Health Serv Outcomes Res Method (2018) 18:227–243
https://doi.org/10.1007/s10742-018-0184-5

A shared parameter location scale mixed effect model for EMA data subject to informative missing

Xiaolei Lin[1] · Robin Mermelstein[2] · Donald Hedeker[1]

Received: 16 March 2018 / Revised: 20 May 2018 / Accepted: 29 May 2018 / Published online: 7 June 2018

Abstract In this paper, we address the problem of accounting for informative missing in the context of ecological momentary assessment studies (sometimes referred to as intensive longitudinal studies), where each study unit gets measured intensively over time and intermittent missing is usually present. We present a shared parameter modeling approach that links the primary longitudinal outcome with potentially informative missingness by a common set of random effects that summarize a subjects' specific traits in terms of their mean (location) and variability (scale). The primary outcome, conditional on the random effects, are allowed to exhibit heterogeneity in terms of both the mean and within subject variance. Unlike previous methods which largely rely on numerical integration or approximation, we estimate the model by a full Bayesian approach using Markov Chain Monte Carlo. An adolescent mood study example is illustrated together with a series of simulation studies. Results in comparison to more conventional approaches suggest that accounting for the common but unobserved random subject mean and variance effects, shared between the primary outcome and missingness models, can significantly improve the model fit, and also provide the benefit of understanding how missingness can affect the inference for the primary outcome.

Keywords Ecological momentary assessments · Informative missing · Mixed effects · Shared parameter model

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10742-018-0184-5) contains supplementary material, which is available to authorized users.

✉ Xiaolei Lin
xlin3@uchicago.edu

Robin Mermelstein
robinm@uic.edu

Donald Hedeker
dhedeker@health.bsd.uchicago.edu

[1] Department of Public Health Sciences, The University of Chicago, Chicago, USA

[2] Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, USA

1 Introduction

Many scientific investigations generate longitudinal data with missing values, either by intermittent missing or in the form of drop out (Laird 1988). In either case, if subjects with missing values behave different in terms of the primary outcome compared to those without, the conventional statistical methods that assumes missing completely at random (MCAR) would yield invalid inference. Furthermore, if missing observations have different distributions conditional on the observed data, models assuming missing at random (MAR) would also not be valid. Modern data collection procedures, such as ecological momentary assessments (EMA), allow researchers to study psychological and behavioral outcomes by repeated sampling in real time fashion (Shiffman et al. 2008). Typically these procedures involve instant short surveys from individuals over the course of hours, days, and weeks, where relatively large numbers of measurements per subject are produced and intermittent missingness due to non-responses can be an issue (Sokolovsky et al. 2014). Subjects with a substantial proportion of non-responses could be systematically different in terms of the outcomes compared to those without. An intuitive suspicion would be that, subjects with worse behavioral outcomes or at the occasions when they are experiencing higher levels of stress might respond less often. In psychological and behavioral sciences, within subject variation is another critical metric in characterizing the primary mental outcomes (Martin and Hofer 2004). Therefore, within subject variation of the primary outcome can further diverge depending on the missingness. For instance, subjects with unstable outcomes could respond less often compared to those with relatively consistent outcomes. However, research about the informative and intermittent missingness with respect to both the mean and within subject variation of the primary outcomes is rather limited.

In general, three modeling frameworks can be used under the scenario of informative missing. Selection models, first originated in econometrics by Heckman (1979), and later formulated in a longitudinal setting by Diggle and Kenward (1994), assume that the observed outcomes are subject to selection bias and a drop out model is introduced to correct for this bias. Pattern mixture models proposed by Little (1993), partition the joint distribution of the primary outcome and missing process into distinct missing patterns, and compute the joint likelihood conditional on each pattern. Shared parameter models assume that there is a set of latent variables U shared between the primary outcome and missing process, which are conditionally independent given U (Vonesh et al. 2006). Unlike time to event studies where there are relatively small numbers of missing patterns, in EMA studies intermittent missingness is common and the number of missing patterns could be intrinsically large, making it hard to employ pattern mixture models. We thus resort to shared parameter models, which turn out to make more intuitive sense. Consider the scenario where the primary outcome of interest is mood assessments and data can be intermittently missing due to non-responses. The shared parameter model would assume that some common but unobserved information is shared between subject's mood and the probability of missing, and conditional on the latent information, mood can be modeled independently of the missing process. Recent work has shown that sharing subject's specific location traits between the primary mood outcome and the missing process can significantly improve the model fit (Cursio et al. 2018). Our model extends this work by additionally allowing covariates to influence the within subject variance, by including random subject scale (variance) effects, by allowing the random location and scale effects to influence missingness, and by adopting a Bayesian model estimation framework instead of maximum likelihood methods. In particular, no work has been done to explore the possible effect of a subject's scale on the missing process. Furthermore, we investigate the use of the model to impute the missing responses.

Shared parameter models can be difficult to implement due to marginalization of the random effects. Wu and Carroll (1988) utilized a maximum likelihood estimation method by numerical integration. Follmann and Wu (1995) approximated the generalized linear model by conditioning on the data that describes missingness. Pulkstenis et al. (1998) derived a closed form expression of the marginal likelihood by specifying conjugate random effects for both the outcome and missing process. Unlike these methods that are either sensitive to starting values or restrictive in terms of the random effects distributional assumptions, we propose to estimate the model by a full Bayesian approach without undue restrictions on the distribution of the primary outcome, missing process, or random effects.

In this paper, we develop a comprehensive Bayesian approach to shared parameter models, where the primary outcome and missing process can be related in terms of both the subject's location and scale random effects. The expansion of sharing to include additional scale information makes practical sense in the context of EMA and psychological studies (Nesselroade 2004). The primary outcomes are allowed to follow a variety of distributions with heterogeneous error variance, where the heterogeneity is characterized by a location and scale random effect, respectively (Hedeker et al. 2008). The intermittent missing process is modelled by a logistic regression model including a random subject effect. The missingness random effect is then linked with the outcome location and scale random effects to allow for informative missingness. With the full Bayesian approach, posterior distributions of the model parameters as well as random effects are obtained by MCMC (Bradley and Siddhartha 1995).

2 Motivating example

This research is motivated by an EMA study investigating the effects of psychosocial factors on mood regulation among adolescents. The entire EMA study was conducted across 6 waves: baseline, 6, 15 months, 2, 5 and 6 years. For illustration purposes, we will focus on data from the baseline wave.

At baseline, 461 adolescents (average age 15.6, minimum 14.4, maximum 16.7) from 9th and 10th grade were asked to carry electronic devices and answer questions when randomly prompted during a 7 day study period. Each individual was prompted multiple times within a single day. Questions included location, activities, companionship, mood and other psychological assessments. The primary outcomes of interest are positive affect (PA) as well as negative affect (NA), which consist of the average of several mood items rated from 1 to 10 that measures subject's positive/negative mood. For PA, questions include: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others; for NA, questions include: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable. Higher PA levels indicates better mood while higher NA indicates worse mood. Each response will be time stamped regardless of missing status. For the analyses presented here, we will consider subject-level covariates of age, gender, smoking status, negative mood regulation, and the occasion-varying indicator of whether they were alone or with others at the time of the prompt.

Intermittent missingness was generated when an individual did not respond to the prompts. Dropout is not as big a concern here as more than 96.5% individuals were still available at the end of the 7 day study period. On average, 22.3% prompts were missing for each individual, with the highest missing proportion per individual being 89.7%. There were a fair number of prompts missing on each study day: 22.6% on day 1, 19.0% on day 2, 20.9% on day 3, 24.6% on day 4, 25.6% on day 5, 25.0% on day 6 and 23.9% on day 7. The proportions of missing prompts were relatively similar during weekdays. In terms of time of day, most missingness occurred between 3 and 9 a.m. (27.2%) and least from 6 to 9 p.m. (20.8%), but the pattern on weekend days was different in that missingness occurred mostly between 9 p.m. and 3 a.m. (52.9% on Saturday and 73.1% on Sunday).

In what follows, we propose a shared parameter model that links the primary outcome with the missing process through a set of random subject effects. For simplicity, we will illustrate the model framework using a normally distributed outcome and binary missing indicator in the context of the example EMA study. However, it can also be extended to other outcome types (e.g., binary or Poisson). We then illustrate a full Bayesian approach for model estimation using MCMC. A series of simulation studies are presented to validate the model estimation procedure, and to examine the use of the model for imputation of missing observations. Finally, the proposed model is applied to the adolescent mood EMA study and results are compared to naive analyses where only the observed data are used.

3 Shared parameter model

We present the methodology along the lines of Follmann and Wu (1995), but for a normally distributed and intensively measured longitudinal outcome and binary intermittent missing indicator, which often arise in the context of EMA studies. The approach is to specify the outcome and missing models that share the same set of random effects for each individual.

3.1 Model for intensively measured longitudinal outcomes

Let Y_{ij} be the outcome for individual i at occasion j, where $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$ (we allow different individuals to have different number of measurements by subscript n with i). Examples might include mood assessments (PA/NA), craving for food, depression scores, or other psychological measurements. Since Y is measured intensively over time, most of the n_i would be large (usually 20–40) compared to traditional longitudinal studies. We specify a mixed effect location scale model for Y_{ij} as described in Hedeker et al. (2008):

$$Y_{ij} \mid \left\{ \nu_{1,i}, \nu_{2,i} \right\} \sim \mathcal{N}\left(X_{ij}^{\top}\beta + P_{ij}^{\top}\nu_{1,i};\ \exp(Z_{ij}^{\top}\alpha + Q_{ij}^{\top}\nu_{2,i}) \right) \tag{1}$$

where X_{ij} and Z_{ij} are the fixed effect covariate vectors in the mean and within subject variance model. Both can include subject and occasion level covariates, and usually Z_{ij} contain a subset of variables in X_{ij}; β and α are the corresponding fixed effect coefficient vectors (α in log scale), which indicate the population average effect of the covariates on the mean and (log of) the within subject variability of the outcome. Similarly, P_{ij} and Q_{ij} are the random effect covariate vectors in the mean and variance model, with $\nu_{1,i}$ and $\nu_{2,i}$ being the corresponding random subject location and scale effects, indicating the effect of subject i on his/her mean and within subject variability of the repeated measurements. Usually P and Q are subsets of X and Z. In the case of a random intercept location scale model, P and Q both consist of a column of 1's. The reason to include both location and scale random effects is to allow for subject heterogeneity in both the mean and within subject variability of the outcome that cannot be fully explained by covariates. This relaxes the homogeneous error variance assumption adopted by most statistical methods.

In the context of the EMA adolescent study example, we will model the mood outcomes (PA and NA) as:

$$Y_{ij} = \beta_0 + \beta_1 smoke_i + \beta_2 gender_i + \beta_3 NMR_i + \beta_4 GPA_i + \beta_5 AloneWS_{ij} + \beta_6 AloneBS_i + \nu_{1,i} + \epsilon_{ij} \tag{2}$$

where smoke (0 = non-smoker, 1 = smoker), gender (0 = Female, 1 = Male), NMR (negative mood regulation) and GPA are all subject level covariates. The occasion level covariate Alone is further decomposed into AloneBS and AloneWS, which are the between-subject and within-subject component of Alone, respectively. The reason for the decomposition is to investigate how the effect of being alone on mood differs when comparing the same subject at different occasions (AloneWS) to different subjects averaged over all occasions (AloneBS). This might help to indicate the appropriate level for clinical interventions to be performed at Piasecki et al. (2014). $\nu_{1,i}$ is the random subject location intercept and reflects the influence of subject i on his/her mood assessments. ϵ_{ij} is the random error and reflects the uncertainty in measuring subject i's mood at occasion j relative to the subject average. The variance of ϵ_{ij} reflects the mood consistency measured for subject i, thus the smaller the variance is, the more stable subject i behaves in terms of his/her mood. To account for the fact that individuals usually exhibit distinct patterns for mood consistency, we additionally model the error variance by

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha_0 + \alpha_1 smoke_i + \alpha_2 gender_i + \alpha_3 NMR_i + \alpha_4 GPA_i + \alpha_5 AloneWS_{ij} + \alpha_6 AloneBS_i + \nu_{2,i} \tag{3}$$

where $\nu_{2,i}$ is the random subject scale intercept and reflects the influence of subject i on the variability of his/her repeated mood assessments. The log function ensures that the error variance is strictly positive. Here we have the same set of covariates in the mean and variance model since the interest is to understand how these covariates affect mood levels as well as the within subject mood variability. The random effects $\{\nu_{1,i}, \nu_{2,i}\}$ are assumed to follow a bivariate normal distribution with mean 0 and some covariance structure. Conditional on $\{\nu_{1,i}, \nu_{2,i}\}$, the mood measurements y_{ij} are *i.i.d.*

3.2 Model for the missing process

We propose a random intercept logistic regression model for the binary missing prompt indicators. Let M_{ij} be the missing indicator for subject i at occasion j, where M_{ij} is 0 if the subject responds to the prompt and 1 if missing. Since all responses are time stamped, we can investigate whether time of day influences these prompt indicators. Here we assume a typical day starts at 3 a.m. in the morning and divide each day into five time bins: 3–9 a.m., 9 a.m.–3 p.m., 3–6 p.m., 6–9 p.m. and 9 p.m.–3 a.m., and use these time bins as covariates in our modeling of the missing process. Empirical analyses indicate that students tend to behave similar during the weekdays, but quite differently on Saturday and Sunday. Thus, to simplify and facilitate model estimation, we combine each of the five time bins from Monday to Friday, resulting in a total of 15 bins: 5 during weekdays, 5 on Saturday and another 5 on Sunday. The proposed random intercept logistic regression model is given by

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau_0 + \sum_{k=2}^{k=15} \tau_k \cdot T^k_{ij} + \lambda_i \tag{4}$$

where $k = 2, \dots, 15$ is the time bin index and T_{ij}^{k} is the indicator of the kth time bin for the prompt individual i received at occasion j. For the purpose of model identifiability, the first bin T_{ij}^{1} is treated as the reference time bin; τ_0 is the fixed intercept, indicating the log odds of missing a response for an individual with $\lambda_i = 0$ during 3–9 p.m. on a weekday when he/she received a prompt. λ_i is subject i's random intercept, indicating the influence of subject i on his/her log odds of missing prompts. Similar to the model in Sect. 3.1, conditional on λ_i, the missing indicators M_{ij} are assumed to be *i.i.d* following a Bernoulli distribution with missing probability $p_{ij} = \frac{\exp(\tau_0 + \sum_{k=2}^{k=15} \tau_k \cdot T_{ij}^{k} + \lambda_i)}{1 + \exp(\tau_0 + \sum_{k=2}^{k=15} \tau_k \cdot T_{ij}^{k} + \lambda_i)}$. Here p_{ij} is modelled by both observed and latent information, with time bins being explicitly measured and the random effect λ_i accounting for all unobserved information at the subject level.

3.3 Parameter sharing and joint model

Up to this point, the outcome and missing process are still separate. It is possible that there exists some common but unobserved information that contributes to both the outcome and missing process. A legitimate example in the above adolescent mood study would be that, an individual's work schedule cannot be measured but is related to both the mood assessments and missing propensity. For example, individuals with tight schedules might have worse and unstable mood, and they might also be less likely to respond to the prompts. In this case, the unmeasurable work schedule might be represented using the random subject effect λ_i extracted from the missing model. This leads to the parameter sharing below:

$$\nu_{1,i} = \gamma \cdot \lambda_i + \eta_{1,i} \tag{5}$$

$$\nu_{2,i} = \delta \cdot \lambda_i + \eta_{2,i} \tag{6}$$

where $\{\nu_{1,i}, \nu_{2,i}\}$ and λ_i are random subject effects in the outcome and missing model, and can both be regarded as traits specific to individual i. A set of linear models are used to link $\{\nu_{1,i}, \nu_{2,i}\}$ with λ_i. In Eq. 5, individuals i's location random intercept $\nu_{1,i}$ can be represented by his/her personal missing trait λ_i and an error term $\eta_{1,i}$ that absorbs the residual variation orthogonal to λ_i. The shared parameter coefficient γ indicates the effect of missingness on the subject's mean outcome. Similarly in Eq. 6, δ represents the effect of missingness on the within subject variability of the outcome. By specifying Eqs. 5 and 6, the primary longitudinal outcome is linked with the missing process through the random subject effects and informative missing can be taken into account. Thus, we call λ the shared random subject effect between the outcome and missingness, and η_1 and η_2 as the residual random subject location and scale effects.

By substituting Eqs. 5 and 6 into Eqs. 2 and 3, we can re-express the shared parameter outcome model for this example below as:

$$Y_{ij} = \beta_0 + \sum_{k=1}^{6} \beta_k \cdot x_{ij}^{k} + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{7}$$

$$\log(\sigma_{\epsilon_{ij}}^{2}) = \alpha_0 + \sum_{k=1}^{6} \alpha_k \cdot x_{ij}^{k} + \delta \cdot \lambda_i + \eta_{2,i}. \tag{8}$$

4 Model estimation

4.1 Maximum likelihood estimation method

The shared parameter location scale model assumes that conditional on the set of random subject effects $\{\lambda_i, \eta_{1,i}, \eta_{2,i}\}$, the primary outcome Y_{ij} and missing indicator M_{ij} are independent. Thus, we can explicitly write out the conditional joint likelihood of $\mathcal{L}(Y, M \mid \lambda, \eta_1, \eta_2)$ as

$$\mathcal{L}(Y, M \mid \lambda, \eta_1, \eta_2) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} f_N(y_{ij} \mid x_{ij}, \lambda_i, \eta_{1,i}, \eta_{2,i}) \cdot f_B(m_{ij} \mid t_{ij}, \lambda_i) \tag{9}$$

where f_N and f_B denote the probability density (mass) function of a Normal and Bernoulli random variable. Specifically, $y_{ij} \mid \lambda_i, \eta_{1,i}, \eta_{2,i} \sim \mathcal{N}\left(x_{ij}^\top \beta + \gamma \cdot \lambda_i + \eta_{1,i}, \exp(x_{ij}^\top \alpha + \delta \cdot \lambda_i + \eta_{2,i})\right)$, and $m_{ij} \mid \lambda_i \sim \mathcal{B}\left(p = \frac{\exp(t_{ij}^\top \tau + \lambda_i)}{1+\exp(t_{ij}^\top \tau + \lambda_i)}\right)$. The marginal joint likelihood $\mathcal{L}(Y, M)$ is then obtained by integrating the conditional joint likelihood $\mathcal{L}(Y, M \mid \lambda, \eta_1, \eta_2)$ over the joint distribution of the random effect vector $\{\lambda, \eta_1, \eta_2\}$.

$$\mathcal{L}(Y, M) = \int \prod_{i=1}^{n} \left\{ \prod_{j=1}^{n_i} f_N(y_{ij} \mid x_{ij}, \lambda_i, \eta_{1,i}, \eta_{2,i}) \cdot f_B(m_{ij} \mid t_{ij}, \lambda_i) \right\} d\mathcal{F}\left(\lambda_i, \eta_{1,i}, \eta_{2,i}\right) \tag{10}$$

where Eq. 10 can be further simplified by model assumptions. As mentioned in Sect. 3.3, λ_i is independent of $\{\eta_{1,i}, \eta_{2,i}\}$, while $\eta_{1,i}$ and $\eta_{2,i}$ are allowed to be correlated. Therefore, $d\mathcal{F}\left(\lambda_i, \eta_{1,i}, \eta_{2,i}\right)$ can be factored into $d\mathcal{F}_N(\lambda_i) d\mathcal{F}_N(\eta_{1,i}, \eta_{2,i})$. Once the marginal joint likelihood is computed, optimization can proceed by obtaining the first and second partial derivatives of the (log) marginal likelihood with respect to all model parameters. However, this procedure involves multi-dimensional numerical integration over the random effect distribution and can be computationally challenging. Also, computing the inverse of the Hessian Matrix can be difficult due to the large number of parameters in the joint model (35 in the above adolescent mood study example).

4.2 Full Bayesian estimation approach

Due to the difficulties in evaluating the marginal joint likelihood as described in Sect. 4.1, we switch to a full Bayesian estimation approach, where parameters and random effects are regarded as random quantities while data are regarded as fixed. To simplify notation, denote $\theta = (\beta, \alpha, \tau, \gamma, \delta)$ as the model parameter vector, $\lambda = \{\lambda_i\}_{i=1}^{n}$ as the random subject effects for the missing process, $\eta = \{\eta_{1,i}, \eta_{2,i}\}_{i=1}^{n}$ as the random subject effect vector in the outcome model and $D = \{Y_i, M_i\}_{i=1}^{n}$ as the data.

Since θ, λ and η are all random, they each follow some prior distribution before we get to observe the data D, and we denote the priors as $\pi(\theta)$, $\pi(\lambda)$, and $\pi(\eta)$ respectively. Since individuals are assumed to be independent, $\pi(\lambda)$ can be written as $\prod_{i=1}^{n} \pi(\lambda_i)$, and similarly for $\pi(\eta)$. Natural choices for $\pi(\lambda_i)$ and $\pi(\eta_i)$ are a univariate standard normal and bivariate standard normal, respectively. For $\pi(\theta)$, one can specify a separate prior for each component in θ provided that a full conditional posterior is obtained for each of them. Given independent priors, one can derive the conditional posterior as

$$P(\theta \mid \lambda_i, \eta_i, D_i) \propto P(D_i \mid \theta, \lambda_i, \eta_i)\pi(\theta) \tag{11}$$

$$P(\lambda_i \mid \theta, \eta_i, D_i) \propto P(D_i \mid \theta, \lambda_i, \eta_i)\pi(\lambda_i) \tag{12}$$

$$P(\eta_i \mid \theta, \lambda_i, D_i) \propto P(D_i \mid \theta, \lambda_i, \eta_i)\pi(\eta_i) \tag{13}$$

$P(D_i \mid \theta, \lambda_i, \eta_i)$ is the conditional joint likelihood given in Eq. 9, and π is the corresponding prior. Once the full conditional posteriors are obtained for θ, λ and η, we can approximate their joint posterior by sampling each variable from its full conditional posterior iteratively using Gibbs sampling (Casella and George 1992). In the case where the conditional posterior is not of a recognized form, one can use the Metropolis–Hastings algorithm, which keeps drawing samples from a proposal distribution and decides whether or not to accept the sample as from the conditional posterior with some acceptance rate (Chib and Greenberg 1995). We devised a MCMC sampling algorithm where component-wise Metropolis–Hastings algorithms are nested within Gibbs sampling. However, a better approach can be taken using Stan, an open source Hamiltonian Monte Carlo sampler, since it can better deal with the trade off between step size and acceptance rate by reducing the correlation between successive samples using a Hamiltonian evolution and target values with a higher acceptance rate than the observed probability distribution (Stan-Development-Team 2014). Both the MCMC derivation and Stan implementation details are provided in the Supplemental Materials. The Hamiltonian Monte Carlo sampling uses improper uniform priors [uniform on $(-\infty, +\infty)$] for regression coefficients, improper bounded uniform priors [uniform on $(0, +\infty)$] for random effect variances, and an LKJ prior for the random effect correlation matrix.

5 Simulation results

To validate the proposed model and estimation procedure, we conducted a series of simulation studies and present the results here. Because of the heavy computation load, we limited the number of simulations to 100 under each scenario. Generally, results became quite consistent after the first 25 simulations and remained consistent till the end.

For each simulation, an intensively measured longitudinal outcome Y was generated via a location scale process for 100 individuals at a total of 30 occasions. Covariates included gender (subject level) and time stamps (occasional level). Once the complete data were generated, observations were set to intermittent missing via a Bernoulli process, where the missing probability was simulated under two scenarios: (1) missing does not depend on potential outcomes (MCAR or MAR), and (2) missing depends on potential outcomes (MNAR). For each scenario, analyses were conducted using two candidate methods: (a) a naive model which assumes MAR and utilizes only the observed outcome, and (b) the proposed model that shares random subject location and scale effects between the outcome and missing process. The detailed model specifications are shown in Eqs. 14 and 15 for the naive model and Eqs. 16–18 for the proposed model, where Y_{ij} denotes the outcome and M_{ij} denotes the missing indicator.

Naive model:

$$Y_{ij} = \beta^{int} + \beta^{gender} \cdot gender_i + \beta^{time} \cdot time_{ij} + \nu_{1,i} + \epsilon_{ij} \tag{14}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha^{int} + \alpha^{gender} \cdot gender_i \tag{15}$$

Proposed model:

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau^{int} + \tau^{gender} \cdot gender_i + \tau^{time} \cdot time_{ij} + \lambda_i \tag{16}$$

$$Y_{ij} = \beta^{int} + \beta^{gender} \cdot gender_i + \beta^{time} \cdot time_{ij} + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{17}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = \alpha^{int} + \alpha^{gender} \cdot gender_i + \delta \cdot \lambda_i + \eta_{2,i} \tag{18}$$

Results are summarized in Table 1. β are the regression coefficients for the outcome mean model, and α are for the within subject variance model (on log scale). $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$ are the variances for the random subject location and scale effects. γ and δ are the coefficients for the shared parameters, indicating the influence of missingness on the mean and within subject variance, as described in Sect. 3.3. The point estimates are obtained as the posterior mean for regression coefficients β, α, γ and δ since their posteriors are approximately

Table 1 Simulation results under two scenarios: MCAR/MAR and MNAR

Parameter	True value	Naive model			Shared parameter model		
		Bias	AIW	COV (%)	Bias	AIW	COV (%)
Scenario MCAR/MAR							
β^{int}	0.30	-2.2×10^{-2}	0.836	98	-2.3×10^{-2}	0.842	95
β^{gender}	0.20	8.5×10^{-4}	1.162	97	1.6×10^{-3}	1.171	96
β^{time}	0.01	-7.1×10^{-5}	0.012	92	-9.2×10^{-5}	0.012	92
α^{int}	0.30	-4.7×10^{-3}	0.181	94	-7.2×10^{-3}	0.192	97
α^{gender}	0.10	1.5×10^{-3}	0.253	96	2.2×10^{-3}	0.265	96
$\sigma^2_{\nu_1}$	2.00	1.26×10^{-1}	1.266	94	1.46×10^{-1}	1.276	93
$\sigma^2_{\nu_2}$	0	–	–	–	7.7×10^{-3}	0.032	–
γ	0	–	–	–	8.9×10^{-4}	0.468	97
δ	0	–	–	–	1.34×10^{-2}	1.886	96
Scenario MNAR							
β^{int}	0.30	4.6×10^{-2}	0.863	96	-8.1×10^{-3}	0.858	97
β^{gender}	0.20	3.0×10^{-2}	1.185	95	2.9×10^{-2}	1.195	93
β^{time}	0.01	5.0×10^{-4}	0.014	96	3.1×10^{-4}	0.009	95
α^{int}	0.30	1.76×10^{-1}	0.185	29	1.2×10^{-3}	0.622	95
α^{gender}	0.10	2.9×10^{-2}	0.260	43	8.3×10^{-3}	0.880	95
$\sigma^2_{\nu_1}$	2.00	1.34×10^{-1}	1.326	93	1.27×10^{-1}	1.378	99
$\sigma^2_{\nu_2}$	1.00	–	–	–	8.1×10^{-2}	0.739	95
γ	− 1.00	–	–	–	-5.8×10^{-2}	0.656	97
δ	1.00	–	–	–	5.6×10^{-2}	0.671	93

symmetric, and as mode for random effect variances $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$ since their posteriors are relatively skewed. Bias is computed for each parameter as the average point deviation from the true value: $Bias = \sum_{k=1}^{100}(\hat{\theta}_k - \theta)/100$, where $\hat{\theta}_k$ denotes the posterior mean for $(\beta, \alpha, \gamma, \delta)$ and mode for $\left(\sigma^2_{\nu_1}, \sigma^2_{\nu_2}\right)$ from the k_{th} simulation. AIW (average interval width) is computed as the average range between the 97.5 and 2.5% quantile of the posterior: $AIW = \sum_{k=1}^{100}(\theta^U_k - \theta^L_k)/100$, where θ^U_k and θ^L_k are the 97.5 and 2.5% quantile of the posterior distribution from the kth simulation. For each parameter, we also calculate the number of times out of 100 that the 95% credible interval contains its true value, thus providing the coverage rate as $COV = \sum_{k=1}^{100} \mathbb{1}\{\theta^L_k \leqslant \theta \leqslant \theta^U_k\}/100$.

Under MAR, the naive model and proposed model are expected to perform similar as the missing process is independent of the potential outcome. This is also confirmed in Table 1 from the small bias, reasonable AIW and correct COV for both models. Under MNAR, however, the two models diverge in terms of the inference for α. The shared parameter model has both smaller bias and much better coverage rate for α^{intcp} and α^{gender} compared to the naive model. Furthermore, the small bias and correct coverage of $\sigma^2_{\nu_2}$, γ and δ also provide evidence of the validity of the proposed model. This clearly shows that the naive analyses, which ignore the association between the primary outcome and the missing process, can lead to invalid inference, particularly for the variance model parameters. The mean model parameters and random location effect variance seem less likely to be affected as all bias is absorbed into the error variance components. Overall, the proposed shared parameter model achieves good estimation precision, correct interval length and asymptotic coverage rate, yet provides insightful information about the missing mechanisms.

In addition, we imputed the missing values under the two scenarios by both candidate models and compared the imputated values with the true (missing) outcomes. Under MAR, both the naive and shared parameter models achieve small imputation bias (0.0005 vs. 0.0007) and correct coverage rate (94.8 vs. 94.9%). Under MNAR, however, the naive model produces greater imputation bias (0.162 vs. 0.005) and an inadequate coverage rate (86.5 vs. 94.9%) as compared to the shared parameter model.

6 Application to adolescent mood study example

In this section, we revisit the example introduced in Sect. 2. An important aim of the study is to identify factors that can potentially influence the mean and within subject variance of individuals' positive/negative mood. Informative missing is likely to occur since mood can only be assessed if individuals respond to the prompt, and whether or not they decide to respond may be affected by their mood at the time of the prompt. For every prompt individual i received, we record the missing indicator vector M_i (1 if missing, 0 if respond), mood assessment vector Y_i (PA or NA) if not missing and the time window T_i ($T^k_{ij} = 1$ if the prompt occurred in the kth window for $k = 1, \ldots, 15$).

Three candidate models are applied to the example EMA data: (1) a random intercept model with heterogeneous variance (HV) that provides covariance adjustment for the correlation among repeated measurements and allows covariates to affect the within subject variance, (2) the shared location model that not only provides inference for the outcome mean and within subject variance, but also allows missingness and the outcome mean to be correlated through shared random subject effect, and (3) the proposed shared location scale model that associates each individual's missingness with both the mean and within subject variance of the primary outcome. The detailed model specifications are shown below, where $X_{ij} = (1, smk_i, gender_i, NMR_i, GPA_i, AloneBS_i, AloneWS_{ij})$ is the covariate vector for the mean and within subject variance model, $\beta = (\beta^{int}, \beta^{smk}, \beta^{gender}, \beta^{NMR}, \beta^{GPA}, \beta^{AloneBS}, \beta^{AloneWS})$ is the regression coefficient vector for the mean model, and $\alpha = (\alpha^{int}, \alpha^{smk}, \alpha^{gender}, \alpha^{NMR}, \alpha^{GPA}, \alpha^{AloneBS}, \alpha^{AloneWS})$ is the regression coefficient vector for the within subject variance model.

Random intercept HV model:

$$Y_{ij} = X_{ij}^{\top}\beta + \nu_{1,i} + \epsilon_{ij} \tag{19}$$

$$\log\left(\sigma^2_{\epsilon_{ij}}\right) = X_{ij}^{\top}\alpha + \nu_{2,i} \tag{20}$$

Shared location model:

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau_0 + \sum_{k=2}^{k=15}\tau_k \cdot T_{ij}^k + \lambda_i \tag{21}$$

$$Y_{ij} = X_{ij}^{\top}\beta + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{22}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = X_{ij}^{\top}\alpha + \nu_{2,i} \tag{23}$$

Shared location scale model:

$$\log\left(\frac{Pr(M_{ij}=1)}{1-Pr(M_{ij}=1)}\right) = \tau_0 + \sum_{k=2}^{k=15}\tau_k \cdot T_{ij}^k + \lambda_i \tag{24}$$

$$Y_{ij} = X_{ij}^{\top}\beta + \gamma \cdot \lambda_i + \eta_{1,i} + \epsilon_{ij} \tag{25}$$

$$\log(\sigma^2_{\epsilon_{ij}}) = X_{ij}^{\top}\alpha + \delta \cdot \lambda_i + \eta_{2,i} \tag{26}$$

To better compare the parameter estimates and credible intervals, we performed the Bayesian approach in Sect. 4.2 for all three candidate models. Results are summarized in Table 2 for positive affect and Table 3 for negative affect. Again, the point estimates are obtained as the posterior mean for regression coefficients β, α, γ and δ, and as mode for random effect variances $\sigma^2_{\nu_1}$ and $\sigma^2_{\nu_2}$. The 95% credible intervals (CI) are obtained as the 2.5 and 97.5% posterior quantiles for all parameters. The model selection criteria $elpd_{LOO}$, proposed by Vehtari et al. (2017), estimates the pointwise leave one out (LOO) prediction accuracy from a fitted Bayesian model by evaluating the log likelihood over the posterior samples. It is preferred over the deviance information criterion (DIC) since

Table 2 Comparison of parameter estimates and credible intervals between the random intercept model with heterogeneous variance, and shared parameter location model, and shared parameter location scale model: positive affect

Parameter	Random intercept HV model		Shared location model		Shared location scale model	
	Estimate	CI	Estimate	CI	Estimate	CI
β^{int}	6.084	(5.463, 6.742)	6.152	(5.532, 6.826)	6.115	(5.422, 6.789)
β^{smk}	− 0.135	(− 0.330, 0.071)	− 0.143	(− 0.358, 0.055)	− 0.146	(− 0.363, 0.076)
β^{gender}	0.204	(− 0.011, 0.440)	0.210	(− 1.528, 0.412)	0.212	(− 0.019, 0.428)
β^{NMR}	0.640	(0.491, 0.803)	0.647	(0.466, 0.811)	0.658	(0.500, 0.821)
β^{GPA}	− 0.148	(− 0.302, − 0.011)	− 0.170	(− 0.315, − 0.022)	− 0.172	(− 0.319, − 0.020)
$\beta^{AloneBS}$	− 1.070	(− 1.597, − 0.536)	− 1.059	(− 1.64, − 0.450)	− 0.999	(− 1.529, − 0.450)
$\beta^{AloneWS}$	− 0.328	(− 0.387, − 0.270)	− 0.329	(− 0.391, − 0.268)	− 0.256	(− 0.307, − 0.205)
α^{int}	1.518	(1.367, 1.670)	1.512	(1.371, 1.661)	1.286	(0.868, 1.713)
α^{smk}	0.006	(− 0.043, 0.054)	0.006	(− 0.043, 0.059)	0.033	(− 0.109, 0.170)
α^{gender}	− 0.220	(− 0.274, − 0.166)	− 0.220	(− 0.271, − 0.167)	− 0.270	(− 0.421, − 0.114)
α^{NMR}	− 0.176	(− 0.214, − 0.136)	− 0.175	(− 0.211, − 0.140)	− 0.160	(− 0.263, − 0.050)
α^{GPA}	− 0.082	(− 0.114, − 0.049)	− 0.081	(− 0.114, − 0.048)	− 0.080	(− 0.179, 0.027)
$\alpha^{AloneBS}$	0.321	(0.189, 0.451)	0.320	(0.191, 0.447)	0.330	(− 0.021, 0.676)
$\alpha^{AloneWS}$	0.049	(− 0.010, 0.105)	0.051	(− 0.007, 0.109)	0.107	(0.047, 0.168)
$\sigma^2_{v_1}$	1.025	(0.998, 1.333)	1.093	(1.013, 1.317)	1.215	(1.013, 1.326)
$\sigma^2_{v_2}$	–	–	–	–	0.530	(0.415, 0.572)
ρ_{v_1,v_2}	–	–	–	–	− 0.294	(− 0.358, − 0.275)
γ	–	–	− 0.109	(− 0.242, 0.006)	− 0.134	(− 0.254, − 0.006)
δ	–	–	–	–	0.061	(− 0.0030, 0.150)
$elpd_{LOO}$	− 23,866	–	− 23,863	–	− 22,815	–

it accounts for the entire posterior distribution, works for singular models and is invariant to parametrization. Higher $elpd_{LOO}$ indicates better model fit adjusting for the model complexity.

For covariate effects on the mean of the mood outcome, the shared location model and shared location scale model give relatively similar estimates that are different from the random intercept HV model, except for *AloneWS*. Specifically for positive affect, higher negative mood regulation is significantly associated with higher PA, while higher GPA and being alone (both at subject and occasion levels) are associated with lower PA. Although not statistically significant, the trend indicates that smokers tend to have lower PA compared to non-smokers and males tend to have better PA than females, adjusting for all other covariates. For all three candidate models, the magnitude of the effect of *AloneBS* on PA is estimated to be three times as big as *AloneWS*, suggesting that the between- and within-subject effects are not equal, though of the same sign. Subjects who are alone more often report lower average PA (between-subject effect), and when subjects are alone they also report lower PA (within-subject effect). For negative affect, in Table 3, the mean effect estimates are of opposite sign and lead to similar conclusions, as higher NA indicates worse mood.

For covariate effects on the within subject variance, PA and NA models provide relatively similar coefficient estimates since both the within subject variance of PA and NA

Table 3 Comparison of parameter estimates and credible intervals between the random intercept model with heterogeneous variance, and shared parameter location model, and shared parameter location scale model: negative affect

Parameter	Random intercept HV model		Shared location model		Shared location scale model	
	Estimate	CI	Estimate	CI	Estimate	CI
β^{int}	4.550	(3.756, 5.293)	4.591	(3.818, 5.349)	4.516	(3.773, 5.307)
β^{smk}	0.374	(0.086, 0.639)	0.366	(0.082, 0.626)	0.359	(0.100, 0.616)
β^{gender}	− 0.354	(− 0.612, − 0.092)	− 0.377	(− 0.652, − 0.105)	− 0.384	(− 0.640, − 0.123)
β^{NMR}	− 0.825	(− 1.020, − 0.628)	− 0.863	(− 1.036, − 0.685)	− 0.853	(− 1.052, − 0.663)
β^{GPA}	0.241	(0.071, 0.415)	0.262	(0.087, 0.424)	0.275	(0.101, 0.440)
$\beta^{AloneBS}$	0.172	(− 0.516, 0.901)	0.175	(− 0.509, 0.850)	0.106	(− 0.569, 0.770)
$\beta^{AloneWS}$	0.199	(0.132, 0.268)	0.201	(0.136, 0.263)	0.090	(0.045, 0.135)
α^{int}	1.803	(1.654, 1.945)	1.803	(1.663, 1.955)	1.689	(1.105, 2.215)
α^{smk}	0.132	(0.082, 0.184)	0.132	(0.078, 0.185)	0.204	(0.020, 0.374)
α^{gender}	− 0.236	(− 0.286, − 0.185)	− 0.236	(− 0.289, − 0.185)	− 0.370	(− 0.570, − 0.169)
α^{NMR}	− 0.183	(− 0.221, − 0.146)	− 0.182	(− 0.221, − 0.144)	− 0.299	(− 0.436, − 0.150)
α^{GPA}	− 0.078	(− 0.111, − 0.045)	− 0.078	(− 0.111, − 0.045)	− 0.047	(− 0.160, 0.069)
$\alpha^{AloneBS}$	− 0.030	(− 0.164, 0.103)	− 0.030	(− 0.170, 0.113)	0.062	(− 0.446, 0.546)
$\alpha^{AloneWS}$	− 0.001	(− 0.061, 0.057)	− 0.001	(− 0.060, 0.056)	0.023	(− 0.041, 0.088)
$\sigma^2_{v_1}$	1.625	(1.544, 2.069)	1.643	(1.551, 2.061)	1.714	(1.527, 2.033)
$\sigma^2_{v_2}$	–	–	–	–	0.758	(0.745, 1.015)
ρ_{v_1,v_2}	–	–	–	–	0.413	(0.375, 0.443)
γ	–	–	0.176	(0.017, 0.343)	0.175	(0.017, 0.338)
δ	–	–	–	–	0.086	(− 0.031, 0.207)
$elpd_{LOO}$	− 25,339	–	− 25,339	–	− 23,840	–

reflect individuals' mood consistency/inconsistency. The random intercept HV model and shared location model give similar effect estimates as well as narrower credible intervals that are different from the proposed shared location scale model. This is as expected since neither of the two former models include scale random effects for the within subject variance and there is no parameter sharing between the outcome variation and the missing process. Therefore we will refer to the proposed shared location scale model for coefficient interpretation. Specifically for positive affect, higher negative mood regulation is significantly associated with more stable mood, which is in agreement with the theory that higher negative mood regulation indicates better mood control. Males tend to have more stable PA compared to females, and subjects with lower GPA tend to have more stable PA. Although not statistically significant, the trend indicates that smokers tend to have more erratic PA compared to non-smokers adjusting for all other covariates. The three candidate models disagree on the effect of *Alone*, where the random intercept HV model and shared location model indicate that the between subject component *AloneBS* contributes significantly, while the shared location scale model indicates increased variation for the within subject component *AloneWS*. Thus, comparing responses from the same subject at different occasions, the subject's PA is more variable when he/she is alone compared to when he/she is with others.

As an extra benefit from the shared parameter model, there seems to be a negative association between the missingness and PA mean (positive for NA), as indicated by the estimate of γ. This is in agreement with our hypothesis that a lower response rate is related to worse mood. Although not significant at the 5% level, the positive estimate of δ on PA and NA indicates that a lower response rate is also associated with unstable mood. For PA, there is a negative association between the random location and scale random effect possibly due to a ceiling effect (i.e., subjects with high PA means tend to have lower scale due to the ceiling of measurement). This association is positive for NA indicating that subjects with lower means have lower variability, possibly due to a floor effect of measurement. The shared location scale model achieves great improvement in terms of the model fit over the other two models adjusting for complexity, as is shown by $elpd_{LOO}$, and is thus preferred.

Table 4 summarizes the estimated effects in the missingness model for all 15 time windows (5 time bins for weekdays as well as for Saturday and Sunday) for both PA and NA. Generally, one would expect to obtain very similar estimates for both PA and NA since, when prompted, students are most likely to answer them both or neither. During weekdays, the most prompts were answered late in the day (from 6 to 9 p.m.) and students become least attentive early in the day (from 3 to 9 a.m.). For the weekend days, students answer most prompts from 9 a.m.to 3 p.m. on Saturday and from 6 to 9 p.m. on Sunday. Students generally behave less responsive on Saturday, and a bit more responsive on Sunday (except from 3 to 9 a.m.) during the same time frame as compared with weekdays.

In addition to the above analyses, we also performed cross validation using the adolescent mood data. Specifically, observations were set to missing according to various missing scenarios, then missing observations were imputed by candidate models trained from the available data (as well as missing patterns), and finally imputed values were compared with the true values based on posterior prediction accuracy as measured by the model likelihood

Table 4 Parameter estimates and credible intervals for the missing process model: positive affect and negative affect

Parameter	Positive affect		Negative affect	
	Estimate	CI	Estimate	CI
τ_0	− 1.281	(− 1.453, − 1.103)	− 1.285	(− 1.477, − 1.085)
τ: Weekday 9 a.m.–3 p.m.	− 0.245	(− 0.428, − 0.064)	− 0.240	(− 0.424, − 0.053)
τ: Weekday 3–6 p.m.	− 0.112	(− 0.310, 0.078)	− 0.104	(− 0.300, 0.089)
τ: Weekday 6–9 p.m.	− 0.270	(− 0.465, − 0.078)	− 0.264	(− 0.460, − 0.059)
τ: Weekday 9 p.m.–3 a.m.	− 0.144	(− 0.360, 0.062)	− 0.138	(− 0.354, 0.080)
τ: Saturday 3–9 a.m.	0.847	(0.290, 1.412)	0.845	(0.277, 1.421)
τ: Saturday 9 a.m.–3 p.m.	− 0.100	(− 0.361, 0.159)	− 0.096	(− 0.359, 0.157)
τ: Saturday 3–6 p.m.	0.117	(− 0.171, 0.404)	0.126	(− 0.151, 0.397)
τ: Saturday 6–9 p.m.	0.032	(− 0.279, 0.326)	0.041	(− 0.240, 0.324)
τ: Saturday 9 p.m.–3 a.m.	0.251	(− 0.057, 0.543)	0.256	(− 0.043, 0.538)
τ: Sunday 3–9 a.m.	1.725	(1.043, 2.463)	1.739	(1.005, 2.521)
τ: Sunday 9 a.m.–3 p.m.	− 0.234	(− 0.495, 0.032)	− 0.229	(− 0.481, 0.042)
τ: Sunday 3–6 p.m.	− 0.341	(− 0.673, − 0.033)	− 0.333	(− 0.667, − 0.025)
τ: Sunday 6–9 p.m.	− 0.465	(− 0.789, − 0.158)	− 0.459	(− 0.777, − 0.157)
τ: Sunday 9 p.m.–3 a.m.	− 0.313	(− 0.647, 0.013)	− 0.303	(− 0.640, 0.032)
σ^2_λ	0.767	(0.666, 0.959)	0.717	(0.636, 0.945)

($elpd_{CV}$). As before, candidate models include (1) a random intercept HV model that assumes MAR, (2) a shared parameter location model that only links the outcome mean with the missing process, and (3) a shared parameter location scale model that links both PA mean and variance with the missing process. In Table 5, γ and δ are the coefficients in the parameter sharing model and denote the effect of missingness on PA mean and variance, respectively. In the first scenario, where both γ and δ were set to 0 and missingness does not depend on potential outcomes (MAR or MCAR), the first two candidate models should achieve similar posterior prediction accuracy (since sharing location or scale does not contribute to model fit). The shared location scale model achieves slightly higher prediction accuracy due to the inclusion of random scale effects. In the second scenario, where δ is set to 0 and missingness only depends on the outcome mean, the shared location model should perform better than the random intercept HV model since it corrects for MNAR by linking information between the PA mean with the missingness. Again, the shared location scale model performs slightly better than the shared location model due to the inclusion of random scale effects. In the third scenario, where δ is allowed to vary and missingness depends on both the PA mean and variance, only the shared location scale model is able to capture the correct missing pattern. The further away δ moves from 0, the greater benefit the shared location scale model achieves over the random intercept HV and shared location models.

The details of implementing the three candidate models in Stan and R (using the package "RStan") are provided in the supplemental materials. One can incorporate the Stan code provided in Supplemental Materials 1.2 and call Stan from R using the sample code in 1.3.

7 Discussion

In this paper, we have developed a general shared parameter framework for normally distributed and intensively measured longitudinal outcomes with informative missingness. We exploit a mixed effect location scale model proposed by Hedeker et al. (2008) for the primary outcome, and a random intercept logistic model for the intermittent missing process. The two models are connected by the random subject location and scale effects, which are assumed to capture the common but unmeasurable information at the subject level that contributes to both the primary outcome and missing process.

Table 5 Comparison of cross validated prediction accuracy (measured by $elpd_{CV}$) under various missing scenarios

Missing scenarios γ	δ	Random intercept HV model	Shared location model	Shared location scale model
0	0	− 6183.44	− 6183.82	− 5953.89
− 1	0	− 7597.93	− 7507.31	− 7316.31
− 2	0	− 9697.15	− 8879.17	− 8606.67
− 4	0	− 12,398.75	− 10,617.76	− 10,468.86
− 1	1	− 9133.91	− 8950.74	− 8500.87
− 1	2	− 10,396.42	− 10,165.195	− 9595.23
− 1	4	− 12,899.99	− 12,299.93	− 11,621.36

Due to the computational burden in multi-dimensional numerical integration, we propose a Bayesian MCMC estimation approach where the joint posterior distribution can be approximated by samples drawn from a conventional Metropolis–Hastings–Gibbs algorithm. Various improvements such as Hamiltonian Monte Carlo can be used to balance the trade off between step size and acceptance rate, and ultimately make the chains converge faster.

For simplicity, we have assumed a linear relationship between the outcome location/scale random effects and the missingness random effect with some measurement error. However, this relationship could be made more general. Since the scale random effects are on the log metric, it could be the case that they are related with the missingness random intercept on the original scale, i.e, $\exp(v_{2,i}) = \delta \cdot \lambda_i + \eta_{2,i}$. However, this might introduce additional difficulties in model estimation. Also, we have assumed a simple random intercept logistic model for the missing process, which might not be a good fit in some situations. As an extension, one might try a random slope model for the missingness where each individual has a bivariate random effect vector $\{intercept_i, slope_i\} = \{\lambda_{1,i}, \lambda_{2,i}\}$, representing the influence of individual i on his/her baseline missing propensity as well as the rate of change. In this case, a more sophisticated sharing mechanism would be required to connect $\{v_{1,i}, v_{2,i}\}$ with $\{\lambda_{1,i}, \lambda_{2,i}\}$, which might not be trivial.

The simulation results under the MAR assumption show almost no difference between the naive and proposed model. One might prefer the naive method given it is more parsimonious and easier to estimate. However, one cannot know whether data are missing at random or not in practice when the true underlying mechanism is unknown. As is shown in the adolescent mood study example, the naive model underperforms in terms of the out of sample prediction accuracy relative to the proposed approach. Also, the estimated γ and δ coefficients both suggest evidence towards non-random missingness. Therefore, the shared parameter model would seem to be a safer choice and might be preferred in practice.

In psychological and behavioral sciences, research interests are usually focused on both the actual magnitude of the outcome as well as the within subject variability (Steven et al. 2014). For example, clinician or patient rated average levels of depression, as well as their variability, are both critical aspects in characterizing depressed and bipolar patients. Identifying factors that can potentially influence the mean and within subject variation of the psychological outcomes can jointly provide deep insights for clinical intervention. Further, as both the outcome mean and variability can be correlated with individuals' propensity of responding, it is essential to share both the location and scale random effects between the outcome and missingness models. For binary and Poisson outcomes, where the variability is a 1–1 function of the mean, one can replace the bivariate $\{v_{1,i}, v_{2,i}\}$ vector in the sharing model with the univariate $v_{1,i}$ scaler. An alternative model specification when there is evidence for over-dispersion is to model the over-dispersion parameter with a set of scale random effects, and adopt a similar sharing mechanism as in the proposed model (Hedeker et al. 2009).

Funding The project described was supported by National Heart Lung and Blood Institute (Grant No. R01 HL121330) and National Cancer Institute (Grant No. 5P01CA09862).

Compliance with ethical standards

Conflict of interest The content is solely the responsibility of the authors and does not necessarily represent the official views of National Heart Lung and Blood Institute or National Cancer Institute.

Ethical approval This article does not contain any studies with animals performed by any of the authors. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Bradley, P.C., Siddhartha, C.: Bayesian model choice via Markov Chain Monte Carlo methods. J. R. Stat. Soc. Ser. B (Methodol.) **57**(3), 473–484 (1995)

Casella, G., George, E.: Explaining the Gibbs sampler. Am. Stat. **46**(3), 167–174 (1992)

Chib, S., Greenberg, E.: Understanding the Metropolis–Hastings algorithm. Am. Stat. Assoc. **49**(4), 327–335 (1995)

Cursio, J., Mermelstein, R., Hedeker, D.: Latent trait shared-parameter mixed-models for missing ecological momentary assessment data. Manuscript submitted for publication (2018)

Diggle, P., Kenward, M.: Informative drop-out in longitudinal data analysis. J. R. Stat. Soc. Ser. C **43**, 49–93 (1994)

Follmann, D., Wu, M.: An approximate generalized linear model with random effects for informative missing data. Biometrics **51**, 151–168 (1995)

Heckman, J.: Sample selection bias as a specification error. Econometrica **47**, 153–161 (1979)

Hedeker, D., Mermelstein, R., Demirtas, H.: An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics **64**, 627–634 (2008)

Hedeker, D., Demirtas, H., Mermelstein, R.: A mixed ordinal location scale model for analysis of ecological momentary assessment (EMA) data. Stat. Interface **2**, 391–402 (2009)

Laird, N.: Missing data in longitudinal studies. Stat. Med. **7**, 305–315 (1988)

Little, J.: Pattern-mixture models for multivariate incomplete data. J. Am. Stat. Assoc. **88**, 125–134 (1993)

Martin, M., Hofer, S.: Intraindividual variability, change, and aging: conceptual and analytical issues. Gerontology **50**, 7–11 (2004)

Nesselroade, J.: Intraindividual variability and short-term change. Gerontology **50**, 44–47 (2004)

Piasecki, T., Trela, C., Hedeker, D., Mermelstein, R.: Smoking antecedents: separating between- and within-person effects of tobacco dependence in a multiwave ecological momentary assessment investigation of adolescent smoking. Nicotine Tob. Res. **16**(Suppl 2), 119–126 (2014)

Pulkstenis, E., TenHave, T., Richard, J.: Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. J. Am. Stat. Assoc. **93**, 438–450 (1998)

Shiffman, S., Stone, A., Hufford, M.: Ecological momentary assessment. Annu. Rev. Clin. Psychol. **4**, 1–32 (2008)

Sokolovsky, A., Mermelstein, R., Hedeker, D.: Factors predicting compliance to ecological momentary assessment among adolescent smokers. Nicotine Tob. Res. **16**, 351–358 (2014)

Stan-Development-Team: Stan Modeling Language: User's Guide and Reference Manual, Technical report. Bell Labs (2014). http://mc-stan.org

Steven, M., Matthew, R., Paul, E., Elizabeth, K., Daniel, F.: Mood instability and psychosis: analyses of British national survey data. Schizophr. Bull. **40**(2), 269–277 (2014)

Vehtari, A., Gelman, A., Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. **27**(5), 1413–1432 (2017)

Vonesh, E., Greene, T., Schluchter, M.: Shared parameter models for the joint analysis of longitudinal data and event times. Stat. Med. **15**, 143–163 (2006)

Wu, M., Carroll, R.: Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics **44**, 175–188 (1988)