Research Article

Statistics in Medicine

Received 22 April 2014, Accepted 3 October 2014 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6345

Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: an application to ecological momentary assessment data

Kush Kapur,[a*†] Xue Li,[b] Emily A. Blood[c] and Donald Hedeker[d]

In the statistical literature, the methods to understand the relationship of explanatory variables on each individual outcome variable are well developed and widely applied. However, in most health-related studies given the technological advancement and sophisticated methods of obtaining and storing data, a need to perform joint analysis of multivariate outcomes while explaining the impact of predictors simultaneously and accounting for all the correlations is in high demand. In this manuscript, we propose a generalized approach within a Bayesian framework that models the changes in the variation in terms of explanatory variables and captures the correlations between the multivariate continuous outcomes by the inclusion of random effects at both the location and scale levels. We describe the use of a spherical transformation for the correlations between the random location and scale effects in order to apply separation strategy for prior elicitation while ensuring positive semi-definiteness of the covariance matrix. We present the details of our approach using an example from an ecological momentary assessment study on adolescents. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: multivariate outcomes; joint models; mixed-effects location scale model; spherical transformation; Gibbs sampling

1. Introduction

Mixed-effects models are commonly employed for testing the efficacy of a treatment or to study the growth patterns of outcomes obtained repeatedly over time. In these models, a basic feature is the inclusion of random subject effects into the regression models to allow for the influence of subjects on their repeated observations. These random effects not only provide information on subject-specific responses over time but also capture the correlations between the repeated outcomes parsimoniously. The advantages and many applications of mixed-effects models are presented by several authors in the literature; for example, see [1–7] and the references therein.

Longitudinal studies often collect data on multiple outcomes simultaneously over the course of the repeated assessments. Although inferences can be derived by modeling the various outcomes separately, some of the special features of multiple outcomes can only be understood by using an analysis that synthesizes all the outcomes into one model. To do this, a class of models has been developed to jointly model the behavior of a sequence of longitudinal measurements or survival outcomes. Various joint models extend the modeling of association between and within outcomes using either a common set of latent

[a]*Clinical Research Center and Department of Neurology, Boston Children's Hospital, Harvard Medical School, 21 Autumn St., Boston, MA 02215, U.S.A.*
[b]*Cooperative Studies Program Coordinating Center, Hines VA Hospital, 5000 South 5th Avenue, Building 1, Hines, IL 60141, U.S.A.*
[c]*Division of Biostatistics, Community and Family Medicine, Geisel School of Medicine at Dartmouth, 1 Rope Ferry Rd, Hanover, NH 03755, U.S.A.*
[d]*Department of Health Studies, The University of Chicago Biological Sciences, 5841 South Maryland Avenue, Room W254, Chicago, IL 60637, U.S.A.*
[*]*Correspondence to: Kush Kapur, Clinical Research Center and Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, U.S.A.*
[†]*E-mail: kush.kapur@childrens.harvard.edu*

variables or by the generalized estimating equation approach to avoid specification of the full joint distribution. These models also allow evaluation of the effect of covariates on different types of responses while capturing the relationship between the responses of the clustered data. Verbeke and Davidian [8] provide a comprehensive review of joint modeling approaches for longitudinal data depending on the use of latent variables, type of outcomes, and structure of the data. For the joint analysis of longitudinal and survival outcomes, we refer to [8–11].

For multilevel continuous outcomes, a common assumption is that the error variance, which captures within-subjects (WS) variance, and the random-effects variance, which represents between-subjects (BS) variance, are homogeneous across subjects. However, there are several instances when the error variance is systematically related to explanatory variables or other characteristics of the subject. The idea of modeling the error variance in terms of explanatory variables has been presented in the statistical literature by Carroll and Ruppert [12]. Furthermore, examples of parametric models to capture heteroscedasticity can also be found in econometric work (see, for example, [13] and [14]). Cleveland *et al.* [15] proposed a general class of mixed-effects models that include both random location and random scale effects. In these models, the scale (error variance) of the measurements was allowed to change randomly with each unit of measurement in addition to the location (mean response) of the measurements.

Recently, Hedeker *et al.* [16] presented an approach to model the WS and BS variances as a function of explanatory variables for two-level continuous data. Here, two-level refers to the hierarchical levels of observations (level-1) that are nested within subjects (level-2). Li and Hedeker [17] extended this approach to three-level data (e.g., observations within days within subjects) while allowing covariates to influence the variances of each level using a log-linear representation. Oksana *et al.* [18] proposed a two-level bivariate mixed-effects location scale model that jointly models two outcomes and allows modeling of the BS and WS variances, as well as the BS and WS covariances of the outcomes. An essential contribution of these approaches is the inclusion of a random subject effect in the error variance model in order to further capture the heterogeneity in the errors at measurement level as well as allow for correlation between the random location and scale effects. These authors also provide complete details on how to obtain maximum-likelihood estimates of the parameters and their standard errors. On the basis of these ideas, Blood *et al.* [19] presented a method for estimating and testing the effect of explanatory variables on the variance of momentary behaviors in three-level dyadic ecological momentary assessments (EMA) data. Lee *et al.* [20] proposed a Kronecker product covariance structure for bivariate longitudinal ordinal data to capture the correlation between processes at a given time and the correlation within a process over time. For the serial correlation, they reparameterized the correlation matrix in terms of partial autocorrelations to obtain a general class of models than standard autoregressive correlation models. Recently, Zhang *et al.* [21] proposed a joint mean-variance correlation modeling approach for longitudinal studies by applying hyper-spherical coordinates that guaranteed positive definiteness of the correlation matrix and developed a regression approach to model the correlation matrix of the longitudinal measurements by exploiting this parametrization.

The aforementioned approaches of modeling heteroscedasticity at different levels have also been discussed within a Bayesian framework. Boscardin and Gelman [22] have presented a fully Bayesian approach, which automatically averages over the uncertainty in the model parameters. Browne [23] and Browne *et al.* [24] discussed models where the level-1 variance depends on predictor variables and contrast two maximum-likelihood methods using iterative generalized least squares with two Markov chain Monte Carlo (MCMC) methods based on adaptive hybrid versions of the Metropolis–Hastings sampling. Browne *et al.* [25] further described a methodology for splitting variance that is attributable to higher levels in multilevel binomial logistic models using a variance partition coefficient. Pourahmadi and Daniels [26] introduced a new class of models, termed dynamic conditionally linear mixed models, by decomposing the WS covariance matrix using Cholesky decomposition and allowing for the past responses to enter as covariates. Hoff and Niu [27] presented an approach to parametrize the covariance matrix of a multivariate response vector as a parsimonious quadratic function of explanatory variables and described parameter estimation using EM-algorithm and MCMC approximation via Gibbs sampling. Rast *et al.* [28] used the software BUGS/JAGS to model individual differences in level and change using the mixed-effects location scale model proposed by Hedeker *et al.* [16]. Congdon [29, Chapter 11] discussed the importance of modeling dependence of the random-effects variance and error variance on the explanatory variables in order to have a robust inference about the model parameters and potentially obtain a better fit.

It is clear that modeling heteroscedasticity leads to efficient estimation of the regression parameters and allows predictive inference to be more precise for some units and less for others in comparison with the models with homogeneous variance assumptions that yield the same precision for all the observations. In

this manuscript, we have generalized the approach presented in Hedeker *et al.* [16] and Li and Hedeker [17] to multivariate continuous outcomes by incorporating the random effects at both the location and scale levels within a Bayesian framework. In addition, we allow for complete correlation between the various random location and scale effects at different levels using a spherical transformation. We present the details of our approach following an example from an EMA study of mood and smoking in adolescents.

Our article is organized as follows: in Section 2, we present our motivating example. In Section 3, we discuss a Bayesian approach to model heterogeneity in the variance and covariance of multivariate continuous outcomes at the location and scale levels. We then present the model likelihood and the priors in Section 4. In Section 5, we present the results based on simulated data in order to understand the properties of the parameter estimates. In Section 6, we discuss the results of the analysis in our motivational example. Section 7 concludes with a discussion. Appendix A contains the conditionals required for the Gibbs sampling. We have also provided a WinBUGS 1.4 code for our proposed model as a part of the Supporting Information of this article.

2. Example

Intensive data collection procedures, such as EMA [30,31], experience sampling [32–34], and diary methods [35], have been developed to record the momentary events and experiences of subjects in daily life [35]. Using these approaches, investigators can obtain in-the-moment information about mood, behavior, and other contextual factors as these measures can fluctuate significantly throughout a day. These methods allow researchers to answer behavioral research questions of an individual or the population of interest and help in reducing the recall bias. In EMA, a study participant is usually provided with a hand-held computer that signals randomly throughout the day, prompting the participants to answer questions about their mood, behavior, and/or context.

The data used to illustrate our proposed mixed-effects location scale model for multivariate outcomes come from a longitudinal study of the natural history of smoking among adolescents. This study has been described elsewhere [36]. Briefly, the study used a multi-method approach to assess adolescents at multiple timepoints (baseline, 6, 9, 15, 24, and 33 months) through various data collection methods including self-report questionnaires, in-person interviews, family observations, psycho-physiological assessments, and week-long EMA sampling via hand-held palmtop computers. In this article, we will focus on the data from the baseline EMA collection only. The participants included in the study were either in 9th or 10th grade at baseline (50.7% 9th graders) who had smoked at least once during the past 12 months but had not yet smoked five or more cigarettes in a day. A total of 461 (55.1% female; 56.8% White; mean age = 15.7 years) participants completed the baseline assessment; 56.1% of them had smoked at least one cigarette in the past month, and 31% smoked weekly or more.

All participants carried hand-held computers for a data collection period of seven consecutive days. They were trained to respond to random prompts from the computers and also to self-initiate event recordings of smoking episodes. Here, we will focus only on the random prompts, which were initiated by the device approximately four times per day. Each random prompt was date-stamped and time-stamped. The device also recorded whether the interview was completed, missed, delayed, or disbanded. A total of 14,105 random prompts were obtained on 3642 days from 461 participants with an approximate average of 30 prompts per participant (range = 7 to 71). We will illustrate our approach using three outcome measures: subject's negative affect (NA), social isolation (SI), and feeling tired and bored (TB). NA, SI, and TB consisted of the average of several mood items, which were rated on the scale from 1 to 10 (with 10 representing very high levels of specific attribute). Higher values of NA indicated higher negative mood, higher values of SI indicated high value for feeling socially isolated, and higher values of TB indicated high levels of feeling tired and bored.

For illustration, we include the covariates described in a previous analysis of these data [16]. Subject-level (level-3) covariates include *Smoker* (defined as presence of at least one smoking event during the EMA baseline data collection period, 1 = yes or 0 = no), *PropSmk* (an individual-level covariate that indicates the level of smoking, and it is defined as the number of smoking events over the total number of random prompts and smoking events), *Gender* (1 = male or 0 = female), *Grade10* (1 = 10th or 0 = 9th grade), *NovSeekC* (a continuous measure of novelty seeking with higher values indicating more novelty seeking), and *NegMoodRegC* (a continuous measure with higher values indicating better negative mood regulation). Among these covariates, *NovSeekC* and *NegMoodRegC* are grand mean centered. As a day-level (level-2) covariate, we include *WeekEnd* (0 = weekday indicating Monday to

Friday or 1 = weekend indicating Saturday and Sunday). At the prompt-level (level-1) covariate, we consider whether the subject was alone or accompanied by others (0 = not alone or 1 = alone) at the time of the random prompt. For this variable, we have created both BS and WS-WD (within subject within day) versions, *AloneBS* and *AloneWS*, using the decomposition $X_{ijk} = \bar{X}_{i..} + (X_{ijk} - \bar{X}_{i..})$ (see [37] for details). Here, $AloneBS = \bar{X}_{i..}$ is the proportion of random prompts in which a subject was alone, and $AloneWS = (X_{ijk} - \bar{X}_{i..})$ is the prompt-specific deviation from the proportion. Note that *AloneBS* is a subject-level covariate and *AloneWS* is a prompt-level covariate. For more details about the data, we refer to [16–18].

3. Bayesian mixed-effects location scale model for multivariate longitudinal outcomes

Our approach of constructing a joint model for multivariate longitudinal outcomes relies on a conditional independence assumption; that is, the outcome measurements within subject and between multiple measures are assumed to be independent conditional on the collection of random subject effects. In what follows, we describe the details of our model and discuss the interpretation of various parameters. Let us assume that for each outcome measure of interest $i = 1 \dots M$, we have $j = 1 \dots n$ subjects. We have further assumed that we have $l = 1 \dots n_{jk}$ prompts (level-1 measurements) nested within $k = 1 \dots n_j$ days (level-2 units). Therefore, we propose a three-level mixed-effects submodel for each outcome with the following specification

$$\begin{aligned} y_{ijkl} &= \mathbf{x}_{ijkl}^T \boldsymbol{\beta}_i + \eta_{ij} + \pi_{ijk} + \epsilon_{ijkl}, \\ \log\left(\sigma^2_{\epsilon_{ijkl}}\right) &= \mathbf{z}_{ijkl}^T \boldsymbol{\alpha}_i + \tau_{ij}, \\ \log\left(\sigma^2_{\pi_{ijk}}\right) &= \mathbf{u}_{ijk}^T \boldsymbol{\gamma}_i, \\ \log\left(\sigma^2_{\eta_{ij}}\right) &= \mathbf{v}_{ij}^T \boldsymbol{\lambda}_i. \end{aligned} \tag{1}$$

The vector $\mathbf{x}_{ijkl}$ contains the covariates at various levels for the ith outcome, and the vector $\boldsymbol{\beta}_i$ contains the $p_i \times 1$ fixed-effects corresponding to this set of covariates for this outcome. The random effects at the subject level (η_{ij}) and day level (π_{ijk}), respectively, for the ith outcome explain the heterogeneity at the mean level. The errors (ϵ_{ijkl}), which explain the variability of the scale, are assumed to be normally distributed and independent of each other and the random effects at various levels, that is, $\epsilon_{ijkl} \sim N\left(0, \sigma^2_{\epsilon_{ijkl}}\right)$. In addition to allowing for heteroscedasticity of random effects and errors, we have allowed for the inclusion of explanatory variables to the model for each of these variances by using a log-transformation. The vector $\boldsymbol{\alpha}_i$ contains the $q_i \times 1$ fixed-effects parameters corresponding to the $\mathbf{z}_{ijkl}$ covariates of the log-transformed error variance $\sigma^2_{\epsilon_{ijkl}}$. Similarly, the vector $\boldsymbol{\gamma}_i$ contains the $r_i \times 1$ fixed-effects parameters corresponding to the $\mathbf{u}_{ijk}$ covariates of the log-transformed variance of π_{ijk}, and the vector $\boldsymbol{\lambda}_i$ contains the $s_i \times 1$ fixed-effects corresponding to the $\mathbf{v}_{ij}$ covariates of the log-transformed variance of η_{ij}. We have also included random level-3 scale effect (τ_{ij}) for the error variance component. In order to capture the correlations among M multivariate outcomes, the random effects, π_{ijk}, at level-2 are assumed to be multivariate normally distributed as follows:

$$\begin{pmatrix} \pi_{1jk} \\ \pi_{2jk} \\ \vdots \\ \pi_{Mjk} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\pi_{1jk}} & \sigma_{\pi_{1jk}\pi_{2jk}} & \cdots & \sigma_{\pi_{1jk}\pi_{Mjk}} \\ \sigma_{\pi_{2jk}\pi_{1jk}} & \sigma^2_{\pi_{2jk}} & \cdots & \sigma_{\pi_{2jk}\pi_{Mjk}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\pi_{Mjk}\pi_{1jk}} & \sigma_{\pi_{Mjk}\pi_{2jk}} & \cdots & \sigma^2_{\pi_{Mjk}} \end{pmatrix}\right], \tag{2}$$

and the random effects at both location (η_{ij}) and scale (τ_{ij}) of level-3 are also assumed to follow a multivariate normal distribution as

$$\begin{pmatrix} \eta_{1j} \\ \eta_{2j} \\ \vdots \\ \eta_{Mj} \\ \tau_{1j} \\ \tau_{2j} \\ \vdots \\ \tau_{Mj} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\eta_{1j}} & \sigma_{\eta_{1j}\eta_{2j}} & \cdots & \sigma_{\eta_{1j}\eta_{Mj}} & \sigma_{\eta_{1j}\tau_{1j}} & \sigma_{\eta_{1j}\tau_{2j}} & \cdots & \sigma_{\eta_{1j}\tau_{Mj}} \\ \sigma_{\eta_{2j}\eta_{1j}} & \sigma^2_{\eta_{2j}} & \cdots & \sigma_{\eta_{2j}\eta_{Mj}} & \sigma_{\eta_{2j}\tau_{1j}} & \sigma_{\eta_{2j}\tau_{2j}} & \cdots & \sigma_{\eta_{2j}\tau_{Mj}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\eta_{Mj}\eta_{1j}} & \sigma_{\eta_{Mj}\eta_{2j}} & \cdots & \sigma^2_{\eta_{Mj}} & \sigma_{\eta_{Mj}\tau_{1j}} & \sigma_{\eta_{Mj}\tau_{2j}} & \cdots & \sigma_{\eta_{Mj}\tau_{Mj}} \\ \sigma_{\tau_{1j}\eta_{1j}} & \sigma_{\tau_{1j}\eta_{2j}} & \cdots & \sigma_{\tau_{1j}\eta_{Mj}} & \sigma^2_{\tau_{1j}} & \sigma_{\tau_{1j}\tau_{2j}} & \cdots & \sigma_{\tau_{1j}\tau_{Mj}} \\ \sigma_{\tau_{2j}\eta_{1j}} & \sigma_{\tau_{2j}\eta_{2j}} & \cdots & \sigma_{\tau_{2j}\eta_{Mj}} & \sigma_{\tau_{2j}\tau_{1j}} & \sigma^2_{\tau_{2j}} & \cdots & \sigma_{\tau_{2j}\tau_{Mj}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\tau_{Mj}\eta_{1j}} & \sigma_{\tau_{Mj}\eta_{2j}} & \cdots & \sigma_{\tau_{Mj}\eta_{Mj}} & \sigma_{\tau_{Mj}\tau_{1j}} & \sigma_{\tau_{Mj}\tau_{2j}} & \cdots & \sigma^2_{\tau_{Mj}} \end{pmatrix} \right]. \tag{3}$$

The variance parameters $\sigma^2_{\epsilon_{ijkl}}$ are the variability within day (level-2) of individual (level-3) measurements. Similarly, the random-effect variances $\sigma^2_{\pi_{ijk}}$ of π_{ijk} are the estimate of variability between level-2 units within level-3 measurements. The variance $\sigma^2_{\eta_{ij}}$ of η_{ij} provides an estimate of variability between level-3 measurements.

We have used a spherical parametrization for the variance–covariance matrix of the random effects π_{ijk} at level-2 and the variance–covariance matrix of the location (η_{ij}) and scale (τ_{ij}) random effects at level-3, which easily accommodates modeling of variances in terms of covariates with the help of a separation strategy (for more details, see [38] and [39]). In addition, it allows for a simpler specification of diffuse priors for the correlation terms and remaining variances that are not dependent on the covariates. With respect to optimization, this re-parametrization ensures positive semi-definiteness of Σ by limiting the search within a positive semi-definite cone (see [40]). Pinheiro [41] has shown via an extensive simulation study that the spherical parametrization has the best combination of performance in terms of time and number of iterations. Further, it leads to simpler interpretations as compared with other parametrization such as the Cholesky and log-Cholesky techniques. Note that Zhang *et al.* [21] proposed a very similar spherical parametrization approach in order to guarantee positive definiteness of the correlation matrix of longitudinal outcomes. They have also provided extensive geometric interpretation of the correlations in terms of spherical parameters. However, one important limitation of this parametrization is that it does not ensure rotational invariance, and hence it is not possible to establish a priori invariance of priors for the correlations to the permutations of indices and ordering of the outcomes.

From the definition of separation strategy, we can rewrite the covariance matrix as a product of the correlation matrix ($\boldsymbol{R}$) and the diagonal matrix containing standard deviations ($\boldsymbol{V}^{1/2}$) in the following form

$$\Sigma = \boldsymbol{V}^{1/2}\boldsymbol{R}\boldsymbol{V}^{1/2}. \tag{4}$$

Using the Cholesky decomposition, we can define $\boldsymbol{R} = \boldsymbol{L}^T\boldsymbol{L}$. Here, $\boldsymbol{L}$ is the upper-triangular matrix. The spherical coordinates of the first p elements of the pth column of $\boldsymbol{L}$ are given by

$$\begin{aligned} l_{p1} &= \cos\left(\delta_{p2}\right), \\ l_{p2} &= \sin\left(\delta_{p2}\right)\cos\left(\delta_{p3}\right), \\ &\vdots \\ l_{p,p-1} &= \sin\left(\delta_{p2}\right)\sin\left(\delta_{p3}\right)\ldots\cos\left(\delta_{pp}\right), \\ l_{p,p} &= \sin\left(\delta_{p2}\right)\sin\left(\delta_{p3}\right)\ldots\sin\left(\delta_{pp}\right). \end{aligned} \tag{5}$$

The first element l_{11} is set to 1. The spherical angles δ_{pr} are restricted to $(0, \pi)$ for uniqueness property. Therefore, the (s, t) element of $\boldsymbol{R}$ denoted by R_{st} is obtained as the inner product of sth column of $\boldsymbol{L}^T$ with tth column of $\boldsymbol{L}$.

In order to reduce the number of parameters, we have fixed the correlation between the multiple outcomes across subjects, that is, $\rho_{\pi_{ijk}\pi_{i'jk}} = \rho_{\pi(ii')}$, so correlations at level-2 are not allowed to vary across subjects or between days.

For the variance–covariance matrix of the random location and scale effects at level-3, Hedeker *et al.* [16] and Li and Hedeker [17] allowed the correlations ($\rho_{\eta_{ij}\tau_{ij}}$) to vary across subjects and fixed the variance of scale parameters, that is, $\sigma^2_{\tau_{ij}} = \sigma^2_{\tau(i)}$ and covariance between location and scale parameters, $\sigma_{\eta_{ij}\tau_{ij}} = \sigma_{\eta\tau(i)}$ in their models. In order to simplify the specification of priors for the correlations obtained

using spherical parametrization and reduce number of parameters in the correlation matrices, we have fixed the variance of scale, that is, $\sigma^2_{\tau_{ij}} = \sigma^2_{\tau(i)}$, and the correlation parameters of the location and scale of each outcome, that is, $\rho_{\eta_{ij}\tau_{ij}} = \rho_{\eta\tau(i)}$, and only allowed the variance of location $\sigma^2_{\eta_{ij}}$ parameters to vary across subjects for all the outcomes of the model in Equation (1). The covariance between the random location and scale vary proportionally to the variance of location parameter across subjects by virtue of the well-defined relationship, $\sigma_{\eta_{ij}\tau_{ij}} = \rho_{\eta\tau(i)}\sigma_{\eta_{ij}}\sigma_{\tau(i)}$. We have used this approach for the correlations and cross-correlations between the two location parameters, two scale parameters, or location and scale parameters of the two different outcomes. Hence, the covariances between multiple outcomes in our current model are influenced indirectly by the specific covariates through the relationships $\sigma_{\pi_{ijk}\pi_{i'jk}} = \rho_{\pi(ii')}\sigma_{\pi_{ijk}}\sigma_{\pi_{i'jk}} = \rho_{\pi(ii')}\sqrt{\exp\left(\mathbf{u}^T_{ijk}\boldsymbol{\gamma}_i\right)\exp\left(\mathbf{u}^T_{i'jk}\boldsymbol{\gamma}_{i'}\right)}$ and $\sigma_{\eta_{ij}\eta_{i'j}} = \rho_{\eta(ii')}\sigma_{\eta_{ij}}\sigma_{\eta_{i'j}} = \rho_{\eta(ii')}\sqrt{\exp\left(\mathbf{v}^T_{ij}\boldsymbol{\lambda}_i\right)\exp\left(\mathbf{v}^T_{i'j}\boldsymbol{\lambda}_{i'}\right)}$. The intra-class correlation in our approach can be estimated approximately using the average of variances over the subjects. Therefore, our approach is very flexible in accommodating various features of joint location and scale models while not being limited by any modeling restrictions.

In the next section, we will present the details of model likelihood and priors for the parameters of our model.

4. Model likelihood and priors

We denote the vector of M outcomes as $\boldsymbol{y}_{jkl} = (y_{1jkl}, y_{2jkl}, \ldots, y_{Mjkl})^T$ and the vector of fixed-effects for M outcomes as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M)^T$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_M)^T$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_M)^T$, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_M)^T$. Similarly, we denote the vector of random effects for M outcomes as $\boldsymbol{\eta}_j = (\eta_{1j}, \eta_{2j}, \ldots, \eta_{Mj})^T$, $\boldsymbol{\tau}_j = (\tau_{1j}, \tau_{2j}, \ldots, \tau_{Mj})^T$, and $\boldsymbol{\pi}_{jk} = (\pi_{1jk}, \pi_{2jk}, \ldots, \pi_{Mjk})^T$. Also, denote $\boldsymbol{X}_{jkl} = \bigoplus_{i=1}^{M} \mathbf{x}_{ijkl}$, $\boldsymbol{Z}_{jkl} = \bigoplus_{i=1}^{M} \mathbf{z}_{ijkl}$, $\boldsymbol{U}_{jk} = \bigoplus_{i=1}^{M} \mathbf{u}_{ijk}$, and $\boldsymbol{V}_j = \bigoplus_{i=1}^{M} \mathbf{v}_{ij}$. We use a log-normal distribution as the prior for the variance of the scale effects. Hence, define $\log(\sigma^2_{\tau_i}) = \boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_M)^T$. Denote the vector of spherical angles for the variance–covariance matrix of location and scale parameters at level-3 by $\boldsymbol{\omega}$ and the vector of spherical angles for the variance–covariance matrix of the random effects π_{ijk} at level-2 by $\boldsymbol{\psi}$. Finally, denote the vector of overall fixed-effects by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda, \boldsymbol{\zeta})^T$, the overall vector of random effects at level-2 and level-3 by $\boldsymbol{\phi} = (\boldsymbol{\eta}_j, \boldsymbol{\tau}_j, \boldsymbol{\pi}_{jk})^T$, and the overall vector of spherical angles by $\boldsymbol{\delta} = (\boldsymbol{\omega}, \boldsymbol{\psi})^T$.

Therefore, the joint posterior distribution for the vector of fixed-effects parameters, spherical parameters, and the random effects assuming ignorable missingness is as follows:

$$p\left(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\delta} | \boldsymbol{y}_{jkl}, \boldsymbol{X}_{jkl}, \boldsymbol{Z}_{jkl}, \boldsymbol{U}_{jk}, \boldsymbol{V}_j\right) \propto$$
$$\prod_{j=1}^{n}\prod_{k=1}^{n_j}\prod_{l=1}^{n_{jk}} p\left(\boldsymbol{y}_{jkl} | \boldsymbol{X}_{jkl}, \boldsymbol{Z}_{jkl}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}_j, \boldsymbol{\tau}_j, \boldsymbol{\pi}_{jk}\right) p\left(\boldsymbol{\pi}_{jk} | \boldsymbol{U}_{jk}, \boldsymbol{\gamma}, \boldsymbol{\psi}\right) p\left(\boldsymbol{\eta}_j, \boldsymbol{\tau}_j | \boldsymbol{V}_j, \lambda, \boldsymbol{\zeta}, \boldsymbol{\omega}\right) \times$$
$$p(\boldsymbol{\beta}) p(\boldsymbol{\alpha}) p(\boldsymbol{\gamma}) p(\lambda) p(\boldsymbol{\zeta}) p(\boldsymbol{\psi}) p(\boldsymbol{\omega}),$$

where

$$p\left(\boldsymbol{y}_{jkl} | \boldsymbol{X}_{jkl}, \boldsymbol{Z}_{jkl}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}_j, \boldsymbol{\tau}_j, \boldsymbol{\pi}_{jk}\right) = N\left(\boldsymbol{X}^T_{jkl}\boldsymbol{\beta} + \boldsymbol{\eta}_j + \boldsymbol{\pi}_{jk}, Diag\left(\exp\left(\boldsymbol{Z}^T_{jkl}\boldsymbol{\alpha} + \boldsymbol{\tau}_j\right)\right)\right),$$
$$p\left(\boldsymbol{\pi}_{jk} | \boldsymbol{U}_{jk}, \boldsymbol{\gamma}, \boldsymbol{\psi}\right) = N\left(\mathbf{0}_M, \boldsymbol{\Sigma}\left(\boldsymbol{U}_{jk}, \boldsymbol{\gamma}, \boldsymbol{\psi}\right)\right),$$
$$p\left(\boldsymbol{\eta}_j, \boldsymbol{\tau}_j | \boldsymbol{V}_j, \lambda, \boldsymbol{\zeta}, \boldsymbol{\omega}\right) = N\left(\mathbf{0}_{2M}, \boldsymbol{\Sigma}\left(\boldsymbol{V}_j, \lambda, \boldsymbol{\zeta}, \boldsymbol{\omega}\right)\right).$$

$\boldsymbol{\Sigma}(\boldsymbol{U}_{jk}, \boldsymbol{\gamma}, \boldsymbol{\psi})$ and $\boldsymbol{\Sigma}(\boldsymbol{V}_j, \lambda, \boldsymbol{\zeta}, \boldsymbol{\omega})$ are the corresponding variance–covariance matrices of the random effects at level-2 and random location and scale effects at level-3 as defined in Equations (2) and (3). They can be represented using the separation strategy described in Equations (4) and (5) along with the respective spherical angles $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ and the model definition in Equation (1) of log-transformed variance parameters as

$$\boldsymbol{\Sigma}(\boldsymbol{V}_j, \lambda, \boldsymbol{\zeta}, \boldsymbol{\omega}) = \left(Diag\left(\exp\left(\boldsymbol{V}_j^T \lambda\right), \exp(\boldsymbol{\zeta})\right)\right)^{\frac{1}{2}} \boldsymbol{L}(\boldsymbol{\omega})^T \boldsymbol{L}(\boldsymbol{\omega}) \left(Diag\left(\exp\left(\boldsymbol{V}_j^T \lambda\right), \exp(\boldsymbol{\zeta})\right)\right)^{\frac{1}{2}},$$
$$\text{and } \boldsymbol{\Sigma}\left(\boldsymbol{U}_{jk}, \boldsymbol{\gamma}, \boldsymbol{\psi}\right) = \left(Diag\left(\exp\left(\boldsymbol{U}_{jk}^T \boldsymbol{\gamma}\right)\right)\right)^{\frac{1}{2}} \boldsymbol{L}(\boldsymbol{\psi})^T \boldsymbol{L}(\boldsymbol{\psi}) \left(Diag\left(\exp\left(\boldsymbol{U}_{jk}^T \boldsymbol{\gamma}\right)\right)\right)^{\frac{1}{2}}, \quad (6)$$

where the matrix $\boldsymbol{L}(\boldsymbol{\omega})$ and $\boldsymbol{L}(\boldsymbol{\psi})$ are the upper-triangular matrix as defined in Equation (5).

We have specified vague priors for the fixed-effects regression coefficients using normal distribution, that is, $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma^2_{\beta_0} I_{\mathbf{p}})$, $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma^2_{\alpha_0} I_{\mathbf{q}})$, $\boldsymbol{\gamma} \sim N(\mathbf{0}, \sigma^2_{\gamma_0} I_{\mathbf{r}})$, and $\lambda \sim N(\mathbf{0}, \sigma^2_{\lambda_0} I_{\mathbf{s}})$ with $\mathbf{p} = \sum_{i=1}^{M} p_i$, $\mathbf{q} = \sum_{i=1}^{M} q_i$, $\mathbf{r} = \sum_{i=1}^{M} r_i$, $\mathbf{s} = \sum_{i=1}^{M} s_i$, and $\sigma^2_{\beta_0} = \sigma^2_{\alpha_0} = \sigma^2_{\gamma_0} = \sigma^2_{\lambda_0} = 10^3$. We have tried two diffuse priors for the variance of the scale effects τ_{ij}: gamma distribution for the inverse of variance with very small scale and shape parameters, that is, $G(0.001, 0.001)$, and log-normal distribution with the variance $(\sigma^2_{\zeta_0})$ of the order 10^3. In this manuscript, we present results based on log-normal priors in order to keep the specification consistent with rest of the model. This choice also leads to scale invariant priors [38]. We have also assumed that $\boldsymbol{R}$ and $\boldsymbol{V}$ are independent in our prior specification scheme and the priors for the spherical parameters are set to the uniform distribution over the interval between $(0, \pi)$. Note that the spherical parametrization is not rotationally invariant.

In order to implement the Gibbs sampling approach for estimation (for more details, see the seminal papers such as [42–51]), we need to specify the conditional distributions for the fixed-effects, random effects, and spherical parameters. These conditional distributions are provided in the Appendix A.

We implemented our model using the WinBUGS 1.4 software [52] and R2WinBUGS package [53] in R 3.0.2 software. The model was run for 50,000 iterations with 20,000 iterations for burn-in followed by 30,000 samples for estimates using three parallel chains. We used thinning of 10 samples for our final model. The convergence of our model was monitored on the basis of MCMC chains, auto-correlations, cross-correlations, density plots, and the Raferty–Lewis diagnostic test.

In the next section, we present the results of simulated data in order to show the impact of the aforementioned prior specification scheme on the parameter estimates.

5. Simulation study

In this section, we demonstrate that the parameters estimated on the basis of prior specification using the separation strategy and the spherical parametrization have desirable properties although the prior specification is not invariant to the rotation of correlation parameters in terms of spherical parameters. In our limited simulation study, we simulated data with bivariate outcomes using the following model parameters and covariates:

- $M = 2, n = 400, n_j = 5$, and $n_{jk} = 5$ for all j and k.
- The covariate vectors are set as $\mathbf{x}_{ijkl} = (1, X_1, X_2, X_3, X_4), \mathbf{z}_{ijkl} = (1, X_1, X_2, X_3, X_4), \mathbf{u}_{ijk} = (1, X_1, X_2, X_4)$, and $\mathbf{v}_{ij} = (1, X_1, X_2)$. X_1 and X_2 are level-3 covariates. X_1 is a binary variable taking value 1 with $p = 0.5$. X_2 is a normally distributed covariate with mean 0.5 and variance 0.04. X_3 is a normally distributed covariate at level-1 with mean 0.25 and variance of 0.01. X_4 is a binary variable at level-2 with success probability of 0.30.
- The fixed-effects are set as $\boldsymbol{\beta}_1 = (3.5, 0.30, 0.70, 0.15, -0.20)^T$, $\boldsymbol{\beta}_2 = (2.5, 0.10, 0.125, 0.20, -0.10)^T$, $\boldsymbol{\alpha}_1 = (0.30, 0.40, 0.50, 0.10, 0.10)^T$, $\boldsymbol{\alpha}_2 = (0.15, 0.25, 0.25, 0.40, 0.05)^T$, $\boldsymbol{\gamma}_1 = (0.00, 0.25, -0.65, -0.10)^T$, $\boldsymbol{\gamma}_2 = (-1.25, 0.25, 0.25, 0.10)^T$, $\lambda_1 = (-0.10, 0.20, 0.90)^T$, $\lambda_2 = (-0.15, 0.45, 0.90)^T$, $\zeta_1 = -0.15$, and $\zeta_2 = 0.35$.
- The correlation of the random location effects at level-2 is set to 0.50, and the correlation terms of the random location and scale effects at level-3 are chosen as $\rho_{\eta(12)} = 0.75$, $\rho_{\eta\tau(11)} = 0.50$, $\rho_{\eta\tau(12)} = 0.60$, $\rho_{\eta\tau(21)} = 0.30$, $\rho_{\eta\tau(22)} = 0.60$, and $\rho_{\tau(12)} = 0.65$.

These parameters are chosen to be very similar to our current example. The posterior probabilities of the parameters estimated using the WinBUGS 1.4 software are presented in Figures 1–3. From the Figures, we note that the true parameters are always within in the 95% high density interval for these data with total sample size of 10,000 ($n = 400$, $n_j = 5$, and $n_{jk} = 5$) for each outcome. The posterior densities of fixed-effect parameters (Figures 1 and 2) are close to normal, and the mode of the distribution is also very close to the true parameters for all the covariates at level-1, level-2, and level-3. Similarly, the posterior densities of the correlation parameters (Figure 3) are also normally distributed; however, the

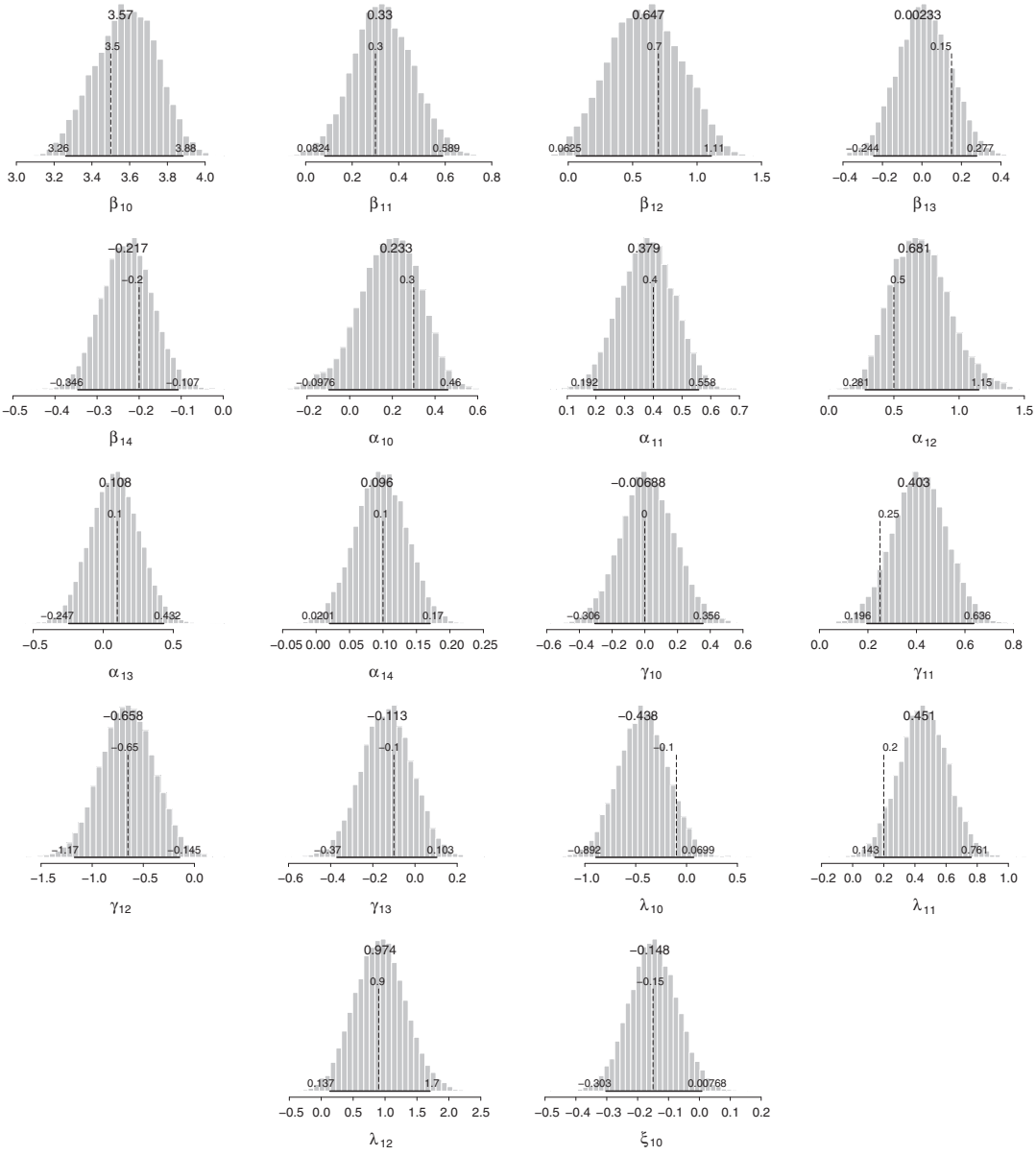

Figure 1. Plots of posterior densities along with 95% high density interval and true fixed-effects for first outcome of simulated data.

mode of all the correlation parameters are shifted leftward. This is because the uniform prior specification of the spherical parameters do not lead to uniform priors for the correlation as spherical parametrization is not rotationally invariant. The marginal density functions of the correlations are U-shaped with different densities on the edges and thickness in the middle in comparison with each other. From the inference standpoint, the true correlation parameter is contained within the 95% high density intervals, and this interval does not contain 0 for our simulated data with the given sample size.

Figure 4 contains the posterior densities of correlation parameters for the same dataset but after switching the indices of outcomes to understand the impact of permutation of indices. Comparing Figures 3 and 4, we see that the posterior densities and the mode of these posterior densities of correlations are very similar to each other. The mode of densities are still shifted leftward. Additionally, the posterior densities of fixed-effects parameters (not shown) for these two data with switched labels are almost identical. Hence, in our current simulated example, the permutation of indices does not seem to have impact on the final estimates of correlations for the given sample size and the correlation parameters. Note that we have only investigated the behavior of permutation of indices using one set of simulated data. Although

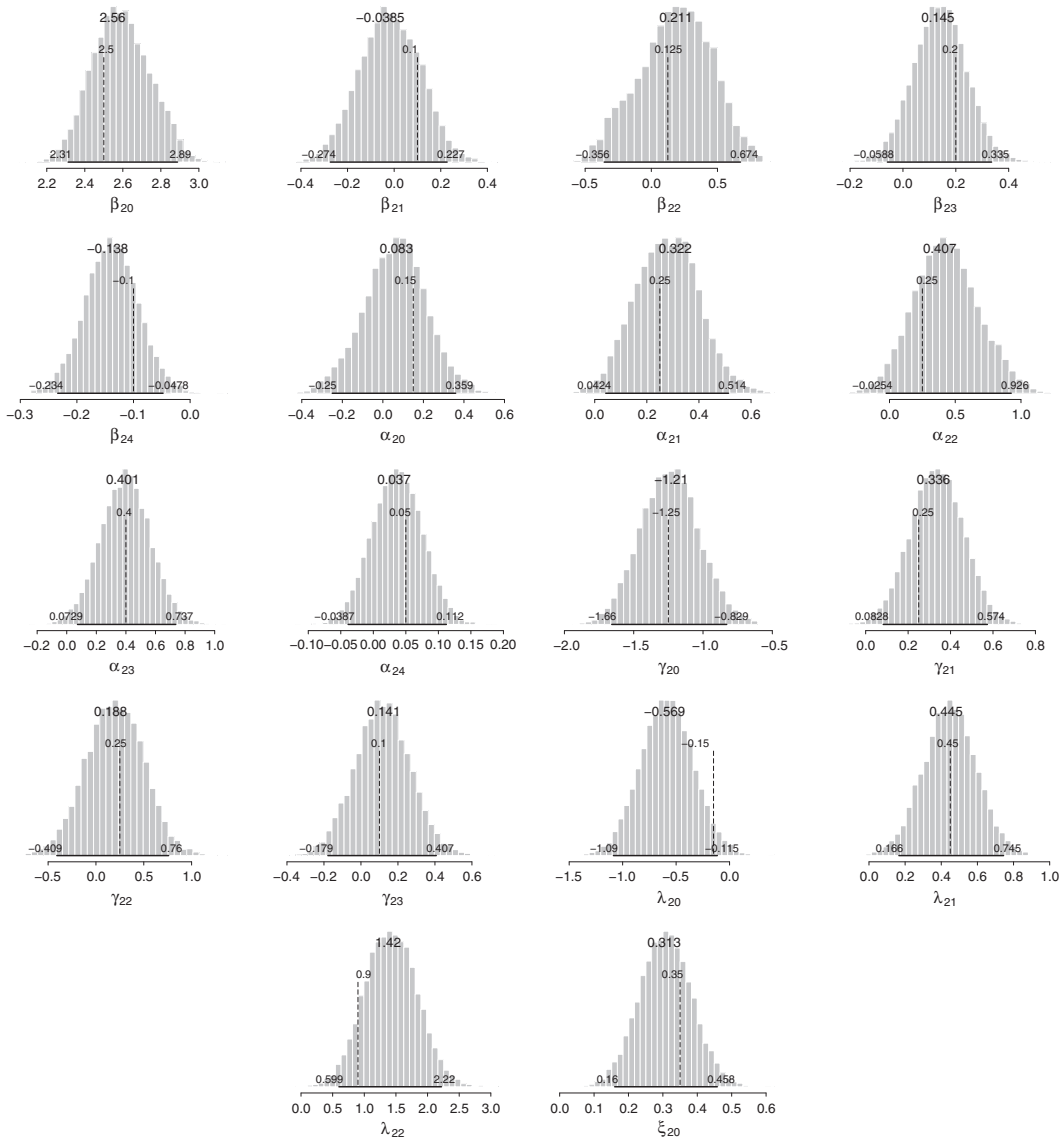

Figure 2. Plots of posterior densities along with 95% high density interval and true fixed-effects for second outcome of simulated data.

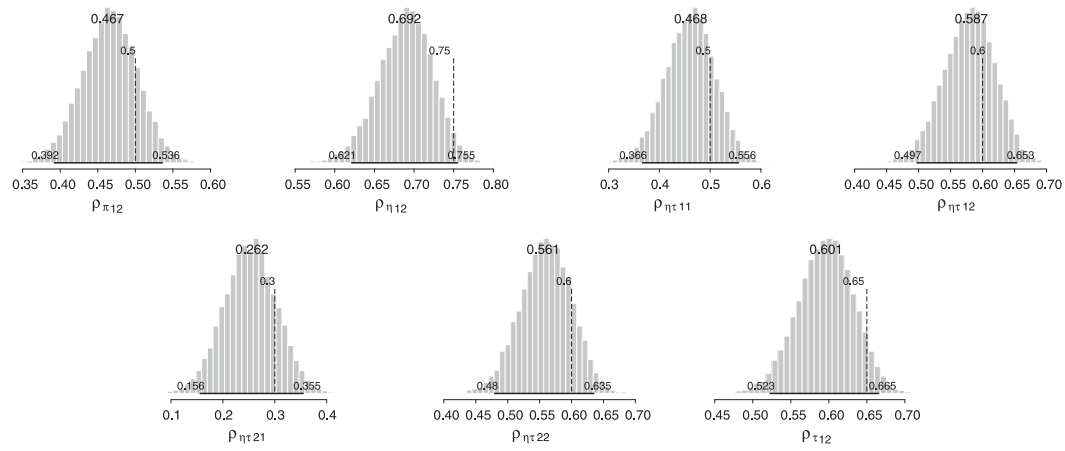

Figure 3. Plot of posterior densities along with 95% high density interval and true correlations of simulated data corresponding to first simulation run.

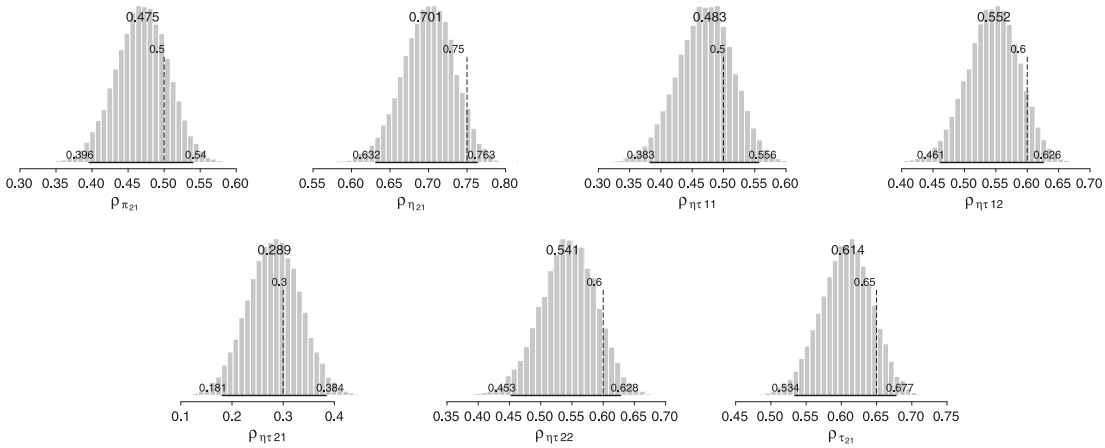

Figure 4. Plot of posterior densities along with 95% high density interval and true correlations of simulated data after switching the labels of two outcomes.

we have not encountered any problem with this approach for our simulated data, we believe that an additional simulation study with a larger number of outcomes and various combinations of parameters and sample sizes is needed to fully understand the behavior of permutation of indices on the final estimates of the correlation parameters using spherical parametrization.

We also refer to Barnard *et al.* [38] for further discussion on the issues relating to choice and impact of different prior specifications on spherical parametrization.

6. Results

For our current example, the outcomes y_{1jkl}, y_{2jkl}, and y_{3jkl} are the NA, SI, and TB, respectively, of subject j on day k for the occasion l. Before considering the three-level model, we consider a simpler two-level model, by ignoring the nesting of occasions within days. Here we have

$$\begin{aligned}
y_{1jk} &= \mathbf{x}_{1jk}^T\boldsymbol{\beta}_1 + \eta_{1j} + \epsilon_{1jk},\\
\log\left(\sigma^2_{\epsilon_{1jk}}\right) &= \mathbf{z}_{1jk}^T\boldsymbol{\alpha}_1 + \tau_{1j},\\
\log\left(\sigma^2_{\eta_{1j}}\right) &= \mathbf{v}_{1j}^T\boldsymbol{\lambda}_1,\\
y_{2jk} &= \mathbf{x}_{2jk}^T\boldsymbol{\beta}_2 + \eta_{2j} + \epsilon_{2jk},\\
\log\left(\sigma^2_{\epsilon_{2jk}}\right) &= \mathbf{z}_{2jk}^T\boldsymbol{\alpha}_2 + \tau_{2j},\\
\log\left(\sigma^2_{\eta_{2j}}\right) &= \mathbf{v}_{2j}^T\boldsymbol{\lambda}_2,\\
y_{3jk} &= \mathbf{x}_{3jk}^T\boldsymbol{\beta}_2 + \eta_{3j} + \epsilon_{3jk},\\
\log\left(\sigma^2_{\epsilon_{3jk}}\right) &= \mathbf{z}_{3jk}^T\boldsymbol{\alpha}_2 + \tau_{3j},\\
\log\left(\sigma^2_{\eta_{3j}}\right) &= \mathbf{v}_{3j}^T\boldsymbol{\lambda}_3.
\end{aligned}$$

As discussed in Section 3, we have allowed the variance of location ($\sigma^2_{\eta_{ij}}$) parameters to vary across subjects, but we have fixed the correlation parameters of the location and scale of each outcome ($\rho_{\eta_{ij}\tau_{ij}}$). We have also fixed the variance of the scale parameter over subjects. Hence, we can write the variance–covariance of the random effects in Equation (3) explicitly as follows:

$$\begin{pmatrix}\eta_{1j}\\ \eta_{2j}\\ \eta_{3j}\\ \tau_{1j}\\ \tau_{2j}\\ \tau_{3j}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix} \sigma^2_{\eta_{1j}} & \rho_{\eta(12)}\sigma_{\eta_{1j}}\sigma_{\eta_{2j}} & \rho_{\eta(13)}\sigma_{\eta_{1j}}\sigma_{\eta_{3j}} & \rho_{\eta\tau(11)}\sigma_{\eta_{1j}}\sigma_{\tau(1)} & \rho_{\eta\tau(12)}\sigma_{\eta_{1j}}\sigma_{\tau(2)} & \rho_{\eta\tau(13)}\sigma_{\eta_{1j}}\sigma_{\tau(3)} \\ \rho_{\eta(21)}\sigma_{\eta_{2j}}\sigma_{\eta_{1j}} & \sigma^2_{\eta_{2j}} & \rho_{\eta(23)}\sigma_{\eta_{2j}}\sigma_{\eta_{3j}} & \rho_{\eta\tau(21)}\sigma_{\eta_{2j}}\sigma_{\tau(1)} & \rho_{\eta\tau(22)}\sigma_{\eta_{2j}}\sigma_{\tau(2)} & \rho_{\eta\tau(23)}\sigma_{\eta_{2j}}\sigma_{\tau(3)} \\ \rho_{\eta(31)}\sigma_{\eta_{3j}}\sigma_{\eta_{1j}} & \rho_{\eta(32)}\sigma_{\eta_{3j}}\sigma_{\eta_{2j}} & \sigma^2_{\eta_{3j}} & \rho_{\eta\tau(31)}\sigma_{\eta_{3j}}\sigma_{\tau(1)} & \rho_{\eta\tau(32)}\sigma_{\eta_{3j}}\sigma_{\tau(2)} & \rho_{\eta\tau(33)}\sigma_{\eta_{3j}}\sigma_{\tau(3)} \\ \rho_{\tau\eta(11)}\sigma_{\tau(1)}\sigma_{\eta_{1j}} & \rho_{\tau\eta(12)}\sigma_{\tau(1)}\sigma_{\eta_{2j}} & \rho_{\tau\eta(13)}\sigma_{\tau(1)}\sigma_{\eta_{3j}} & \sigma^2_{\tau(1)} & \rho_{\tau(12)}\sigma_{\tau(1)}\sigma_{\tau(2)} & \rho_{\tau(13)}\sigma_{\tau(1)}\sigma_{\tau(3)} \\ \rho_{\tau\eta(21)}\sigma_{\tau(2)}\sigma_{\eta_{1j}} & \rho_{\tau\eta(22)}\sigma_{\tau(2)}\sigma_{\eta_{2j}} & \rho_{\tau\eta(23)}\sigma_{\tau(2)}\sigma_{\eta_{3j}} & \rho_{\tau(21)}\sigma_{\tau(2)}\sigma_{\tau(1)} & \sigma^2_{\tau(2)} & \rho_{\tau(23)}\sigma_{\tau(2)}\sigma_{\tau(3)} \\ \rho_{\tau\eta(31)}\sigma_{\tau(3)}\sigma_{\eta_{1j}} & \rho_{\tau\eta(32)}\sigma_{\tau(3)}\sigma_{\eta_{2j}} & \rho_{\tau\eta(33)}\sigma_{\tau(3)}\sigma_{\eta_{3j}} & \rho_{\tau(31)}\sigma_{\tau(3)}\sigma_{\tau(1)} & \rho_{\tau(32)}\sigma_{\tau(3)}\sigma_{\tau(2)} & \sigma^2_{\tau(3)} \end{pmatrix}\right]. \tag{7}$$

The parameter estimates along with their standard errors (SE) and 95% credible intervals (CI) of a two-level model for trivariate outcomes (NA, SI, and TB) are presented in Table I.

The three-level model taking the additional nesting of occasions within days for our current example can be written as

$$\begin{aligned} y_{1jkl} &= \mathbf{x}^T_{1jkl}\boldsymbol{\beta}_1 + \eta_{1j} + \pi_{1jk} + \epsilon_{1jkl}, \\ \log\left(\sigma^2_{\epsilon_{1jkl}}\right) &= \mathbf{z}^T_{1jkl}\boldsymbol{\alpha}_1 + \tau_{1j}, \\ \log\left(\sigma^2_{\pi_{1jk}}\right) &= \mathbf{u}^T_{1jk}\boldsymbol{\gamma}_1, \\ \log\left(\sigma^2_{\eta_{1j}}\right) &= \mathbf{v}^T_{1j}\boldsymbol{\lambda}_1, \\ y_{2jkl} &= \mathbf{x}^T_{2jkl}\boldsymbol{\beta}_2 + \eta_{2j} + \pi_{2jk} + \epsilon_{2jkl}, \\ \log\left(\sigma^2_{\epsilon_{2jkl}}\right) &= \mathbf{z}^T_{2jkl}\boldsymbol{\alpha}_2 + \tau_{2j}, \\ \log\left(\sigma^2_{\pi_{2jk}}\right) &= \mathbf{u}^T_{2jk}\boldsymbol{\gamma}_2, \\ \log\left(\sigma^2_{\eta_{2j}}\right) &= \mathbf{v}^T_{2j}\boldsymbol{\lambda}_2, \\ y_{3jkl} &= \mathbf{x}^T_{3jkl}\boldsymbol{\beta}_2 + \eta_{3j} + \pi_{3jk} + \epsilon_{3jkl}, \\ \log\left(\sigma^2_{\epsilon_{3jkl}}\right) &= \mathbf{z}^T_{3jkl}\boldsymbol{\alpha}_2 + \tau_{3j}, \\ \log\left(\sigma^2_{\pi_{3jk}}\right) &= \mathbf{u}^T_{3jk}\boldsymbol{\gamma}_3, \\ \log\left(\sigma^2_{\eta_{3j}}\right) &= \mathbf{v}^T_{3j}\boldsymbol{\lambda}_3, \end{aligned}$$

the variance–covariance matrix of the random location effects at level-2 can be written as

$$\begin{pmatrix}\pi_{1jk}\\ \pi_{2jk}\\ \pi_{3jk}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix} \sigma^2_{\pi_{1jk}} & \rho_{\pi(12)}\sigma_{\pi_{1jk}}\sigma_{\pi_{2jk}} & \rho_{\pi(13)}\sigma_{\pi_{1jk}}\sigma_{\pi_{3jk}} \\ \rho_{\pi(21)}\sigma_{\pi_{2jk}}\sigma_{\pi_{1jk}} & \sigma^2_{\pi_{2jk}} & \rho_{\pi(23)}\sigma_{\pi_{2jk}}\sigma_{\pi_{3jk}} \\ \rho_{\pi(31)}\sigma_{\pi_{3jk}}\sigma_{\pi_{1jk}} & \rho_{\pi(32)}\sigma_{\pi_{3jk}}\sigma_{\pi_{2jk}} & \sigma^2_{\pi_{3jk}} \end{pmatrix}\right], \tag{8}$$

and the variance–covariance matrix of the random location and scale effects remains the same as before for the two-level model,

$$\begin{pmatrix}\eta_{1j}\\ \eta_{2j}\\ \eta_{3j}\\ \tau_{1j}\\ \tau_{2j}\\ \tau_{3j}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix} \sigma^2_{\eta_{1j}} & \rho_{\eta(12)}\sigma_{\eta_{1j}}\sigma_{\eta_{2j}} & \rho_{\eta(13)}\sigma_{\eta_{1j}}\sigma_{\eta_{3j}} & \rho_{\eta\tau(11)}\sigma_{\eta_{1j}}\sigma_{\tau(1)} & \rho_{\eta\tau(12)}\sigma_{\eta_{1j}}\sigma_{\tau(2)} & \rho_{\eta\tau(13)}\sigma_{\eta_{1j}}\sigma_{\tau(3)} \\ \rho_{\eta(21)}\sigma_{\eta_{2j}}\sigma_{\eta_{1j}} & \sigma^2_{\eta_{2j}} & \rho_{\eta(23)}\sigma_{\eta_{2j}}\sigma_{\eta_{3j}} & \rho_{\eta\tau(21)}\sigma_{\eta_{2j}}\sigma_{\tau(1)} & \rho_{\eta\tau(22)}\sigma_{\eta_{2j}}\sigma_{\tau(2)} & \rho_{\eta\tau(23)}\sigma_{\eta_{2j}}\sigma_{\tau(3)} \\ \rho_{\eta(31)}\sigma_{\eta_{3j}}\sigma_{\eta_{1j}} & \rho_{\eta(32)}\sigma_{\eta_{3j}}\sigma_{\eta_{2j}} & \sigma^2_{\eta_{3j}} & \rho_{\eta\tau(31)}\sigma_{\eta_{3j}}\sigma_{\tau(1)} & \rho_{\eta\tau(32)}\sigma_{\eta_{3j}}\sigma_{\tau(2)} & \rho_{\eta\tau(33)}\sigma_{\eta_{3j}}\sigma_{\tau(3)} \\ \rho_{\tau\eta(11)}\sigma_{\tau(1)}\sigma_{\eta_{1j}} & \rho_{\tau\eta(12)}\sigma_{\tau(1)}\sigma_{\eta_{2j}} & \rho_{\tau\eta(13)}\sigma_{\tau(1)}\sigma_{\eta_{3j}} & \sigma^2_{\tau(1)} & \rho_{\tau(12)}\sigma_{\tau(1)}\sigma_{\tau(2)} & \rho_{\tau(13)}\sigma_{\tau(1)}\sigma_{\tau(3)} \\ \rho_{\tau\eta(21)}\sigma_{\tau(2)}\sigma_{\eta_{1j}} & \rho_{\tau\eta(22)}\sigma_{\tau(2)}\sigma_{\eta_{2j}} & \rho_{\tau\eta(23)}\sigma_{\tau(2)}\sigma_{\eta_{3j}} & \rho_{\tau(21)}\sigma_{\tau(2)}\sigma_{\tau(1)} & \sigma^2_{\tau(2)} & \rho_{\tau(23)}\sigma_{\tau(2)}\sigma_{\tau(3)} \\ \rho_{\tau\eta(31)}\sigma_{\tau(3)}\sigma_{\eta_{1j}} & \rho_{\tau\eta(32)}\sigma_{\tau(3)}\sigma_{\eta_{2j}} & \rho_{\tau\eta(33)}\sigma_{\tau(3)}\sigma_{\eta_{3j}} & \rho_{\tau(31)}\sigma_{\tau(3)}\sigma_{\tau(1)} & \rho_{\tau(32)}\sigma_{\tau(3)}\sigma_{\tau(2)} & \sigma^2_{\tau(3)} \end{pmatrix}\right]. \tag{9}$$

Table I. Two-level random scale and random location model for negative affect, social isolation, and tired and bored.

	Mean NA			Mean SI			Mean TB		
	$\hat{\beta}_1$	SE	95% CI	$\hat{\beta}_2$	SE	95% CI	$\hat{\beta}_3$	SE	95% CI
Intercept	3.215	(0.115)	(2.994, 3.435)	2.267	(0.111)	(2.090, 2.513)	5.033	(0.136)	(4.775, 5.309)
Smoker	0.456	(0.155)	(0.102, 0.728)	0.310	(0.124)	(0.030, 0.550)	0.291	(0.160)	(−0.061, 0.583)
PropSmk	−0.969	(0.797)	(−2.662, 0.790)	0.136	(0.763)	(−1.502, 1.686)	−0.584	(0.819)	(−2.161, 1.059)
NovSeekC	0.192	(0.072)	(0.0523, 0.337)	0.143	(0.063)	(0.023, 0.278)	0.461	(0.086)	(0.298, 0.630)
NegMoodRegC	−0.684	(0.099)	(−0.867, −0.490)	−0.545	(0.087)	(−0.696, −0.369)	−0.591	(0.097)	(−0.763, −0.388)
Gender	−0.293	(0.107)	(−0.507, −0.090)	−0.011	(0.083)	(−0.168, 0.150)	−0.467	(0.122)	(−0.700, −0.226)
Grade10	0.174	(0.107)	(−0.012, 0.374)	0.081	(0.100)	(−0.103, 0.270)	0.052	(0.108)	(−0.142, 0.298)
AloneBS	0.259	(0.211)	(−0.099, 0.787)	0.306	(0.193)	(−0.033, 0.729)	−0.196	(0.249)	(−0.667, 0.280)
AloneWS	0.175	(0.021)	(0.134, 0.217)	0.197	(0.016)	(0.167, 0.228)	0.401	(0.030)	(0.342, 0.460)
WeekEnd	−0.111	(0.020)	(−0.152, −0.072)	−0.014	(0.012)	(−0.038, 0.011)	−0.454	(0.030)	(−0.513, −0.396)
	WS NA			WS SI			WS TB		
	$\hat{\alpha}_1$	SE	95% CI	$\hat{\alpha}_2$	SE	95% CI	$\hat{\alpha}_3$	SE	95% CI
Intercept	0.712	(0.097)	(0.541, 0.912)	0.166	(0.119)	(−0.070, 0.388)	1.072	(0.082)	(0.920, 1.248)
Smoker	0.455	(0.107)	(0.239, 0.646)	0.369	(0.137)	(0.119, 0.629)	0.148	(0.064)	(0.010, 0.284)
PropSmk	−1.369	(0.520)	(−2.255, −0.348)	−0.597	(0.698)	(−2.057, 0.788)	−0.420	(0.305)	(−1.011, 0.192)
NovSeekC	0.205	(0.061)	(0.087, 0.327)	0.163	(0.070)	(0.024, 0.304)	0.149	(0.041)	(0.067, 0.227)
NegMoodRegC	−0.243	(0.069)	(−0.395, −0.111)	−0.499	(0.094)	(−0.682, −0.314)	−0.044	(0.041)	(−0.128, 0.032)
Gender	−0.305	(0.079)	(−0.450, −0.127)	−0.089	(0.098)	(−0.276, 0.126)	−0.030	(0.055)	(−0.138, 0.077)
Grade10	−0.046	(0.077)	(−0.194, 0.103)	−0.116	(0.105)	(−0.319, 0.096)	−0.106	(0.054)	(−0.216, 0.001)
AloneBS	0.081	(0.186)	(−0.335, 0.412)	0.171	(0.199)	(−0.150, 0.570)	−0.010	(0.141)	(−0.310, 0.233)
AloneWS	0.057	(0.028)	(0.002, 0.114)	0.457	(0.029)	(0.401, 0.515)	−0.210	(0.028)	(−0.267, −0.156)
WeekEnd	0.020	(0.032)	(−0.040, 0.084)	0.016	(0.031)	(−0.045, 0.078)	−0.117	(0.030)	(−0.178, −0.059)

Table I. Continue.

	BS NA			BS SI			BS TB		
	$\hat{\lambda}_1$	SE	95% CI	$\hat{\lambda}_2$	SE	95% CI	$\hat{\lambda}_3$	SE	95% CI
Intercept	0.306	(0.154)	(−0.023, 0.602)	−0.428	(0.169)	(−0.743, −0.108)	0.839	(0.192)	(0.482, 1.240)
Smoker	0.186	(0.150)	(−0.120, 0.485)	0.507	(0.155)	(0.204, 0.825)	−0.192	(0.173)	(−0.533, 0.146)
PropSmk	−0.830	(0.651)	(−2.131, 0.390)	−0.693	(0.717)	(−2.107, 0.692)	0.433	(0.805)	(−1.090, 1.966)
NovSeekC	−0.064	(0.084)	(−0.225, 0.096)	0.096	(0.090)	(−0.078, 0.277)	0.061	(0.100)	(−0.137, 0.251)
NegMoodRegC	−0.310	(0.089)	(−0.477, −0.133)	−0.479	(0.092)	(−0.666, −0.298)	−0.138	(0.112)	(−0.365, 0.085)
Gender	−0.471	(0.111)	(−0.682, −0.247)	−0.339	(0.125)	(−0.570, −0.090)	−0.292	(0.138)	(−0.563, −0.025)
Grade10	0.023	(0.113)	(−0.202, 0.247)	0.003	(0.114)	(−0.228, 0.220)	−0.126	(0.133)	(−0.380, 0.134)
AloneBS	0.838	(0.233)	(0.382, 1.295)	1.098	(0.277)	(0.559, 1.618)	0.066	(0.304)	(−0.554, 0.659)
Scale variance	$\hat{\xi}$	SE	95% CI						
$\log(\sigma^2_{\tau(1)})$	−0.196	(0.074)	(−0.351, −0.058)						
$\log(\sigma^2_{\tau(2)})$	0.369	(0.073)	(0.227, 0.511)						
$\log(\sigma^2_{\tau(3)})$	−1.372	(0.085)	(−1.538, −1.211)						
Correlations	$\hat{\rho}$	SE	95% CI						
$\rho_{\eta(12)}$	0.852	(0.016)	(0.818, 0.882)						
$\rho_{\eta(13)}$	0.700	(0.027)	(0.645, 0.750)						
$\rho_{\eta(23)}$	0.482	(0.040)	(0.401, 0.556)						
$\rho_{\eta\tau(11)}$	0.593	(0.033)	(0.526, 0.655)						
$\rho_{\eta\tau(12)}$	−0.030	(0.055)	(−0.137, 0.080)						
$\rho_{\eta\tau(13)}$	0.570	(0.034)	(0.504, 0.633)						
$\rho_{\eta\tau(21)}$	0.336	(0.041)	(0.254, 0.417)						
$\rho_{\eta\tau(22)}$	0.693	(0.027)	(0.637, 0.745)						
$\rho_{\eta\tau(23)}$	−0.047	(0.053)	(−0.152, 0.058)						
$\rho_{\eta\tau(31)}$	0.350	(0.044)	(0.259, 0.434)						
$\rho_{\eta\tau(32)}$	0.424	(0.041)	(0.342, 0.501)						
$\rho_{\eta\tau(33)}$	0.065	(0.055)	(−0.041, 0.171)						
$\rho_{\tau(12)}$	0.679	(0.028)	(0.622, 0.730)						
$\rho_{\tau(13)}$	0.597	(0.039)	(0.514, 0.669)						
$\rho_{\tau(23)}$	0.385	(0.048)	(0.287, 0.475)						

The results for the three-level model (accounting for the nesting of observations within days) are presented in Table II.

WinBUGS 1.4 software provides Deviance Information Criterion (DIC) based on the conditional likelihood. Spiegelhalter *et al.* [54], Ntzoufras [55, Chapter 11], and Lesaffre and Lawson [56, Chapter 10] have extensively discussed differences between conditional and marginal inference for hierarchical models to explain the importance of DIC based on the conditional likelihood for the model comparison. However, Trevisani and Gelfand [57] have shown that, a posteriori, the first-stage likelihood is expected to be bigger than the marginal likelihood for any given model and discussed the impact of this relationship on model comparison. Celeux *et al.* [58] have further investigated the problem with DIC based on the conditional likelihood for the latent variable models and have discussed various alternatives. Computing DIC using the marginal likelihood entails computing the marginal likelihood, which necessitates additional sampling of parameters and random effects at the posterior ordinates [59, 60]. There are other methods in the literature that rely on samples from prior, posterior, or proposal distribution, but these estimates are not stable and can have infinite variance. The approaches presented by Chib [59] and Chib and Jeliazkov [60] are not readily available in WinBUGS 1.4 software. Li *et al.*[61] have proposed Robust DIC (RDIC), which is the first-order approximation for DIC based on marginal likelihood. The expression for RDIC is as follows:

$$\text{RDIC} = D(\bar{\theta}) + 2\text{tr}\left(I(\bar{\theta})V(\bar{\theta})\right), \tag{10}$$

where $D(\bar{\theta}) = -2\ln p(y|\bar{\theta})$ is defined in terms of log of marginal likelihood after integrating out the random effects with the parameters set to the posterior mode ($\bar{\theta}$), $V(\bar{\theta}) = E[(\theta - \bar{\theta})(\theta - \bar{\theta})^T|y]$ is the posterior variance of the parameters, and $I(\bar{\theta})$ is the Fisher information matrix (FIM) based on observed data again after integrating out the random effects. We implemented an approximate procedure to calculate RDIC in R software version 3.1.0 using Gaussian quadratures (GQ) with five quadrature points for each random effect. Note that for the two-level model we have six random effects and for the three-level model we have nine random effects. Additionally, the computation and memory requirements are huge even with just five quadrature points for the model with random effects > 5. An alternative approach based on the EM algorithm for computing term 1 in Equation (10) is discussed by Li *et al.* [61]; however, this approach also requires additional sampling of random effects at the posterior ordinates of parameters. Hence, we have used the same level of accuracy by restricting to five quadrature points for both the two-level and three-level models. We computed the FIM in the expression of RDIC using the samples from posterior distribution of random effects. The derivatives of the log of marginal likelihood with respect to each parameter were approximated using finite difference methods as GQ is not a practical approach for computing FIM based on observed likelihood. For the two-level models, we obtained an approximate value of RDIC = 353,462 for the model containing no covariates, whereas for the model that included all the covariates in the model, we obtained RDIC = 328,483. Similarly, in the case of the three-level models, we obtained RDIC = 365,212 for the model without any covariates and RDIC = 338,675 for the model with all the covariates. Clearly, the covariates in both the two-level and three-level models respectively improved the overall fit. This example is only constructed for illustration; hence, we have not carefully selected the fixed-effects parameters in both the models.

Comparing the two-level and three-level models in Tables I and II, respectively, we note that the fixed-effects parameter estimates are very similar across the two models, but the three-level model also allows covariates to characterize the within-subject within-day (WS-WD) and within-subject between-day (WS-BD) variability. Furthermore, the additional random effects at level-2 allow us to describe the correlations between the multiple outcomes at the day level ($\rho_{\pi(12)}$, $\rho_{\pi(13)}$, and $\rho_{\pi(23)}$), while the variance–covariance matrix of random location and scale captures the relationships between these outcomes at the subject level. Therefore, the three-level model clearly takes into account the additional nesting of observations and hence correctly captures the correlations within and between observations.

Inspecting the results in Table II, we find that being alone (AloneWS) at the time of signal has a negative effect on the scales of NA, SI, and TB, while having better negative mood regulation (NegMoodRegC) has a beneficial mood effect (lower NA, SI, and TB means) in terms of mean levels. The corresponding AloneWS estimates are 0.166 (0.021), 0.202 (0.016), and 0.339 (0.028) for NA, SI, and TB, respectively. Similarly, the NegMoodRegC estimates are −0.717 (0.057), −0.558 (0.057), and −0.615 (0.073) for NA, SI, and TB, respectively. Being a loner (AloneBS, with an estimate of 0.707 (0.188)) also has a negative effect on the scale for NA. Weekend has beneficial mood effect on NA (−0.190 (0.037)) and SI (−0.485 (0.040)) means. Smokers have worse mood (higher NA mean, with an estimate of 0.300 (0.118)) in comparison with nonsmokers, while men have significantly better mood (lower NA and TB means, with

Table II. Three-level random scale and random location model for negative affect, social isolation, and tired and bored.

	Mean NA			Mean SI			Mean TB		
	$\hat{\beta}_1$	SE	95% CI	$\hat{\beta}_2$	SE	95% CI	$\hat{\beta}_3$	SE	95% CI
Intercept	3.198	(0.132)	(2.938, 3.422)	2.657	(0.078)	(2.496, 2.800)	4.921	(0.094)	(4.711, 5.083)
Smoker	0.300	(0.118)	(0.044, 0.525)	0.095	(0.113)	(−0.092, 0.307)	0.223	(0.138)	(−0.047, 0.499)
PropSmk	−0.975	(0.423)	(−1.833, −0.191)	0.060	(0.448)	(−0.705, 0.880)	−0.664	(0.626)	(−1.881, 0.727)
NovSeekC	0.223	(0.060)	(0.096, 0.329)	0.183	(0.050)	(0.080, 0.276)	0.503	(0.076)	(0.353, 0.650)
NegMoodRegC	−0.717	(0.057)	(−0.828, −0.606)	−0.558	(0.057)	(−0.670, −0.468)	−0.615	(0.073)	(−0.754, −0.464)
Gender	−0.364	(0.086)	(−0.527, −0.209)	−0.120	(0.068)	(−0.251, 0.023)	−0.535	(0.097)	(−0.726, −0.345)
Grade10	0.035	(0.114)	(−0.181, 0.207)	−0.155	(0.058)	(−0.253, −0.029)	−0.013	(0.118)	(−0.248, 0.197)
AloneBS	0.707	(0.188)	(0.417, 1.098)	0.128	(0.100)	(−0.103, 0.293)	0.246	(0.168)	(−0.116, 0.550)
AloneWS	0.166	(0.021)	(0.127, 0.207)	0.202	(0.016)	(0.171, 0.232)	0.339	(0.028)	(0.284, 0.396)
WeekEnd	−0.190	(0.037)	(−0.260, −0.116)	−0.016	(0.025)	(−0.063, 0.033)	−0.485	(0.040)	(−0.564, −0.407)
	WS-WD NA			WS-WD SI			WS-WD TB		
	$\hat{\alpha}_1$	SE	95% CI	$\hat{\alpha}_2$	SE	95% CI	$\hat{\alpha}_3$	SE	95% CI
Intercept	0.292	(0.171)	(−0.050, 0.616)	0.124	(0.128)	(−0.134, 0.347)	0.870	(0.111)	(0.636, 1.056)
Smoker	0.402	(0.107)	(0.222, 0.650)	0.222	(0.139)	(−0.022, 0.510)	0.120	(0.069)	(−0.012, 0.253)
PropSmk	−1.356	(0.478)	(−2.279, −0.431)	−0.558	(0.557)	(−1.616, 0.459)	−0.284	(0.333)	(−0.968, 0.343)
NovSeekC	0.189	(0.067)	(0.064, 0.329)	0.185	(0.072)	(0.041, 0.331)	0.154	(0.045)	(0.067, 0.243)
NegMoodRegC	−0.185	(0.067)	(−0.311, −0.044)	−0.472	(0.079)	(−0.624, −0.286)	−0.007	(0.043)	(−0.087, 0.083)
Gender	−0.370	(0.102)	(−0.568, −0.155)	−0.224	(0.118)	(−0.442, −0.001)	−0.031	(0.064)	(−0.146, 0.099)
Grade10	−0.019	(0.100)	(−0.214, 0.181)	−0.198	(0.105)	(−0.400, −0.017)	−0.069	(0.062)	(−0.184, 0.049)
AloneBS	0.470	(0.273)	(−0.113, 0.941)	0.255	(0.176)	(−0.168, 0.538)	−0.019	(0.157)	(−0.298, 0.316)
AloneWS	0.062	(0.031)	(0.001, 0.121)	0.417	(0.031)	(0.354, 0.477)	−0.229	(0.030)	(−0.288, −0.169)
WeekEnd	0.013	(0.033)	(−0.051, 0.078)	0.028	(0.033)	(−0.038, 0.091)	−0.093	(0.032)	(−0.155, −0.032)

Table II. Continue.

	WS-BD NA			WS-BD SI			WS-BD TB		
	$\hat{\gamma}_1$	SE	95% CI	$\hat{\gamma}_2$	SE	95% CI	$\hat{\gamma}_3$	SE	95% CI
Intercept	−0.003	(0.249)	(−0.353, 0.631)	−1.383	(0.283)	(−1.806, −0.615)	−0.292	(0.176)	(−0.594, 0.066)
Smoker	0.237	(0.116)	(0.016, 0.426)	0.245	(0.155)	(−0.009, 0.606)	0.191	(0.138)	(−0.054, 0.484)
PropSmk	−1.259	(0.566)	(−2.220, −0.040)	−1.823	(0.834)	(−3.643, −0.358)	−1.166	(0.737)	(−2.722, 0.197)
NovSeekC	0.225	(0.080)	(0.071, 0.377)	0.011	(0.101)	(−0.196, 0.217)	0.036	(0.082)	(−0.127, 0.197)
NegMoodRegC	−0.521	(0.075)	(−0.653, −0.362)	−0.746	(0.116)	(−0.965, −0.519)	−0.164	(0.085)	(−0.331, 0.001)
Gender	−0.327	(0.126)	(−0.546, −0.082)	−0.126	(0.130)	(−0.398, 0.107)	−0.132	(0.113)	(−0.366, 0.091)
Grade10	−0.114	(0.116)	(−0.392, 0.078)	−0.335	(0.108)	(−0.565, −0.149)	−0.110	(0.121)	(−0.328, 0.126)
AloneBS	−0.642	(0.317)	(−1.485, −0.204)	0.265	(0.468)	(−0.730, 1.105)	−0.280	(0.288)	(−0.869, 0.244)
WeekEnd	−0.088	(0.109)	(−0.332, 0.108)	0.057	(0.151)	(−0.312, 0.304)	−0.226	(0.147)	(−0.520, 0.053)
	BS NA			BS SI			BS TB		
	$\hat{\lambda}_1$	SE	95% CI	$\hat{\lambda}_2$	SE	95% CI	$\hat{\lambda}_3$	SE	95% CI
Intercept	0.198	(0.182)	(−0.162, 0.547)	−0.158	(0.188)	(−0.534, 0.207)	0.755	(0.192)	(0.400, 1.145)
Smoker	0.210	(0.148)	(−0.062, 0.492)	0.453	(0.152)	(0.130, 0.732)	−0.139	(0.171)	(−0.473, 0.195)
PropSmk	−1.002	(0.650)	(−2.235, 0.285)	−0.814	(0.638)	(−2.162, 0.449)	0.117	(0.766)	(−1.335, 1.598)
NovSeekC	−0.092	(0.090)	(−0.273, 0.088)	0.094	(0.085)	(−0.080, 0.253)	0.075	(0.105)	(−0.137, 0.283)
NegMoodRegC	−0.290	(0.089)	(−0.464, −0.119)	−0.399	(0.095)	(−0.588, −0.213)	−0.166	(0.114)	(−0.389, 0.065)
Gender	−0.444	(0.117)	(−0.687, −0.219)	−0.445	(0.125)	(−0.697, −0.211)	−0.276	(0.139)	(−0.550, −0.008)
Grade10	−0.027	(0.108)	(−0.233, 0.195)	−0.123	(0.111)	(−0.337, 0.096)	−0.129	(0.135)	(−0.403, 0.136)
AloneBS	0.932	(0.272)	(0.416, 1.528)	0.920	(0.268)	(0.430, 1.509)	0.088	(0.308)	(−0.534, 0.666)

Table II. Continue.

Scale variance	$\hat{\zeta}$	SE	95% CI
$\log(\sigma^2_{\tau(1)})$	−0.172	(0.073)	(−0.314, −0.026)
$\log(\sigma^2_{\tau(2)})$	0.369	(0.070)	(0.233, 0.505)
$\log(\sigma^2_{\tau(3)})$	−1.259	(0.087)	(−1.425, −1.083)
Correlations	$\hat{\rho}$	SE	95% CI
$\rho_{\eta(12)}$	0.825	(0.021)	(0.779, 0.863)
$\rho_{\eta(13)}$	0.685	(0.029)	(0.625, 0.740)
$\rho_{\eta\tau(11)}$	0.441	(0.043)	(0.355, 0.522)
$\rho_{\eta\tau(12)}$	0.590	(0.034)	(0.518, 0.654)
$\rho_{\eta\tau(13)}$	−0.066	(0.055)	(−0.174, 0.043)
$\rho_{\eta(23)}$	0.540	(0.037)	(0.465, 0.610)
$\rho_{\eta\tau(21)}$	0.278	(0.044)	(0.193, 0.363)
$\rho_{\eta\tau(22)}$	0.678	(0.028)	(0.619, 0.731)
$\rho_{\eta\tau(23)}$	−0.090	(0.054)	(−0.196, 0.015)
$\rho_{\eta\tau(31)}$	0.333	(0.046)	(0.243, 0.419)
$\rho_{\eta\tau(32)}$	0.423	(0.043)	(0.335, 0.504)
$\rho_{\eta\tau(33)}$	0.068	(0.054)	(−0.037, 0.175)
$\rho_{\tau(12)}$	0.653	(0.030)	(0.592, 0.707)
$\rho_{\tau(13)}$	0.609	(0.039)	(0.529, 0.683)
$\rho_{\tau(23)}$	0.390	(0.048)	(0.294, 0.480)
$\rho_{\pi(12)}$	0.962	(0.010)	(0.941, 0.981)
$\rho_{\pi(13)}$	0.805	(0.020)	(0.766, 0.843)
$\rho_{\pi(23)}$	0.747	(0.022)	(0.700, 0.790)

estimates of −0.120 (0.068) and −0.535 (0.097)) compared with women. Novelty seeking is also found to have a negative effect on the NA (0.223 (0.060)), SI (0.183 (0.050)), and TB (0.503 (0.076)) scales, that is, subjects who have higher novelty seeking scores also tend to have higher NA, SI, and TB mean levels. Finally, higher grade (Grade10, with an estimate of −0.155 (0.058)) also shows improvement in SI mean level.

In terms of BS variation, men have less subject-to-subject mood variation (for NA, SI, and TB, with estimates of −0.444 (0.117), −0.445 (0.125), and −0.276 (0.139)) and behave more homogeneously, while loners (AloneBS) exhibit more variation in NA and SI means (with corresponding estimates of 0.932 (0.272) and 0.920 (0.268)). Subjects who have greater negative mood regulation have less variation in NA (−0.290 (0.089)) and SI (−0.399 (0.095)) means. Finally, smokers have higher variation in SI means (0.453 (0.152)).

For WS-BD variation, smoking results in more variation in NA (0.237 (0.116)) and SI (0.245 (0.155)) means, but this variation tends to be more homogenous for subjects with a high level of smoking (i.e., PropSmk estimates for NA and SI are −1.259 (0.566) and −1.823 (0.834), respectively). Subjects with negative mood regulation vary less and are more consistent in NA (−0.521 (0.075)), SI (−0.746 (0.116)), and TB (−0.164 (0.084)) variability. Again, novelty seekers have high variability in NA (0.225 (0.080)), while being a loner (AloneBS) decreases NA variability (−0.642 (0.109)). Finally, men (−0.327 (0.126)) tend to have lower WS-BD variability in comparison with women, and higher grade subjects tend to have lower SI (−0.335 (0.108)) variability.

For both NA and SI, the WS-WD variation is less for men (−0.370 (0.102) and −0.224 (0.118)) and negative mood regulators (−0.185 (0.067) and −0.472 (0.079)). As before, novelty seekers show a high variability in WS-WD for all the outcomes (NA, SI, and TB with corresponding estimates as 0.189 (0.067), 0.185 (0.072), and 0.154(0.045), respectively). Additionally, among smokers, higher smoking level is associated with less variability in NA (−1.356 (0.478)), but smoking itself increases this variability in NA (0.402 (0.107)). Subjects tend to have lower WS-WD TB variability on weekends (−0.093 (0.032)). Higher grade subjects tend to have lower SI (−0.198 (0.105)) WS-WD variability. Finally, being alone (AloneWS) showed increased WS-WD variation of NA and SI means but decreased WS-WD variation for TB means (0.062 (0.031), 0.417 (0.031) and −0.229 (0.030), respectively).

The random scale variance for NA , SI, and TB are estimated to be $\exp(-0.172) = 0.842$, $\exp(0.369) = 1.446$, and $\exp(-1.259) = 0.284$, respectively. This suggests that error variance varies across individuals, above and beyond the contribution of the many covariates.

Regarding the correlations between random location and scale parameters at level-3, we obtain the positive correlation estimates of 0.441 (0.043) between NA location and NA scale and 0.590 (0.034) between NA location and SI scale. This shows that subjects with higher NA mean (worse mood) fluctuate more in both NA and SI responses. Similarly, we obtain positive correlation estimates of 0.278 (0.044) between SI location and NA scale and 0.678 (0.028) between SI location and SI scale. These correlation estimates indicate that subjects with higher SI mean (better mood) vary more across prompts in both their NA and SI responses. We also obtain the positive correlation estimates of 0.333 (0.046) between TB location and NA scale and 0.423 (0.043) between TB location and SI scale. Alternatively, the positive correlation estimates between location and scale of these outcomes might also suggest a floor effect of measurements.

The significant positive correlation estimates of 0.825 (0.021), 0.685 (0.029), and 0.540 (0.037) between the scales of NA and SI, SI and TB, and NA and TB, respectively, suggest that subjects share a three-way consistent relationship in the variation of NA, SI, and TB. Subjects who have more variation in their NA response also have more variation in their SI and TB responses; subjects who have more variation in their SI response also have more variation in their TB and NA responses and so forth.

The significant positive correlation estimate of 0.653 (0.030) between NA location and SI location suggests that subjects with higher NA mean also have higher SI mean. Similarly, the positive correlation estimate of 0.609 (0.039) between NA location and TB location indicates that subjects with higher NA means also have higher TB means. Finally, the positive correlation of 0.390 (0.048) between SI location and TB location implies that subjects with higher SI means also have higher TB means and vice versa.

The correlations between the random location effects at level-2 between NA, SI, and TB are also highly significant. The estimated correlation between the NA and SI, SI and TB, and NA and TB at level-2 are 0.962 (0.010), 0.805 (0.020), and 0.747 (0.022), respectively, which is consistent with our findings at level-3. Subjects with higher NA mean levels also have higher SI and higher TB mean levels.

In summary, the proposed joint model for the NA, SI, and TB allows us to make inference about the multiple outcomes accounting for variability and correlations at both the location and scale levels. Although not presented, separate models for NA, SI, and TB yielded parameter estimates that were very similar to the joint models. However, the difference between the two is that the 95% CIs were, in general, much smaller for the joint model in comparison with CIs of the separate model for each outcome. Furthermore, the separate models do not allow any inference on how the outcomes behave jointly.

7. Discussion

In this manuscript we have generalized the approach of jointly modeling the multivariate continuous outcomes by inclusion of both random location and scale effects within a Bayesian framework. We have further allowed covariates to explain heterogeneity at various levels in the model. We have used the separation strategy and spherical parametrization for the variances and covariances associated with the location and scale random effects to simplify the specification of vague priors. We have let the variances of the location term and the covariances within and between location and scale terms of the multivariate outcomes to vary across subjects. The method presented in this manuscript can also be generalized for other types of outcomes with the help of appropriate link functions and specification of distribution within exponential family as routinely performed in case of generalized linear mixed-effects models. From the numerical stability standpoint and in order to avoid ill-defined matrices for variance–covariance of random effects, it is necessary to have all multivariate outcomes to be within the same reasonable scale range. Hence, as a first step, we recommend transforming the outcomes using simple shifting and scaling before specifying a joint model for the mutivariate outcomes. As a cautionary note, we do not recommend our approach of prior elicitation in case the primary interest lies in modeling the correlations between multivariate outcomes in terms of covariates. In general, the joint modeling reveals the relationship between the multiple outcomes in terms of the covariates while accounting for the correlations. This leads to efficient inference and provides summaries that are relevant to research questions that may otherwise be left unanswered.

Appendix A. Conditional distribution of fixed-effects parameters, random effects, and spherical parameters

A.1. Conditional distribution of fixed-effects parameters

The conditional distribution of $\boldsymbol{\beta}$ can be derived from the product of two normal distributions,

$$\begin{aligned} &p\left(\boldsymbol{\beta}|\boldsymbol{y}_{jkl},\boldsymbol{X}_{jkl},\boldsymbol{Z}_{jkl},\boldsymbol{\alpha},\boldsymbol{\eta}_j,\boldsymbol{\tau}_j,\boldsymbol{\pi}_{jk}\right)\\ &= N\left(\boldsymbol{\Sigma}_1\sum_{j=1}^{n}\sum_{k=1}^{n_j}\sum_{l=1}^{n_{jk}}\boldsymbol{X}_{jkl}\boldsymbol{\Sigma}(\boldsymbol{\alpha},\boldsymbol{\tau}_j)^{-1}(\boldsymbol{y}_{jkl}-\boldsymbol{\eta}_j-\boldsymbol{\pi}_{jk}),\boldsymbol{\Sigma}_1\right),\\ &\text{where}\ \ \boldsymbol{\Sigma}_1=\left(\sum_{j=1}^{n}\sum_{k=1}^{n_j}\sum_{l=1}^{n_{jk}}\boldsymbol{X}_{jkl}\boldsymbol{\Sigma}(\boldsymbol{\alpha},\boldsymbol{\tau}_j)^{-1}\boldsymbol{X}_{jkl}^T+\left(\sigma_{\beta_0}^2 I_{\mathbf{p}}\right)^{-1}\right)^{-1},\\ &\text{and}\ \ \boldsymbol{\Sigma}(\boldsymbol{\alpha},\boldsymbol{\tau}_j)=Diag\left(\exp\left(\boldsymbol{Z}_{jkl}^T\boldsymbol{\alpha}+\boldsymbol{\tau}_j\right)\right). \end{aligned} \tag{A.1}$$

The conditional distribution of $\boldsymbol{\alpha}$ up to a proportionality constant can be written as

$$\begin{aligned} &p\left(\boldsymbol{\alpha}|\boldsymbol{y}_{jkl},\boldsymbol{X}_{jkl},\boldsymbol{Z}_{jkl},\boldsymbol{\beta},\boldsymbol{\eta}_j,\boldsymbol{\tau}_j,\boldsymbol{\pi}_{jk}\right)\propto\\ &\prod_{j=1}^{n}\prod_{k=1}^{n_j}\prod_{l=1}^{n_{jk}}|\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}tr\left(T_1+\left(\sigma_{\alpha_0}^2 I_{\mathbf{q}}\right)^{-1}\boldsymbol{\alpha}\boldsymbol{\alpha}^T\right)\right),\\ &\text{where}\ \ T_1=\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)^{-1}\left(\boldsymbol{y}_{jkl}-\boldsymbol{X}_{jkl}^T\boldsymbol{\beta}-\boldsymbol{\eta}_j-\boldsymbol{\pi}_{jk}\right)\left(\boldsymbol{y}_{jkl}-\boldsymbol{X}_{jkl}^T\boldsymbol{\beta}-\boldsymbol{\eta}_j-\boldsymbol{\pi}_{jk}\right)^T. \end{aligned} \tag{A.2}$$

Similarly, the conditional distribution of $\boldsymbol{\gamma}$ can be derived as

$$\begin{aligned} &p\left(\boldsymbol{\gamma}|\boldsymbol{\pi}_{jk},\boldsymbol{U}_{jk},\boldsymbol{\psi}\right)\propto\\ &\prod_{j=1}^{n}\prod_{k=1}^{n_j}\left|\boldsymbol{\Sigma}\left(\boldsymbol{U}_{jk},\boldsymbol{\gamma},\boldsymbol{\psi}\right)\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}tr\left(\boldsymbol{\Sigma}\left(\boldsymbol{U}_{jk},\boldsymbol{\gamma},\boldsymbol{\psi}\right)^{-1}\boldsymbol{\pi}_{jk}\boldsymbol{\pi}_{jk}^T+\left(\sigma_{\gamma_0}^2 I_{\mathbf{r}}\right)^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^T\right)\right). \end{aligned} \tag{A.3}$$

Finally, the conditional distributions of λ and $\boldsymbol{\zeta}$ up to a proportionality constant also appear to be very similar,

$$\begin{aligned} &p\left(\lambda|\boldsymbol{\eta}_j,\boldsymbol{\tau}_j,\boldsymbol{V}_j,\boldsymbol{\zeta},\boldsymbol{\omega}\right)\propto\\ &\prod_{j=1}^{n}\left|\boldsymbol{\Sigma}\left(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(tr\left(\boldsymbol{\Sigma}\left(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega}\right)^{-1}\left(\boldsymbol{\eta}_j^T,\boldsymbol{\tau}_j^T\right)\left(\boldsymbol{\eta}_j^T,\boldsymbol{\tau}_j^T\right)^T\right)+tr\left(\left(\sigma_{\lambda_0}^2 I_{\mathbf{s}}\right)^{-1}\lambda\lambda^T\right)\right)\right), \end{aligned} \tag{A.4}$$

$$\begin{aligned} &p\left(\boldsymbol{\zeta}|\boldsymbol{\eta}_j,\boldsymbol{\tau}_j,\boldsymbol{V}_j,\lambda,\boldsymbol{\omega}\right)\propto\\ &\prod_{j=1}^{n}\left|\boldsymbol{\Sigma}\left(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(tr\left(\boldsymbol{\Sigma}\left(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega}\right)^{-1}\left(\boldsymbol{\eta}_j^T,\boldsymbol{\tau}_j^T\right)\left(\boldsymbol{\eta}_j^T,\boldsymbol{\tau}_j^T\right)^T\right)+tr\left(\left(\sigma_{\zeta_0}^2 I_{\mathbf{M}}\right)^{-1}\boldsymbol{\zeta}\boldsymbol{\zeta}^T\right)\right)\right), \end{aligned} \tag{A.5}$$

where $\boldsymbol{\Sigma}(\boldsymbol{U}_{jk},\boldsymbol{\gamma},\boldsymbol{\psi})$ and $\boldsymbol{\Sigma}(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega})$ are defined in Equation (6).

A.2. Conditional distribution of the random effects

The conditional distribution of the random effect $\boldsymbol{\pi}_{jk}$ can be also derived from the product of two normal distributions,

$$
\begin{aligned}
&p\left(\boldsymbol{\pi}_{jk}|\boldsymbol{X}_{jkl},\boldsymbol{Z}_{jkl},\boldsymbol{U}_{jk},\boldsymbol{y}_{jkl},\boldsymbol{\eta}_j,\boldsymbol{\tau}_j,\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\psi}\right)\\
&=N\left(\boldsymbol{\Sigma}_2\sum_{l=1}^{n_{jk}}\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)^{-1}\left(\boldsymbol{y}_{jkl}-\boldsymbol{X}_{jkl}^T\boldsymbol{\beta}-\boldsymbol{\eta}_j\right),\boldsymbol{\Sigma}_2\right),\\
&\text{where }\ \boldsymbol{\Sigma}_2=\left(\sum_{l=1}^{n_{jk}}\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)^{-1}+\left(\boldsymbol{\Sigma}\left(\boldsymbol{U}_{jk},\boldsymbol{\gamma},\boldsymbol{\psi}\right)\right)^{-1}\right)^{-1},\\
&\text{and as before }\ \boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)=Diag\left(\exp\left(\boldsymbol{Z}_{jkl}^T\boldsymbol{\alpha}+\boldsymbol{\tau}_j\right)\right).
\end{aligned}
\tag{A.6}
$$

The variance–covariance matrix of random location and scale effects $\boldsymbol{\Sigma}(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega})$ can be further represented in terms of sub-matrices as $\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\eta\tau}}\\ \boldsymbol{\Sigma}_{\boldsymbol{\tau\eta}} & \boldsymbol{\Sigma}_{\boldsymbol{\tau}}\end{pmatrix}$. Note that the sub-matrices $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\eta\tau}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\tau\eta}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\tau}}$ are functions of their respective parameters λ, $\boldsymbol{\omega}$, and $\boldsymbol{\zeta}$. Hence, the conditional distribution of the random effect $\boldsymbol{\eta}_j$ can be written as

$$
\begin{aligned}
&p\left(\boldsymbol{\eta}_j|\boldsymbol{X}_{jkl},\boldsymbol{Z}_{jkl},\boldsymbol{V}_j,\boldsymbol{y}_{jkl},\boldsymbol{\pi}_{jk},\boldsymbol{\tau}_j,\boldsymbol{\beta},\boldsymbol{\alpha},\lambda,\boldsymbol{\omega},\boldsymbol{\zeta}\right)\propto\\
&N\left(\boldsymbol{\Sigma}_3\sum_{k=1}^{n_j}\sum_{l=1}^{n_{jk}}\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)^{-1}\left(\boldsymbol{y}_{jkl}-\boldsymbol{X}_{jkl}^T\boldsymbol{\beta}-\boldsymbol{\pi}_{jk}\right)+\boldsymbol{\Sigma}_{\boldsymbol{\eta}|\boldsymbol{\tau}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\eta\tau}}\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\boldsymbol{\tau}_j,\boldsymbol{\Sigma}_3\right)^{-1},\\
&\text{where }\ \boldsymbol{\Sigma}_3=\left(\sum_{k=1}^{n_j}\sum_{l=1}^{n_{jk}}\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)^{-1}+\left(\boldsymbol{\Sigma}_{\boldsymbol{\eta}|\boldsymbol{\tau}}\right)^{-1}\right)^{-1}\\
&\text{and }\ \boldsymbol{\Sigma}_{\boldsymbol{\eta}|\boldsymbol{\tau}}=\boldsymbol{\Sigma}_{\boldsymbol{\eta}}-\boldsymbol{\Sigma}_{\boldsymbol{\eta\tau}}\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\tau\eta}},
\end{aligned}
\tag{A.7}
$$

and the conditional distribution of the random effect $\boldsymbol{\tau}_j$ up to a proportionality constant can be derived as

$$
\begin{aligned}
&p\left(\boldsymbol{\tau}_j|\boldsymbol{X}_{jkl},\boldsymbol{Z}_{jkl},\boldsymbol{V}_j,\boldsymbol{y}_{jkl},\boldsymbol{\pi}_{jk},\boldsymbol{\eta}_j,\boldsymbol{\beta},\boldsymbol{\alpha},\lambda,\boldsymbol{\omega},\boldsymbol{\zeta}\right)\propto\\
&\prod_{k=1}^{n_j}\prod_{l=1}^{n_{jk}}\left|\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)\right|^{-\frac{1}{2}}\left|\boldsymbol{\Sigma}_{\boldsymbol{\tau}|\boldsymbol{\eta}}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}tr\left(T_1+\boldsymbol{\Sigma}_{\boldsymbol{\tau}|\boldsymbol{\eta}}^{-1}\left(\boldsymbol{\tau}_j-\boldsymbol{\Sigma}_{\boldsymbol{\tau\eta}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\boldsymbol{\eta}_j\right)\left(\boldsymbol{\tau}_j-\boldsymbol{\Sigma}_{\boldsymbol{\tau\eta}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\boldsymbol{\eta}_j\right)^T\right)\right),\\
&\text{where }\ T_1=\boldsymbol{\Sigma}\left(\boldsymbol{\alpha},\boldsymbol{\tau}_j\right)^{-1}\left(\boldsymbol{y}_{jkl}-\boldsymbol{X}_{jkl}^T\boldsymbol{\beta}-\boldsymbol{\eta}_j-\boldsymbol{\pi}_{jk}\right)\left(\boldsymbol{y}_{jkl}-\boldsymbol{X}_{jkl}^T\boldsymbol{\beta}-\boldsymbol{\eta}_j-\boldsymbol{\pi}_{jk}\right)^T,\\
&\text{and }\ \boldsymbol{\Sigma}_{\boldsymbol{\tau}|\boldsymbol{\eta}}=\boldsymbol{\Sigma}_{\boldsymbol{\tau}}-\boldsymbol{\Sigma}_{\boldsymbol{\tau\eta}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\eta\tau}}.
\end{aligned}
\tag{A.8}
$$

A.3. Conditional distribution of the spherical parameters

The conditional distribution of the spherical parameters $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ can be written up to a proportionality constant as

$$
\begin{aligned}
p\left(\boldsymbol{\psi}|\boldsymbol{\pi}_{jk},\boldsymbol{U}_{jk},\boldsymbol{\gamma}\right)&\propto\frac{1}{\boldsymbol{\pi}}N\left(\boldsymbol{0}_M,\boldsymbol{\Sigma}\left(\boldsymbol{U}_{jk},\boldsymbol{\gamma},\boldsymbol{\psi}\right)\right),\\
p\left(\boldsymbol{\omega},|\boldsymbol{\eta}_j,\boldsymbol{\tau}_j\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta}\right)&\propto\frac{1}{\boldsymbol{\pi}}N\left(\boldsymbol{0}_{2M},\boldsymbol{\Sigma}\left(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega}\right)\right),
\end{aligned}
\tag{A.9}
$$

where $\boldsymbol{\Sigma}(\boldsymbol{U}_{jk},\boldsymbol{\gamma},\boldsymbol{\psi})$ and $\boldsymbol{\Sigma}(\boldsymbol{V}_j,\lambda,\boldsymbol{\zeta},\boldsymbol{\omega})$ are defined in Equation (6).

Acknowledgements

The authors would like to thank the Editor and two anonymous reviewers for their encouraging and valuable comments to help improve the quality of the manuscript.

References

1. McCulloch CE, Searle SR. *Generalized, Linear, and Mixed Models*. Wiley: New York, 2001.
2. Raudenbush SW, Bryk AS. *Hierarchical Linear Models, 2nd Edition*. Sage: Thousand Oaks, CA, 2002.

3. Demidenko E. *Mixed Models: Theory and Applications*. Wiley: Hoboken, New Jersey, 2004.
4. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley: New York, 2004.
5. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman and Hall: Boca Raton, Florida, 2004.
6. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley: Hoboken, New Jersey, 2006.
7. Goldstein H. *Multilevel Statistical Models* 4th edition. Wiley: West Sussex, United Kingdom, 2011.
8. Verbeke G, Davidian M. Joint models for longitudinal data: introduction and overview. In *Handbooks of Modern Statistical Methods: Longitudinal Data Analysis*, Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds). Chapman and Hall: Boca Raton, Florida, 2009; 319–326.
9. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **4**:465–480.
10. Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001; **96**(455):895–905.
11. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**:809–834.
12. Carroll RJ, Ruppert D. *Transformations and Weighting in Regression*. Chapman and Hall: New York, 1988.
13. Amemiya T. *Advanced Econometrics*. Harvard University Press: Cambridge, Massachusetts, 1985.
14. Greene WH. *Econometric Analysis*. Mcmillan: New York, 1990.
15. Cleveland WS, Denby L, Liu C. Random location and scale effects: model building methods for a general class of models. *Computing Science and Statistics* 2002; **32**:3–10.
16. Hedeker D, Mermelstein RJ, Demirtas H. An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics* 2008; **64**:627–634.
17. Li X, Hedeker D. A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine* 2012; **31**(26):3192–3210.
18. Oksana P, Hedeker D, Richmond M, Sokolovsky A, Mermelstein R. Modeling mood variation and covariation among adolescent smokers: application of a bivariate location-scale mixed-effects model. *Nicotine and Tobacco Research* 2014; **16**:S151–S158.
19. Blood EA, Kaliash LA, Shrier LA. Estimating heterogeneous intra-class correlation coefficients in dyadic ecological momentary assessment. *Journal of Modern Applied Statistical Methods* 2013; **12**:207–219.
20. Lee K, Daniels MJ, Joo Y. Flexible marginalized models for bivariate longitudinal ordinal data. *Biostatistics* 2013; **14**(3):462–476.
21. Zhang W, Leng C, Tang CY. A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society:* Series B (Statistical Methodology) 2014. http://dx.doi.org/10.1111/rssb.12065.
22. Boscardin WJ, Gelman A. Bayesian computation for parametric models of heteroscedasticity in the linear model. *Advances in Econometrics* 1994; **11**:A87–A110.
23. Browne WJ. Applying MCMC methods to multi-level models. *Ph.D. Thesis*, University of Bath, Bath, United Kingdom, 1998.
24. Browne WJ, Draper D, Goldstein H, Rasbash J. Bayesian and likelihood methods for fitting multilevel models with complex level 1 variation. *Computational Statistics and Data Analysis* 2002; **39**:203–225.
25. Browne WJ, Subramanian SV, Jones K, Goldstein H. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society:* Series A (Statistics in Society) 2005; **168**(3):599–613.
26. Pourahmadi M, Daniels MJ. Dynamic conditionally linear mixed models for longitudinal data. *Biometrics* 2002; **58**: 225–231.
27. Hoff PD, Niu X. A covariance regression model. *Statistica Sinica* 2012; **22**:729–753.
28. Rast P, Hofer SM, Sparks C. Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research* 2012; **47**:177–200.
29. Congdon P. *Bayesian Statistical Modelling* 2nd edition. John Wiley: West Sussex, England, 2006.
30. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine* 1994; **16**:199–202.
31. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies* 2003; **4**:35–52.
32. de Vries MW. *The Experience of Psychopathology: Investigating Mental Disorders in Their Natural Settings*. Cambridge University Press: New York, 1992.
33. Scollon C N, Kim-Prieto C, Diener E. Experience sampling: promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies* 2003; **4**:5–34.
34. Feldman Barrett L, Barrett D. An introduction to computerized experience sampling in psychology. *Social Science Computer Review* 2001; **19**:175–185.
35. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. *Annual Review of Psychology* 2003; **54**: 579–616.
36. Dierker L, Mermelstein RJ. Early emerging nicotine-dependence symptoms: a signal of propensity for chronic smoking behavior in adolescents. *Journal of Pediatrics* 2010; **156**:818–822.
37. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**:638–645.
38. Barnard J, McCulloch R, Meng X. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 2000; **10**:1281–1311.

39. Lu G, Ades AE. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; **10**(4): 792–805.
40. Pinheiro JC, Bates DM. Unconstrained parametrizations for variance–covariance matrices. *Statistics and Computing* 1996; **6**:289–296.
41. Pinheiro JC. Topics in mixed effects models. *Ph.D. Thesis*, University of Wisconsin-Madison, Madison, Wisconsin, 1994.
42. Geman S, Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; **6**:721–741.
43. Gelfand AE, Smith AFM. Sampling-based approaches to calculation marginal densities. *Journal of the American Statistical Association* 1990; **85**:398–409.
44. Gelfand AE, Hills SE, Racine-Poon A, Smith AFM. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* 1990; **85**:972–985.
45. Casella G, George EI. Explaining the Gibbs sampler. *The American Statistician* 1992; **46**(3):167–174.
46. Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society:* Series B 1995; **57**(3):473–484.
47. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society.* Series C (Applied Statistics) 1992; **41**(2):337–348.
48. Gilks WR, Clayton DG, Spiegelhalter DJ, Best NG, McNeil AJ. Modelling complexity: applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society.* Series B (Methodological) 1993; **55**(1):39–52.
49. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 1953; **21**:1087–1092.
50. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970; **57**(1).
51. Chib S, Greenberg E. Understanding the Metropolis–Hastings algorithm. *The American Statistician* 1995; **49**(4):327–335.
52. Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50*. Technical Report, MRC Biostatistics Unit: Cambridge, UK, 1995.
53. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis* Second. Chapman & Hall/CRC, 2004.
54. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society:* Series B (Statistical Methodology) 2002; **64**(4):583–639.
55. Ntzoufras I. *Bayesian Modeling Using Winbugs*. John Wiley & Sons: Hoboken, NJ, 2009.
56. Lesaffre E, Lawson A. *Bayesian Biostatistics*. John Wiley & Sons: Hoboken, NJ, 2012.
57. Trevisani M, Gelfand AE. Inequalities between expected marginal log-likelihoods, with implications for likelihood-based model complexity and comparison measures. *The Canadian Journal of Statistics* 2003; **31**(3):1–12.
58. Celeux G, Forbes F, Robert CP, Titterington DM. Deviance information criteria for missing data models. *Bayesian Analysis* 2006; **1**(4):651–674.
59. Chib S. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 1995; **90**(432): 1312–1321.
60. Chib S, Jeliazkov I. Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association* 2001; **96**(453):270–281.
61. Li Y, Zheng T, Yu J. Robust deviation information criterion for latent variable models. *Working Paper Series*, Singapore Management University, School of Economics, 2014.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.