620 | 0.638 (0.609, 0.667) | 0.582 (0.551, 0.612) | 0.796 | 0.812 (0.790, 0.834) | 0.753 (0.726, 0.779) |
| 18 | 0.563 | 0.565 (0.536, 0.595) | 0.550 (0.519, 0.581) | 0.758 | 0.763 (0.738, 0.786) | 0.756 (0.730, 0.782) |
| 24 | 0.505 | 0.496 (0.462, 0.528) | 0.533 (0.500, 0.564) | 0.710 | 0.703 (0.675, 0.733) | 0.759 (0.733, 0.784) |

**Table 9** Smoking cessation data: Reduced MNAR model performance under different set-ups of hyper-parameters

| | | | MNAR | |
|---|---|---|---|---|
| $\varsigma_j^2$ | $c_\beta$ | $c_\alpha$ | LPML | WAIC |
| $10^3$ | $5^2$ | $4^2$ | −5020 | −4827 |
| $10^2$ | $5^2$ | $4^2$ | −5019 | −4827 |
| $10^4$ | $5^2$ | $4^2$ | −5019 | −4827 |
| $10^3$ | $4^2$ | $4^2$ | −5019 | −4828 |
| $10^3$ | $6^2$ | $4^2$ | −5021 | −4827 |
| $10^3$ | $5^2$ | $3^2$ | −5021 | −4831 |
| $10^3$ | $5^2$ | $5^2$ | −5019 | −4825 |

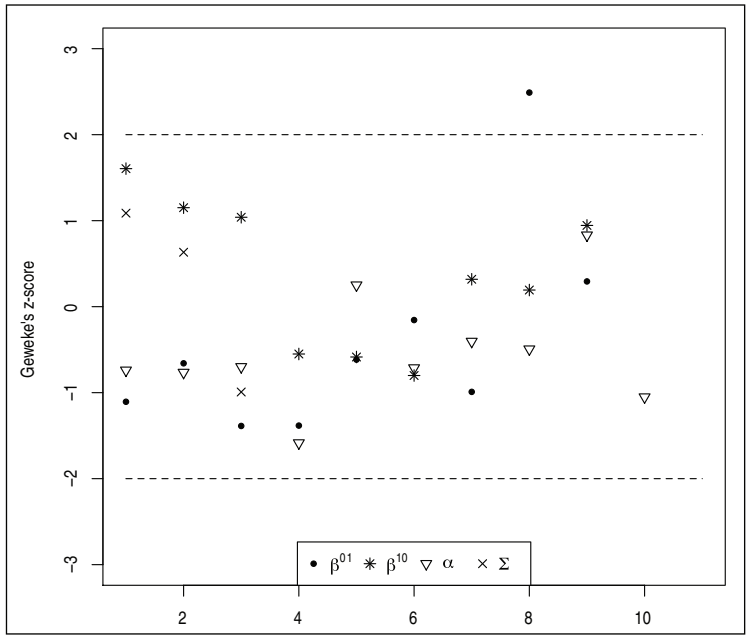


**Figure 4** Smoking cessation data: Geweke' z-statistic for each parameter in the reduced MNAR model using the first 10% and last 50% of the MCMC chain

consistent with the hypothesis and conclusions from the previous study (Brandon et al., 2016) in which the IRM group (intensive intervention) produced the best outcomes, followed by the SRM group (standard intervention) and then the TSH group (usual care). But unlike the original analyses, our analyses provide a further understanding of the transitions of smoking statuses between each follow-up point. For example, the pair-wise comparisons at each time point in the original analyses reported that outcome of the 7-day-point-prevalence abstinence rate of the IRM group is significantly better than that of the SRM group at 18 months. Our analyses suggest that the difference is mostly due to the significant difference in the

**Table 10** Smoking cessation data: Posterior mean (95% credible interval) for model parameters with treatments interacting with marital status based on the reduced MNAR model under $c_\alpha = 3^2$ and $5^2$

| | $c_\alpha = 3^2$ | | | $c_\alpha = 5^2$ | | |
|---|---|---|---|---|---|---|
| | $\beta^{01}$ | $\beta^{10}$ | $\alpha$ | $\beta^{01}$ | $\beta^{10}$ | $\alpha$ |
| Intercept (IRM) | 0.42 (−1.13, 1.87) | −2.27 (−2.60, −1.95) | 3.68 ( 2.34, 5.60) | 0.58 (−0.84, 1.93) | −2.28 (−2.61, −1.97) | 4.58 ( 2.64, 8.53) |
| Intercept (TSH) | 1.46 (−0.30, 3.26) | −3.09 (−3.47, −2.73) | 3.87 ( 2.51, 5.81) | 1.55 ( 0.06, 3.26) | −3.10 (−3.48, −2.74) | 4.77 ( 2.80, 8.73) |
| Intercept (SRM) | 0.15 (−1.76, 1.86) | −2.74 (−3.10, −2.40) | 3.91 ( 2.56, 5.85) | 0.35 (−1.39, 1.94) | −2.74 (−3.10, −2.40) | 4.81 ( 2.85, 8.77) |
| Age | −0.47 (−0.96, −0.06) | −0.07 (−0.06, 0.21) | 0.17 ( 0.10, 0.24) | −0.47 (−0.90, −0.07) | −0.07 (−0.06, 0.21) | 0.17 ( 0.11, 0.24) |
| Sex | −0.40 (−1.39, 0.48) | 0.27 (−0.01, 0.56) | −0.36 (−0.50, −0.23) | −0.31 (−1.22, 0.51) | 0.27 (−0.01, 0.56) | −0.36 (−0.50, −0.23) |
| Marital status (IRM) | −0.02 (−1.39, 1.34) | −0.18 (−0.66, 0.30) | 0.04 (−0.12, 0.19) | −0.00 (−1.29, 1.27) | −0.18 (−0.66, 0.30) | 0.04 (−0.12, 0.19) |
| Marital status (TSH) | −0.05 (−1.77, 1.60) | 0.11 (−0.42, 0.64) | * | −0.01 (−1.58, 1.57) | 0.12 (−0.41, 0.64) | * |
| Marital status (SRM) | 2.15 ( 0.51, 4.11) | 0.24 (−0.30, 0.78) | * | 2.00 ( 0.44, 3.84) | 0.23 (−0.31, 0.77) | * |
| Income | 0.03 (−0.43, 0.52) | 0.11 (−0.04, 0.26) | −0.15 (−0.23, −0.08) | 0.03 (−0.40, 0.47) | 0.11 (−0.04, 0.26) | −0.16 (−0.23, −0.09) |
| Time $t$ | * | * | −0.35 (−0.41, −0.29) | * | * | −0.35 (−0.41, −0.29) |
| $I(y_{ti} = 1)$ | * | * | −3.03 (−5.11, −1.42) | * | * | −4.00 (−8.04, −1.81) |
| $I(y_{t-1,i} = 1)$ | * | * | 0.64 ( 0.30, 0.94) | * | * | 0.70 ( 0.38, 0.98) |

**Note:** Blank indicates that values are not defined. $I(\cdot)$ is an indicator function. * means that the corresponding parameter is not estimated.

transition probability $P(S \to A)$, instead of the difference in the transition probability $P(A \to S)$ or $P(A \to A)$. In other words, IRM is superior to SRM in helping people quit smoking, but the intervention is not significantly better at helping people to remain abstinent. The original study has many other covariates in addition to those used in our models. One of the challenges of including many covariates is dealing with missing observations. In a Bayesian framework, the challenge lies in the fact that we need to specify a flexible joint distribution for incomplete categorical and continuous variables that will be used for imputation. Our future research will focus on extending the current model to allow missing covariates. We are also interested in extending the current missing model of the response to a flexible class of missing data models proposed in Ibrahim et al. (2001). Instead of assuming independent binary distribution of missing indicators $\delta_{1i}, \ldots, \delta_{mi}$, one can assume a joint distribution of the missing indicators, for example, $p(\delta_{1i}, \ldots, \delta_{mi}|\mathbf{y}_i, \mathbf{z}_i) = p(\delta_{1i}|\mathbf{y}_i, \mathbf{z}_{1i}, \boldsymbol{\gamma}_1)p(\delta_{2i}|\delta_{1i}, \mathbf{y}_i, \mathbf{z}_{1i}, \boldsymbol{\gamma}_2)\ldots p(\delta_{mi}|\delta_{1i}, \ldots, \delta_{m-1,i}, \mathbf{y}_i, \mathbf{z}_{mi}, \boldsymbol{\gamma}_m)$ where each conditional distribution is modeled using a logistic regression.

## Supplementary material

We provide the codes for a simulation example via the link `http://www.statmod.org/smij/archive.html`.

## Acknowledgement

The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Declaration of conflicting interests

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

## Funding

The research time of Li and Lee was supported in part by the National Cancer Institute grant (5P30CA118100-11; the National Cancer Institute, USA; PI: Willman). The dataset analyzed herein was from the grant R01-CA134347 (the National Cancer Institute, USA; PI: T Brandon). Original data collection was also supported in part by the Biostatistics and Survey Methods Core Facility at the H. Lee Moffitt Cancer Center and Research Institute, a National Cancer Institute-designated Comprehensive Cancer Center (P30CA76292).

## References


Azzalini A (1994) Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, **81**, 767–75.

Antonelli J, Trippa L and Haneuse S (2016) Mitigating bias in generalized linear mixed models: The case for Bayesian nonparametrics. *Statistical Science*, **31**, 80–95.

Beunckens C, Sotto C and Molenberghs G (2008) A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics and Data Analysis*, **52**, 1533–48.

Brandon TH, Simmons VN, Sutton SK, Unrod M, Harrell PT, Meade CD, Craig BM, Lee JH and Meltzer LR (2016) Extended self-help for smoking cessation: A randomized controlled trial. *American Journal of Preventive Medicine*, **51**, 54–62.

Carpenter JR, Kenward MG and Vansteelandt S (2006) A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **169**, 571–84.

Diggle P and Kenward MG (1994) Informative drop-out in longitudinal data analysis. *Applied Statistics*, 49–93.

Diggle P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data, 2nd edition*. New York, NY: Oxford university press.

Daniels MJ and Hogan JW (2008) *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. CRC Press.

Fiore MC, Jaen CR, Baker T, Bailey WC, Benowitz NL, Curry, SEEA, ... and Henderson PN (2008) *Treating Tobacco Use and Dependence: 2008 update*. Rockville, MD: US Department of Health and Human Services.

Gelfand AE and Dey DK (1994) Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **56**, 501–14.

Gelman A, Hwang J, and Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, **24**, 997–1016.

Gelman A (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, **1**, 515–34.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB (2014) *Bayesian Data Analysis*, vol. 2. Boca Raton, FL: CRC press.

Geisser S and Eddy WF (1979) A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–60.

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics*, edited by JM Bernado, JO Berger, AP Dawid and AFM Smith. Vol. 4, pages 169–193. Oxford: Clarendon Press.

Haario H, Saksman E and Tamminen J (2005) Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, **20**, 265–73.

Heagerty PJ (2002) Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics*, **58**, 342–51.

Huang A and Wand MP (2013) Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, **8**, 439–52.

Hartmann-Boyce J, Lancaster T and Stead LF (2014) Print-based self-help interventions for smoking cessation. The Cochrane Library. doi: 10.1002/14651858.

Ibrahim JG, Chen MH and Lipsitz SR (2001) Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, **88**, 551–64.

Ibrahim JG and Molenberghs G (2009) Missing data methods in longitudinal studies: A review. *TEST*, **18**, 1–43.

Jha P and Peto R (2014) Global effects of smoking, of quitting, and of taxing tobacco. *New England Journal of Medicine*, **370**, 60–8.

Jansen I, Beunckens C, Molenberghs G, Verbeke G and Mallinckrodt C (2006) Analyzing incomplete discrete longitudinal clinical trial data. *Statistical Science*, **21**, 52–69.

Kinney SK and Dunson DB (2007) Fixed and random effects selection in linear and logistic models. *Biometrics*, **63**, 690–98.

Kurland BF and Heagerty PJ (2004) Marginalized transition models for longitudinal binary data with ignorable and non-ignorable drop?out. *Statistics in Medicine*, **23**, 2673–95.

Linero AR and Daniels MJ (2015) A flexible Bayesian approach to monotone missing data in longitudinal studies with nonignorable missingness with application to an acute schizophrenia clinical trial. *Journal of the American Statistical Association*, **110**, 45–55.

McCulloch CE and Neuhaus JM (2011) Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, **26**, 388–402.

Polson NG, Scott JG and Windle J (2013) Bayesian inference for logistic models using Polya–Gamma latent variables. *Journal of the American Statistical Association*, **108**, 1339–49.

Papadakis S, McDonald P, Mullen KA, Reid R, Skulsky K and Pipe A (2010) Strategies to increase the delivery of smoking cessation treatments in primary care settings: A systematic review and meta-analysis. *Preventive Medicine*, **51**, 199–213.

Robins JM, Rotnitzky A and Zhao LP (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106–21.

Steele F (2011) Multilevel discrete-time event history models with applications to the analysis of recurrent employment transitions. *Australian and New Zealand Journal of Statistics*, **53**, 1–20.

Steele F, Goldstein H and Browne W (2004) A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, **4**, 145–59.

Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**, 3571–94.

Xu D, Chatterjee A and Daniels M (2016) A note on posterior predictive checks to assess model fit for incomplete data. *Statistics in Medicine*, **35**, 5029–39.

Yeh HW, Ellerbeck EF and Mahnken JD (2012) Simultaneous evaluation of abstinence and relapse using a Markov chain model in smokers enrolled in a two-year randomized trial. *BMC Medical Research Methodology*, **12**, 95.