Robust Bayesian hierarchical model using normal/independent distributions

Geng Chen[1] and **Sheng Luo***,[2]

[1] Clinical Statistics, GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, PA, 19426, USA
[2] Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, USA

Received 4 December 2014; revised 3 June 2015; accepted 29 July 2015

The multilevel item response theory (MLIRT) models have been increasingly used in longitudinal clinical studies that collect multiple outcomes. The MLIRT models account for all the information from multiple longitudinal outcomes of mixed types (e.g., continuous, binary, and ordinal) and can provide valid inference for the overall treatment effects. However, the continuous outcomes and the random effects in the MLIRT models are often assumed to be normally distributed. The normality assumption can sometimes be unrealistic and thus may produce misleading results. The normal/independent (NI) distributions have been increasingly used to handle the outlier and heavy tail problems in order to produce robust inference. In this article, we developed a Bayesian approach that implemented the NI distributions on both continuous outcomes and random effects in the MLIRT models and discussed different strategies of implementing the NI distributions. Extensive simulation studies were conducted to demonstrate the advantage of our proposed models, which provided parameter estimates with smaller bias and more reasonable coverage probabilities. Our proposed models were applied to a motivating Parkinson's disease study, the DATATOP study, to investigate the effect of deprenyl in slowing down the disease progression.

Keywords: Clinical trial; Item-response theory; Latent variable; MCMC; Outliers.

1 Introduction

Parkinson's disease (PD) is a chronic progressive disease with multidimensional impairments. Symptoms such as tremors, stiffness, slowness of movements, and loss of cognitive function are often observed from PD patients (Cummings, 1992; Fahn et al., 2004). Due to the multidimensional nature of the disease, it is difficult to identify a single outcome to summarize or represent the overall disease severity (Huang et al., 2005). Therefore, clinical trials that search for a neuroprotective treatment for PD patients usually measure multiple outcomes at different visits. Examples of such PD studies include the Deprenyl And Tocopherol Antioxidative Therapy Of Parkinsonism (DATATOP) study (Parkinson Study Group, 1989), Earlier versus Later Levodopa Therapy in Parkinson Disease (ELLDOPA) study (Fahn et al., 2004), and Neuroprotection Exploratory Trials in Parkinson's Disease (NET-PD) study (Elm and The NINDS NET-PD Investigators, 2012). The PD symptoms can be measured by the outcomes of mixed types, for example binary, ordinal, and continuous. Moreover, this type of multivariate longitudinal data structure has three sources of correlation within and between outcomes of the same patient, that is intersource (different measures at the same visit), longitudinal (same measure at different visits), and cross correlation (different measures at different visits). For valid analysis of PD progression, the model needs to account for these three sources of correlation.

*Corresponding author: e-mail: sheng.t.luo@uth.tmc.edu, Phone: +1-713-5009554

To address the challenges of analyzing this type of multivariate longitudinal data, many approaches have been developed, such as a linear combination of all outcomes, choosing one outcome as primary and other outcomes as secondary, and global statistical tests (GST). Among these methods, the implementation of the linear combination of all outcomes is relatively easy. However, it substantially reduces the information and the interpretation may be difficult (Bandyopadhyay et al., 2011). The method of choosing one outcome as primary and other outcomes as secondary may encounter problems when the conclusions from the primary analysis and secondary analysis are quite different. The GST method has some attractive properties such as maintaining high power while controlling the overall type I error (Huang et al., 2005, 2009). However, unless certain assumptions regarding the variances and covariance structure are met, the GST can neither adjust for covariates of interest nor utilize the full longitudinal data information.

An alternative approach is the latent variable model approach that assumes all the outcomes are measurements of some unobservable latent variable (Mungas and Reed, 2000; Wang et al., 2002; Reise and Waller, 2009). To this end, multilevel item response theory (MLIRT) models, which are based on latent variables, have been increasingly used (Douglas, 1999; Glas et al., 2009; Luo et al., 2013; He and Luo, 2013). The MLIRT models consists of two levels. The first level of the MLIRT models describes the outcome measurements as functions of a univariate subject-specific latent variable (denoting disease severity) and measurement-specific parameters, while in the second level, the latent variable is regressed on the covariates of interest and the subject-specific random effects. Advantages of the MLIRT model include: (1) it uses the full longitudinal information and accounts for the three sources of correlations within subjects via the subject-specific random effects; (2) it has a better reflection of the multilevel data structure; and (3) it simultaneously estimates the measurement-specific parameters, the covariate effects, as well as the subject-specific disease progression characteristics (Maier, 2001; Kamata, 2001; He and Luo, 2013). The MLIRT model has been increasingly used in studying many diseases such as PD disability (Weisscher et al., 2010; Luo et al., 2012), Alzheimer's disease (Snitz et al., 2012), Huntington's disease (Vaccarino et al., 2011), and dementia (Miller et al., 2012).

In MLIRT models, normal distributions are usually assumed for continuous outcomes and random effects. However, the parameter estimation may be biased in the presence of outliers and heavy tails in the continuous outcomes and/or random effects. One way to handle the outlier problem is to identify the outliers and exclude them from the analysis. However, the primary efficacy assessments in clinical trial studies are often required to follow the intent-to-treat (ITT) principle (the analysis has to include all randomized individuals). By following ITT, the analysis preserves the benefits of randomization, and it is recommended as the most unbiased approach (Little and Yau, 1996; Lachin, 2000). Thus, the exclusion of outliers is inappropriate under the ITT principle. Data transformation methods (e.g., log, square-root, Box-Cox) might generate distributions close to normality. But the disadvantages include: (1) reduced information on an underlying data generation scheme; (2) reduced interpretability on a transformed scale; and (3) transformations may not be universal and vary with datasets. The outlier issue is further complicated when both continuous outcomes and random effects in the model are subject to "departure from normality" (Lachos et al., 2011). While the transformation for outcomes might be feasible, the transformation for random effects may not be straightforward. To illustrate this, Fig. 1 displays the density histograms and the Q–Q plots of the residuals of a continuous outcome Unified Parkinson's Disease Rating Scale (UPDRS) (left panels, column a), subject-specific random intercepts (center panels, column b), and subject-specific random slopes (right panels, column c) by fitting the MLIRT model (5) with normal assumptions to the motivating DATATOP study. The plots suggest the presence of outliers in the continuous outcome UPDRS, the random intercepts, and the random slopes, as manifested by the large departure from the diagonal line in the tail areas. While McCulloch and Neuhaus (2011) concluded that the misspecification of the random effects distribution does not severely affect the parameter estimation in the generalized linear-mixed models (GLMM), its influence in the MLIRT modeling framework is unclear and how it interacts with the outlying outcome measurements requires further investigation.

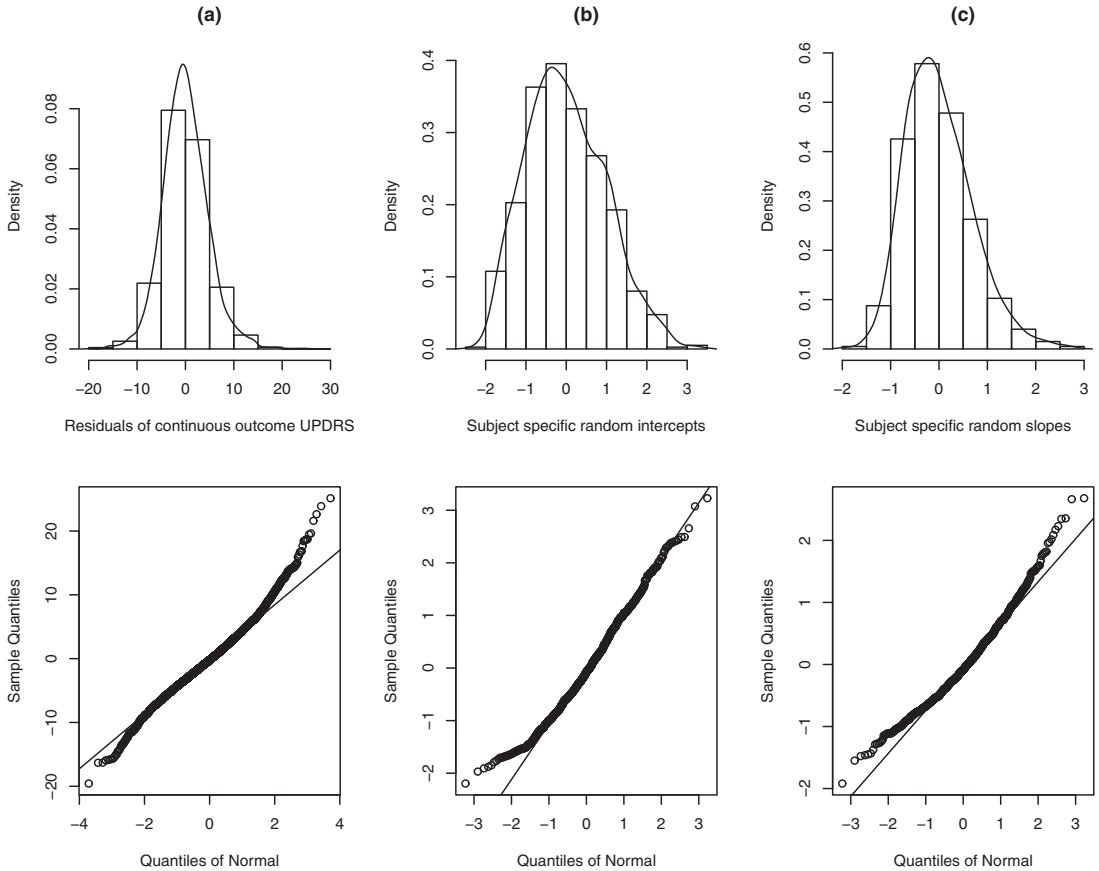

Figure 1 Histogram and normal Q-Q plot of residuals of UPDRS (column a), subject-specific random intercepts (column b), and subject-specific random slopes (column c) obtained by fitting the MLIRT model (5) with normal assumptions.

To address the issue of nonnormality due to outliers and heavy-tails, one solution is to replace the normality assumption by the more robust normal/independent (NI) distributions. The NI distributions are an attractive class of symmetric heavy-tailed densities that include normal, Student's t, slash, and contaminated normal distributions as special cases. The NI distributions have been applied to multivariate linear regression (Liu, 1996), linear mixed effect (LME) model (Rosa et al., 2003; Lin and Lee, 2007), nonlinear mixed effect model (Lachos et al., 2013), linear-mixed effect model with censored data (LMEC), and nonlinear-mixed effect model with censored data (NLMEC) (Lachos et al., 2011). Specifically, Lachos et al. (2011, 2013) used a NI distribution with a shared weight variable ω on both the continuous outcome and random effects that were nonnormal due to outliers and heavy tails. In their methods, when conditional on the shared weight, the distributions of the continuous outcome and random effects are independent (Lachos et al., 2011). However, the continuous outcome and random effects may have different scales of nonnormality so that the strong assumption of sharing the same weight variable may be unreasonable in practice. Alternatively, a more flexible approach to be developed in this article is to assume different NI distributions for the continuous outcome and random effects. To the best of our knowledge, this is no studies on Bayesian MLIRT models using the NI distributions for both the continuous outcome and random effects. In this article, we propose

robust Bayesian parametric MLIRT models using the NI distributions. We then apply our methods to a motivating DATATOP study on PD.

The rest of the article proceeds as follow. In Sections 2, we discuss the MLIRT models, the NI distributions, likelihood formulation, Bayesian inference, and model selection criterion. Section 3 presents an extensive simulation study comparing the performance of various models. In Section 4, we apply the proposed models to the DATATOP study dataset. Section 5 summarizes the main findings and discusses the possible directions in our future research.

2 Model and estimation

2.1 The multilevel item response theory (MLIRT) model

We first introduce the MLIRT model. Let y_{ijk} be the observed outcome k ($k = 1, \ldots, K$) for patient i ($i = 1, \ldots, N$) at visit j ($j = 1, \ldots, J_i$), where $j = 1$ is baseline. Let $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijk}, \ldots, y_{ijK})'$ be the vector of observation for patient i at visit j and let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iK})'$ be the outcome vector across visits. Let θ_{ij} be the continuous latent variable denoting the unobserved disease severity for patient i at visit j with higher value representing more severe disease status. In the first level measurement model, the observed measurements are viewed as imperfect clinical manifestations of the interaction between a univariate subject-specific latent disease severity and the measurement-specific parameters (e.g., the measurements' ability to distinguish PD patients in disease severity). Specifically, we model the binary outcomes using a two-parameter submodel (Fox, 2010), the cumulative probabilities of ordinal outcomes using a graded response submodel (Samejima, 1997), and the continues outcomes using a common factor submodel (Lord et al., 1968).

$$\text{logit}\{\text{p}(y_{ijk} = 1|\theta_{ij})\} = a_k + b_k\theta_{ij}, \tag{1}$$

$$\text{logit}\{\text{p}(y_{ijk} \le l|\theta_{ij})\} = a_{kl} - b_k\theta_{ij}, \quad \text{with } l = 1, 2, \ldots, n_k - 1, \tag{2}$$

$$y_{ijk} = a_k + b_k\theta_{ij} + \epsilon_{ijk}, \tag{3}$$

where random error $\epsilon_{ijk} \sim \text{N}(0, \sigma_k^2)$ with σ_k^2 being variance of continuous outcome k, a_k is the outcome-specific "difficulty" parameter and b_k is the outcome-specific "discriminating" parameter that is always positive and describes the ability that outcome k discriminates between patients with latent disease severity θ_{ij}. Suppose the ordinal outcome k in model (2) has n_k categories and $n_k - 1$ thresholds $a_{k1}, \ldots, a_{kl}, \ldots, a_{kn_{k-1}}$ that satisfy the order constraint $a_{k1} < \cdots < a_{kl} < \ldots < a_{kn_{k-1}}$. The probability of patient i being in category l on outcome k at visit j is $\text{p}(y_{ijk} = l|\theta_{ij}) = \text{p}(y_{ijk} \le l|\theta_{ij}) - \text{p}(y_{ijk} \le l - 1|\theta_{ij})$.

In the second level structural model, the latent disease severity θ_{ij} is regressed on predictors of interest (e.g. treatment, disease duration, and time) and subject-specific random effects.

$$\theta_{ij} = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0 + u_{i0} + (\boldsymbol{X}_{i1}\boldsymbol{\beta}_1 + u_{i1})t_{ij}, \tag{4}$$

where $\boldsymbol{X}_{i0}$ and $\boldsymbol{X}_{i1}$ are the covariate vectors that may share all or part of the covariates, u_{i0} is the random intercept that determines the subject-specific disease severity and u_{i1} is the random slope that determines the subject-specific disease progression rate. We now give an example to further illustrate model (4). If no covariate is in $\boldsymbol{X}_{i0}$ and only the treatment assignment variable is included in $\boldsymbol{X}_{i1}$, $\theta_{ij} = u_{i0} + [\beta_{10} + \beta_{11}I_i(\text{trt}) + u_{i1}]t_{ij}$, where $I(\cdot)$ is an indicator function (1 if treatment and 0 otherwise). In this model, β_{10} and $\beta_{10} + \beta_{11}$ denote the disease progression rates for placebo and

treatment patients, respectively, with β_{11} being the change in disease progression rate introduced by the treatment. The significant negative coefficient β_{11} indicates that the treatment slows down the disease progression. In this context, the null hypothesis of no overall treatment effect is $H_0 : \beta_{11} = 0$. We let the random effects vector $\boldsymbol{u}_i = (u_{i0}, u_{i1})'$ and assume $u_{i0} \sim \mathrm{N}(0, 1)$, $u_{i1} \sim \mathrm{N}(0, \sigma_u^2)$, and $\mathrm{corr}(u_{i0}, u_{i1}) = \rho$. It is well-known that the item-response models are overparameterized (Samejima, 1997) and some constraints need to be imposed to make the models identifiable. To this end, we set $\mathrm{Var}(u_{i0}) = 1$ to ensure model identifiability. Under the local independence assumption (i.e., conditional on the random effects vector $\boldsymbol{u}_i$, all outcome measures for each patient are independent) (Fox, 2010), the full likelihood of patient i across all visits is

$$\mathrm{L}(\boldsymbol{y}_i, \boldsymbol{u}_i) = \left[\prod_{j=1}^{J_i} \prod_{k=1}^{K} \mathrm{p}(y_{ijk}|\boldsymbol{u}_i) \right] \mathrm{p}(\boldsymbol{u}_i). \tag{5}$$

For notational ease, we let the difficulty parameter vector be $\boldsymbol{a} = (\boldsymbol{a}_1', \ldots, \boldsymbol{a}_k', \ldots, \boldsymbol{a}_K')'$, with $\boldsymbol{a}_k' = (a_{k1}, \ldots, a_{kn_k-1})$ for ordinal outcomes. Let the discrimination vector be $\boldsymbol{b} = (b_1, \ldots, b_K)'$, and $\boldsymbol{\beta} = (\boldsymbol{\beta_0}', \tilde{\boldsymbol{\beta}}_{\mathbf{1}}')'$. Because we assume normal distributions for the random errors ϵ_{ijk} for the continuous outcome in model (3) and the random effects vector $\boldsymbol{u}_i$, in addition to their independence, we refer to this model as model Indep-N with the parameter vector $\boldsymbol{\Phi} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k)'$.

2.2 The normal/independent (NI) distribution family

The normal/independent (NI) distribution is a family of symmetric distributions with heavier tails. Extensive discussion about the NI distributions can be found in the literature (e.g. Lange and Sinsheimer, 1993; Liu, 1996; Rosa et al., 2003; Lachos et al., 2011; Luo et al., 2013; Baghfalaki et al., 2013). An element of the univariate NI family is defined as the distribution of random variable $y = \mu + e/\sqrt{\omega}$, where μ is a location vector, random error e is normally distributed with mean zero and variance σ^2, ω is a positive weight variable that has density function $\mathrm{p}(\omega|\nu)$ with tuning parameter ν and is independent of random error e. The NI distribution controls the impact of outliers on the overall inference by stochastically assigning lower weights ω for potential outliers or influencing points (Lange and Sinsheimer, 1993). In practice, the weight variable ω can be estimated and be used for outlier detection. Specifically, if the posterior distribution of ω has high density close to 0 (or if the estimate of ω is close to 0), it indicates that the corresponding observation can be a potential outlier (Rosa et al., 2003). Given ω, y follows a normal distribution $\mathrm{N}(\mu, \omega^{-1}\sigma^2)$ with the marginal pdf of y given by $\mathrm{NI}(y|\mu, \sigma^2, \nu) = \int \mathrm{p}(y|\mu, \sigma^2, \omega)\mathrm{p}(\omega|\nu)d\omega$. When $\omega = 1$ (or equivalently when $\nu \to \infty$), $\mathrm{NI}(y|\mu, \sigma^2, \nu)$ becomes a normal distribution (Lange and Sinsheimer, 1993; Rosa et al., 2003).

The NI distributions provide a family of symmetric heavy-tailed distributions with various specifications of the density function $\mathrm{p}(\omega|\nu)$. We consider the continuous outcome y_{ijk} in model (3) as an example. When a univariate NI distribution is applied to model (3), we have $y_{ijk} = a_k + b_k\theta_{ij} + \epsilon_{ijk}'$ where $\epsilon_{ijk}' = \epsilon_{ijk}/\sqrt{\omega_i}$ with $\epsilon_{ijk} \sim \mathrm{N}(0, \sigma_k^2)$. The weight variable ω_i has density function $\mathrm{p}(\omega_i|\nu)$ with positive tuning parameter ν. When $\omega_i \sim$ Gamma$(\nu/2, \nu/2)$, ϵ_{ijk}' follows a Student's t distribution with parameter ν being the degree of freedom and when $\omega_i \sim \mathrm{Beta}(\nu, 1)$, ϵ_{ijk}' follows a slash distribution with tuning parameter ν. In addition, ϵ_{ijk}' follows a contaminated normal (CN) distribution when ω_i takes one of the two discrete values with pdf $\mathrm{p}(\omega_i|\nu) = \nu I_{(\omega_i=\gamma)} + (1-\nu)I_{(\omega_i=1)}$, where ν $(0 < \nu \leq 1)$ is the proportion of contamination (or the percentage of outliers deviating from the normal distribution) and γ $(0 < \gamma \leq 1)$ is the scale of contamination (how severe the outliers deviate from the normal distribution with smaller γ denoting more deviation). When $\omega_i = 1$, then $\mathrm{p}(\omega_i|\nu) = 1 - \nu$ and $\epsilon_{ijk}' \sim \mathrm{N}(0, \sigma_k^2)$ with probability $1 - \nu$; when $\omega_i = \gamma$, then $\mathrm{p}(\omega_i|\nu) = \nu$ and ϵ_{ijk}' is contaminated with probability ν and

$\epsilon_{ijk}^{'} \sim \mathrm{N}(0, \sigma_k^2/\gamma)$ (Lange and Sinsheimer, 1993; Rosa et al., 2003). Therefore, the CN distribution $\epsilon_{ijk}^{'}$ follows a two-component mixture distribution with pdf $p(\epsilon_{ijk}^{'}) = \nu\mathrm{N}(0, \sigma_k^2/\gamma) + (1-\nu)\mathrm{N}(0, \sigma_k^2)$.

2.3 The NI distributions in the MLIRT model

In this section, we apply the NI distributions to the random error ϵ_{ijk} in model (3) and the random effects vector $\boldsymbol{u}_i = (u_{i0}, u_{i1})^{'}$ in model (4). We first discuss the method in Lachos et al. (2011, 2013) used in linear and nonlinear-mixed models. Lachos et al. (2011, 2013) assume that $(\boldsymbol{u}_i, \epsilon_i) \sim NI(\mathbf{0}, \{(\boldsymbol{\Sigma}_u, 0), (0, \sigma^2)\}, \omega_i)$, where ϵ_i is the random error of a continuous outcome for patient i, where $i = 1, ..., N$. The fact that both $\boldsymbol{u}_i$ and ϵ_i are scaled by the same weight variable ω_i allows conditional independence between $\boldsymbol{u}_i$ and ϵ_i, given ω_i, but not marginal independence. To use their methods in the MLIRT model, we specify $\boldsymbol{u}_i$, ϵ_{ijk} and ω_i hierarchically as

$$\begin{aligned}
\boldsymbol{u}_i|\omega_i &\sim \mathrm{N}\left(\mathbf{0}, \omega_i^{-1}\Sigma_u\right), \\
\epsilon_{ijk}|\omega_i &\sim \mathrm{N}\left(\mathbf{0}, \omega_i^{-1}\sigma_k^2\right), \\
\omega_i &\sim \mathrm{p}(\omega_i|\nu).
\end{aligned}$$

Then the continuous outcome y_{ijk} follows $y_{ijk}|\boldsymbol{u}_i, \omega_i \sim \mathrm{N}(a_k + b_k\theta_{ij}, \omega_i^{-1}\sigma_k^2)$. The full likelihood of patient i across all visits is

$$L(\boldsymbol{y}_i, \omega_i, \boldsymbol{u}_i) = \left[\prod_{j=1}^{J_i}\prod_{k=1}^{K}\mathrm{p}(y_{ijk}|\boldsymbol{u}_i, \omega_i)\right]\mathrm{p}(\omega_i)\mathrm{p}(\boldsymbol{u}_i). \tag{6}$$

We refer to this model as model Dep-NI because $\boldsymbol{u}_i$ and ϵ_{ijk} are marginally dependent. The corresponding parameter vector is $\boldsymbol{\Phi} = (\boldsymbol{a}^{'}, \boldsymbol{b}^{'}, \boldsymbol{\beta}^{'}, \rho, \sigma_u, \sigma_k, \nu)^{'}$. We also refer to model Dep-NI with the Student's t, slash, and contaminated normal distributions as models Dep-T, Dep-SL, and Dep-CN, respectively.

Model Dep-NI assumes that the continuous variable and random effects vector share the same scale of outliers and heavy tails, which may not be true and may negatively affect the model inference. Alternatively, we assume that $\boldsymbol{u}_i$ and ϵ_{ijk} are scaled by different weight variables: $\boldsymbol{u}_i \sim \mathrm{NI}(\mathbf{0}, \Sigma_u, \omega_{1i})$, where ω_{1i} is a subject-level weight variable for $\boldsymbol{u}_i$, and $\epsilon_{ijk} \sim \mathrm{NI}(0, \sigma_k^2, \omega_{2ijk})$, where ω_{2ijk} is a weight variable (specific to outcome k from patient i at visit j) for continuous outcome y_{ijk}, and ω_{1i} and ω_{2ijk} are independent. Applying to the MLIRT model, we specify $\boldsymbol{u}_i$, ϵ_{ijk}, ω_{1i}, and ω_{2ijk} hierarchically as

$$\begin{aligned}
\boldsymbol{u}_i|\omega_{1i} &\sim \mathrm{N}\left(\mathbf{0}, \omega_{1i}^{-1}\Sigma_u\right), \\
\epsilon_{ijk}|\omega_{2ijk} &\sim \mathrm{N}\left(\mathbf{0}, \omega_{2ijk}^{-1}\sigma_k^2\right), \\
\omega_{1i} &\sim \mathrm{p}(\omega_{1i}|\nu_1), \\
\omega_{2ijk} &\sim \mathrm{p}(\omega_{2ijk}|\nu_2).
\end{aligned}$$

Then the continuous outcome y_{ijk} follows $y_{ijk}|\boldsymbol{u}_i, \omega_{2ijk} \sim \mathrm{N}(a_k + b_k\theta_{ij}, \omega_{2ijk}^{-1}\sigma_k^2)$. Let $\boldsymbol{\omega}_i$=$(\omega_{1i}, \boldsymbol{\omega}_{2i})$, where $\boldsymbol{\omega}_{2i} = \{\omega_{2ijk}\}$ for $j = 1, \ldots, J_i$ and $k = 1, \ldots, K$. The full likelihood of patient i is

$$L(\boldsymbol{y}_i, \boldsymbol{\omega}_i, \boldsymbol{u}_i) = \prod_{j=1}^{J_i}\left[\prod_{k=1}^{K}\mathrm{p}(y_{ijk}|\boldsymbol{u}_i, \omega_{2ijk})\mathrm{p}(\omega_{2ijk})\right]\mathrm{p}(\omega_{1i})\mathrm{p}(\boldsymbol{u}_i). \tag{7}$$

We refer to this model as model Indep-NI because $\boldsymbol{u}_i$ and ϵ_{ijk} are marginally independent. The corresponding parameter vector is $\boldsymbol{\Phi} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k, \nu_1, \nu_2)'$. We also refer to model Indep-NI with the Student's t, slash, and contaminated normal distributions as models Indep-T, Indep-SL, and Indep-CN, respectively. Note that models Dep-CN and Indep-CN are Bayesian mixture models because the CN distribution is a two-component mixture distribution. As discussed in Gelman et al. (2013) and Fox (2010), there may be an identifiability issue named label switching in the Bayesian mixture models, i.e., the resulting posterior distribution is invariant to permutations in the labeling of mixture components. Jasra et al. (2005) gave an excellent review of the label switching problem and some solutions. However, models Dep-CN and Indep-CN do not have the identifiability issue because of the increasing order of the variances of the two component distributions, that is $\sigma_k^2/\gamma \geq \sigma_k^2$ as $0 < \gamma \leq 1$. The simulation results from models Dep-NI and Indep-NI with all three NI distributions (Student's t, slash, and CN) are displayed in Web Tables 1 and 2. The fact that all parameters can be successfully recovered in all these models suggest that our proposed models are identifiable.

2.4 Bayesian inference and model selection criteria

We develop a Bayesian approach based on MCMC posterior simulations to analyze the multivariate longitudinal data by applying the NI distributions to the MLIRT models. The fully Bayesian inference has many advantages. First, MCMC algorithms can be used to estimate exact posterior distributions of the parameters, while likelihood-based estimation only produces a point estimate of the parameters, with asymptotic standard errors (Dunson, 2007). Second, Bayesian inference provides better performance in small samples compared to likelihood-based estimation (Lee and Song, 2004). In addition, it is more straightforward to deal with more complicated models using Bayesian inference via MCMC. The model fitting is conducted using the `BUGS` language implemented in `OpenBUGS` (`OpenBUGS` version 3.2.3).

We assume vague prior distributions on all elements in the parameter vectors $\boldsymbol{\Phi}$. Specifically, the prior distribution of all parameters in $\boldsymbol{\beta}$ is $\mathrm{N}(0, 100)$. We use the prior distribution $\mathrm{Gamma}(0.001, 0.001)$ for σ_u and for all components in $\boldsymbol{b}$ to ensure positivity, and use $\mathrm{Uniform}[-1, 1]$ for ρ. The prior distribution for the difficulty parameter a_k of the continuous outcomes is $a_k \sim \mathrm{N}(0, 100)$. For the ordinal outcomes, we let $a_{k1} \sim \mathrm{N}(0, 100)$, and $a_{kl} = a_{k,l-1} + \delta_l$ for $l = 2, n_k - 1$ with $\delta_l \sim \mathrm{N}(0, 100)I(0,)$ (normal distribution left truncated at 0). For the parameters related to the NI distributions, we use $\mathrm{Gamma}(0.001, 0.001)$ for parameter ν in the Student's t and slash distributions and use $\mathrm{Beta}(1, 1)$ for parameters γ and ν in the contaminated normal distribution. We have explored other prior distributions in the simulation study and data analysis and have obtained very similar results. Multiple chains with dispersed initial values are run. To assess convergence, we use the history plots to ensure there are no appearance of trend for all parameters. In addition, we use Gelman-Rubin diagnostic statistics to ensure the scale reduction $\hat{R}$ of all parameters are smaller than 1.1 (Gelman et al., 2013). To facilitate easy reading and implementation of the proposed models, a sample `BUGS` code for fitting model Indep-CN has been posted in the Web supplement.

Among many model selection criteria for Bayesian inference, we have selected the log pseudo-marginal likelihood (LPML) and Bayes factor (BF) to assess model performance. Conditional predictive ordinate (CPO) (Geisser, 1993; Carlin and Louis, 2011) is a cross-validated predictive method that assesses the predictive distribution conditional on the data but with single data point deleted. Let $\boldsymbol{y}$ be the full observed data and $\boldsymbol{y}_{(i)}$ be the data with patient i deleted. Then the CPO for patient i is defined as $\mathrm{CPO}_i = \mathrm{p}(y_i|\boldsymbol{y}_{(i)}) = \int \mathrm{p}(y_i|\boldsymbol{\Phi})\mathrm{p}(\boldsymbol{\Phi}|\boldsymbol{y}_{(i)})d\boldsymbol{\Phi}$. Large CPO indicates that the data for patient i can be well predicted by the model using posterior density of $\boldsymbol{\Phi}$ based on $\boldsymbol{y}_{(i)}$. For our proposed models, there is no close form for CPO_i, thus a Monte Carlo estimation method is used to obtain $\widehat{\mathrm{CPO}}_i$ (Chen et al., 2000). A summary statistics of CPO_i is log pseudo-marginal likelihood (LPML), defined as $\mathrm{LPML} = \sum_{i=1}^{N} \log(\widehat{\mathrm{CPO}}_i)$. A larger value of LPML indicates a better model fitting.

Table 1 Simulation results from models Indep-N, Dep-CN, Indep-CN in setting I in which there were no outliers in the continuous outcome and random effects.

		Indep-N				Dep-CN				Indep-CN			
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
a_1	25.000	0.031	0.501	0.514	0.945	0.009	0.494	0.516	0.950	0.043	0.500	0.510	0.955
b_1	10.000	0.033	0.392	0.376	0.945	0.013	0.393	0.376	0.939	0.027	0.395	0.377	0.939
a_2	80.000	0.024	0.980	1.007	0.954	0.007	0.994	1.010	0.950	0.053	0.970	1.001	0.952
b_2	18.000	0.047	0.732	0.700	0.939	0.013	0.735	0.699	0.944	0.028	0.737	0.701	0.930
a_{31}	−2.700	−0.016	0.131	0.139	0.970	−0.011	0.132	0.139	0.980	−0.019	0.135	0.139	0.964
a_{32}	−0.600	−0.011	0.122	0.122	0.939	−0.008	0.122	0.122	0.950	−0.016	0.123	0.121	0.942
a_{33}	2.000	0.007	0.129	0.132	0.954	0.010	0.127	0.132	0.959	0.003	0.129	0.131	0.952
a_{34}	2.800	0.012	0.136	0.141	0.960	0.017	0.134	0.141	0.965	0.006	0.137	0.140	0.967
a_{35}	5.000	0.043	0.187	0.187	0.939	0.042	0.183	0.187	0.944	0.033	0.189	0.187	0.948
a_{36}	6.000	0.055	0.214	0.214	0.939	0.057	0.212	0.215	0.953	0.044	0.218	0.214	0.936
b_3	2.000	0.023	0.094	0.095	0.942	0.017	0.095	0.094	0.942	0.019	0.096	0.095	0.958
a_{41}	−0.100	0.003	0.052	0.052	0.957	−0.000	0.053	0.052	0.956	0.001	0.055	0.052	0.942
a_{42}	1.000	0.001	0.060	0.057	0.921	0.001	0.059	0.057	0.942	−0.000	0.060	0.057	0.921
a_{43}	1.800	0.004	0.069	0.068	0.945	0.005	0.069	0.068	0.947	0.004	0.071	0.068	0.939
a_{44}	2.600	0.007	0.089	0.087	0.951	0.009	0.088	0.087	0.947	0.006	0.089	0.086	0.942
a_{45}	3.300	0.016	0.114	0.111	0.945	0.020	0.112	0.111	0.944	0.016	0.116	0.111	0.939
a_{46}	4.000	0.037	0.151	0.147	0.939	0.040	0.148	0.147	0.936	0.041	0.152	0.147	0.930
b_4	0.400	0.001	0.029	0.027	0.936	0.001	0.028	0.027	0.936	0.001	0.028	0.027	0.936
β_{10}	0.400	0.005	0.085	0.090	0.954	0.007	0.084	0.091	0.959	0.009	0.085	0.090	0.967
β_{11}	−0.500	−0.005	0.122	0.124	0.948	0.001	0.123	0.124	0.950	−0.014	0.123	0.124	0.930
ρ	0.400	−0.002	0.044	0.045	0.951	−0.001	0.045	0.045	0.942	−0.002	0.044	0.045	0.955
σ_u	1.300	0.004	0.070	0.064	0.942	0.004	0.071	0.064	0.933	0.002	0.068	0.064	0.936

Table 2 Simulation results from models Indep-N, Dep-CN, Indep-CN in setting II in which there were 5% outliers in both the continuous outcome and random effects.

		Indep-N				Dep-CN				Indep-CN			
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
a_1	25.000	1.337	0.625	0.888	0.742	0.146	0.554	0.532	0.927	0.154	0.581	0.526	0.930
b_1	10.000	6.369	0.762	0.632	0.000	−0.165	0.397	0.384	0.920	−0.021	0.419	0.389	0.953
a_2	80.000	3.852	1.104	1.589	0.216	−0.015	1.031	1.038	0.953	0.075	1.130	1.026	0.922
b_2	18.000	12.537	0.951	1.163	0.000	−0.268	0.732	0.707	0.916	−0.057	0.755	0.707	0.938
a_{31}	−2.700	−0.444	0.149	0.198	0.338	−0.013	0.138	0.142	0.953	−0.016	0.143	0.141	0.950
a_{32}	−0.600	−0.413	0.127	0.180	0.282	0.005	0.119	0.124	0.949	−0.008	0.127	0.123	0.950
a_{33}	2.000	−0.375	0.144	0.184	0.467	0.024	0.138	0.135	0.949	−0.002	0.149	0.133	0.919
a_{34}	2.800	−0.369	0.151	0.191	0.512	0.024	0.146	0.145	0.935	0.001	0.155	0.143	0.942
a_{35}	5.000	−0.378	0.193	0.228	0.652	0.029	0.188	0.191	0.942	0.004	0.197	0.189	0.950
a_{36}	6.000	−0.373	0.223	0.253	0.711	0.046	0.217	0.220	0.949	0.027	0.233	0.218	0.953
b_3	2.000	1.321	0.150	0.165	0.000	−0.028	0.104	0.096	0.924	0.001	0.110	0.097	0.911
a_{41}	−0.100	−0.082	0.053	0.059	0.791	0.004	0.051	0.053	0.938	−0.001	0.055	0.052	0.950
a_{42}	1.000	−0.081	0.052	0.063	0.798	0.000	0.053	0.057	0.967	−0.001	0.052	0.057	0.977
a_{43}	1.800	−0.077	0.063	0.073	0.868	0.008	0.062	0.069	0.967	0.003	0.065	0.069	0.961
a_{44}	2.600	−0.075	0.085	0.090	0.861	0.012	0.087	0.088	0.931	0.004	0.085	0.087	0.942
a_{45}	3.300	−0.070	0.105	0.113	0.895	0.025	0.107	0.112	0.956	0.011	0.103	0.111	0.965
a_{46}	4.000	−0.067	0.150	0.145	0.913	0.035	0.155	0.145	0.920	0.018	0.149	0.144	0.942
b_4	0.400	0.274	0.035	0.042	0.000	−0.004	0.026	0.025	0.942	0.001	0.026	0.025	0.938
β_{10}	0.400	0.021	0.064	0.075	0.969	−0.018	0.111	0.098	0.898	0.010	0.096	0.092	0.930
β_{11}	−0.500	0.109	0.083	0.089	0.753	0.022	0.136	0.127	0.945	−0.008	0.121	0.126	0.950
ρ	0.400	0.353	0.021	0.025	0.000	0.037	0.042	0.045	0.884	0.029	0.040	0.046	0.942
σ_u	1.300	−0.108	0.055	0.047	0.387	−0.004	0.062	0.063	0.938	0.001	0.064	0.064	0.938

Furthermore, Bayes factor (BF) is a Bayesian alternative to p-value for hypothesis testing among competing models. The BF quantifies the the degree to which whether the observed data support a hypothesis (Kass and Raftery, 1995; Lewis and Raftery, 1997). Let two competing models be M_1 and M_2 and observed data be $\boldsymbol{y}$, then BF in favor of model M_1 over M_2 is defined as $\mathrm{BF}(M_1; M_2) = \frac{f(\boldsymbol{y}|M_1)}{f(\boldsymbol{y}|M_2)} = \frac{\int f(\boldsymbol{y}|\boldsymbol{\Phi}_1, M_1) f(\boldsymbol{\Phi}_1|M_1) d\boldsymbol{\Phi}_1}{\int f(\boldsymbol{y}|\boldsymbol{\Phi}_2, M_2) f(\boldsymbol{\Phi}_2|M_2) d\boldsymbol{\Phi}_2}$, where $\boldsymbol{\Phi}_i$ is the parameter vectors for model M_i for $i = 1, 2$; $f(\boldsymbol{y}|\boldsymbol{\Phi}_i, M_i)$ is the likelihood of model M_i; and $f(\boldsymbol{\Phi}_i|M_i)$ is the posterior density of $\boldsymbol{\Phi}_i$ for model M_i. The Laplace-Metropolis estimator based on normal distribution is used to approximate the marginal likelihood $f(\boldsymbol{y}|M_i)$. In particular, the $f(\boldsymbol{y}|M_i) \approx (2\pi)^{d_i/2}|\Sigma_i|^{1/2} f(\boldsymbol{y}|\overline{\boldsymbol{\Phi}}_i, M_i) f(\overline{\boldsymbol{\Phi}}_i|M_i)$, where d_i is the number of parameters in $\boldsymbol{\Phi}_i$, Σ_i is the posterior covariance matrix of $\boldsymbol{\Phi}_i$, $\overline{\boldsymbol{\Phi}}_i$ is the posterior mean of $\boldsymbol{\Phi}_i$, $f(\overline{\boldsymbol{\Phi}}_i|M_i)$ is the prior probability of parameters evaluated at $\overline{\boldsymbol{\Phi}}_i$, and $f(\boldsymbol{y}|\overline{\boldsymbol{\Phi}}_i, M_i)$ is the likelihood evaluated at the posterior mean $\overline{\boldsymbol{\Phi}}_i$. When BF is greater than 100, decisive evidence is shown in favor of model M_1 over M_2 (Kass and Raftery, 1995).

3 Simulation studies

In this section, we conducted an extensive simulation study under three settings to compare the performance of the models Indep-N, Dep-NI, and Indep-NI in different scenarios. For all settings, we generated 350 datasets with a sample size of 400 patients (200 in both treatment and placebo groups). The data structure was similar to the motivating DATATOP study, and it had two continuous outcomes and two ordinal outcomes (both with seven categories) at five visits (months 0, 1, 3, 9, 15).

We generated data from model (4) with $\boldsymbol{X}_{i0} = 0$ and $\boldsymbol{X}_{i1} = x_i$, where the covariate x_i took value 0 or 1 each with probability 1/2 to mimic treatment assignment. We set the coefficients to be $\boldsymbol{\beta} = (\beta_{10}, \beta_{11})' = (0.4, -0.5)'$. The parameters for the continuous outcomes were set to be $a_1 = 25, b_1 = 10, \sigma_1 = 5$ and $a_2 = 80, b_2 = 18, \sigma_2 = 20$. The parameters for the ordinal outcomes were set to be $\boldsymbol{a}_3 = (-2.7, -0.6, 2, 2.8, 5, 6)$, $b_3 = 2$, $\boldsymbol{a}_4 = (-0.1, 1, 1.8, 2.6, 3.3, 4)$, $b_4 = 0.4$. We assumed that the subject-specific random effects vector $\boldsymbol{u}_i = (u_{i0}, u_{i1})' \sim N_2(\boldsymbol{0}, \Sigma_u)$, where $\Sigma_u = \{(1, \rho\sigma_u), (\rho\sigma_u, \sigma_u^2)\}$ with $\rho = 0.4$ and $\sigma_u = 1.3$. We applied the Bayesian framework in Section 2.3 and we ran two MCMC chains with dispersed initial values. Each MCMC chain was run for 30,000 iterations with the first 15,000 iterations discarded as burn-in. We computed the average of the posterior mean minus the true values (Bias), the standard deviation of the posterior means (SD), the square root of the average of the posterior variance (SE), and the coverage probabilities (CP) of the 95% equal-tail credible intervals.

To ensure that the all parameters including NI-distribution components can be correctly estimated, we first simulated data from each of the models Dep-NI and Indep-NI with all three NI distributions (Student's t, slash, and CN) and estimated parameters using the corresponding true models. The simulation results are displayed in Web Tables 1 and 2. The fact that the bias was negligible, SE was close to SD, and the coverage probability was close to the nominal value suggests that all parameters can be successfully recovered in the proposed models Dep-NI and Indep-NI with all three NI distributions and that our models were identifiable.

In setting I, both the continuous outcomes and the random effects followed normal distributions without outliers and model Indep-N was the true model. The results were summarized in Table 1. Due to space constraints, we only presented the MLIRT models using the CN distribution (i.e., models Dep-CN, and Indep-CN). The results suggested that all three models (Indep-N, Dep-CN, and Indep-CN) generated comparable and reasonable results, that is negligible bias, SE being close to SD, and CP's being close to nominal level of 95%.

In setting II, we evaluated the model performance in the presence of outliers and heavy tails. The simulation setting was similar to setting I, but with the first continuous outcome and the random effects being generated from normal distributions with 5% outliers, while the second continuous outcome still followed a normal distribution. To generate outliers, we randomly selected 5% data

Table 3 Simulation results of parameter estimation on $\boldsymbol{\beta}$ from models Indep-SL, Indep-T, Indep-CN in setting III in which there were 5% outliers in both the continuous outcome and random effects. The best fitting models are highlighted in bold.

Models	Parameters	True	Bias	SD	SE	CP
Dep-T	β_{10}	0.400	0.041	0.123	0.106	0.876
	β_{11}	−0.500	−0.039	0.152	0.135	0.897
Dep-SL	β_{10}	0.400	0.180	0.156	0.139	0.707
	β_{11}	−0.500	−0.205	0.200	0.177	0.751
Dep-CN	β_{10}	0.400	**−0.018**	**0.111**	**0.098**	**0.898**
	β_{11}	−0.500	**0.022**	**0.136**	**0.127**	**0.945**
Indep-T	β_{10}	0.400	0.076	0.119	0.108	0.848
	β_{11}	−0.500	−0.075	0.151	0.143	0.917
Indep-SL	β_{10}	0.400	0.167	0.142	0.132	0.756
	β_{11}	−0.500	−0.202	0.176	0.175	0.799
Indep-CN	β_{10}	0.400	**0.010**	**0.096**	**0.092**	**0.930**
	β_{11}	−0.500	**−0.008**	**0.121**	**0.126**	**0.950**

from the first continuous outcome and added the noise generated from either Uniform$[3\sigma_1, 6\sigma_1]$ or Uniform$[-6\sigma_1, -3\sigma_1]$ with equal probability. Similarly, we randomly selected 5% of the generated random effects $\boldsymbol{u}_i$ and replaced by data generated from either Uniform[5, 15] or Uniform[−15, −5] with equal probability.

Table 2 displays the results of setting II. The results from models Dep-CN and Indep-CN were reasonably good with negligible bias, SE being close to SD, and the CP's all close to the nominal value, suggesting that accounting for the outliers and heavy tails leads to valid inference. In contrast, model Indep-N provided severely biased estimates and low coverage probabilities for all parameters. This was because that the misspecification of the random effects distribution affected the estimation of $\boldsymbol{u}_i$ and thus of the parameters in model (4) and of θ_{ij}. As shown in models (2) and (3), the imprecise estimation on θ_{ij} affected the estimation of the outcome-specific parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ for all outcomes. Additionally, we also evaluated the situation where the outliers on the random effects were ignored and only the outliers on the first continuous outcome were accounted for using CN distribution. Due to space constraints, the results were summarized in Table A1 of the Appendix. Large bias and poor CPs were observed on most of the outcome-specific parameter vectors ($\boldsymbol{a}$ and $\boldsymbol{b}$) as well as the treatment effect parameter (β_{11}). In sum, setting II suggested that the misspecification of the distributions of the continuous outcome and random effects severely impacted the inference in the MLIRT modeling framework, while the models accounting for the outliers and heavy tails using the NI distributions provided valid inference.

As pointed out by Gelman et al. (2006), the posterior distributions resulting from the vague gamma and inverse gamma prior distributions are sensitive to the choice of hyper-parameters and hence half-normal distributions are recommended. Per the suggestion from one reviewer, we have reanalyzed all simulated datasets in setting II using models Indep-N, Dep-CN, and Indep-CN while replacing the gamma prior distributions by half-normal prior distribution $N(0, 100)$ with support being larger than zero. The simulation results are displayed in Web Table 3. The results from all three models were very similar to their counterparts using gamma prior distributions in Table 2, suggesting that both gamma and half-normal prior distributions work sufficiently well in our models.

In setting III, we compared the performance of models Dep-NI and Indep-NI with all three NI distributions (Student's t, slash, and CN) when the datasets had 5% outliers as in setting II. Due to space constraints, only the estimations on the treatment effect parameter vector $\boldsymbol{\beta}$ were summarized in

Table 3. The results suggested that both models Dep-CN and Indep-CN performed better than their counterparts with Student's t and slash distributions, as indicated by smaller bias, smaller SD and SE, and the CP's closer to the nominal levels. It is of note that model Indep-CN had slightly more accurate parameter estimation than model Dep-NI in this setting because the scales of outliers were different in the first continuous outcome and in the random effects (i.e., the noises were generated from different uniform distributions).

From the simulation study, we concluded that models Indep-N, Dep-NI, and Indep-NI all provided comparable and satisfactory results when both the continuous outcomes and the random effects followed normal distributions. However, when outliers existed in both the continuous outcome and the random effects, both models Dep-NI and Indep-NI provided more reasonable results than model Indep-N, which provides severely biased estimates to all parameters.

4 Application to the DATATOP study

In this section, we applied the proposed models Dep-CN and Indep-CN to the motivating Deprenyl And Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) study, which is a double-blind, placebo-conrolled, multicenter clinical trial. A factorial design in this study was used to test the hypothesis that patients with early Parkinson's disease with deprenyl 10 mg/d and/or tocopherol (vitamin E) 2000 IU/d will delay the time until the initiation of levodopa therapy. There were 800 eligible patients enrolled in DATATOP and randomized to one of the four treatment arms: active deprenyl alone, active tocopherol alone, both active deprenyl and tocopherol, and double placebo. Only deprenyl was found to be effective in delaying the time until the initiation of levodopa therapy (Parkinson Study Group, 1989, 1993). In our analysis, we define the treatment group as the patients who received deprenyl (active deprenyl alone and both active deprenyl and tocopherol), and define the placebo group as the patients who did not receive deprenyl (active tocopherol alone and double placebo). We considered three longitudinal outcomes, that is the Unified Parkinson's Disease Rating Scale (UPDRS) total score, Hoehn and Yahr scale (HY), and Schwab and England activities of daily living (SEADL), which were collected at baseline, month 1, every three months from month 3 to month 24. The UPDRS total score evaluates patients' mentation, behavior, and activities of daily life and it is approximated by a continuous variable with integer value from 0 (not affected) to 176 (most severely affected) (Bushnell and Martin, 1999). Outcome HY measures the disability level in daily activities and it is an ordinal variable ranging from 1 to 5 with higher values indicating worse conditions (Müller et al., 2000). Outcome SEADL assesses patients' daily activities and it is an ordinal variable with integer values from 0 to 100 incrementing by 5 with larger values indicating better clinical conditions (McRae et al., 2000). We combined some categories with zero or small counts so that the ordinal variables HY and SEADL have 5 and 6 categories, respectively. We also recoded the SEADL variable so that higher values in all three outcomes are worse clinical condition. We removed one patient who has no UPDRS measurements in any visit so that there were 398 and 401 patients in the treatment and placebo groups, respectively.

Figure 2 displays the longitudinal profile of the observed outcome UPDRS. PD is a slow progression disease, so slow progression in UPDRS score such as patient 621 (solid line in left panel) is often observed. It is unexpected to observe sudden value change in UPDRS measurements. However, patients 105, 108, and 749 (dashed lines) had some potential outlying measurements indicated by their dramatic value changes in their UPDRS profiles.

To fit our proposed models to the DATATOP dataset, we let $\boldsymbol{X}_{i0} = 0$ and considered the treatment assignment as the only covariate in $\boldsymbol{X}_{i1}$. So model (4) became $\theta_{ij} = u_{i0} + (\beta_{10} + \beta_{11}x_i + u_{i1})t_{ij}$. We used two parallel MCMC chains with overdispersed initial values, and ran each chain for 30,000 iterations. The first 15,000 iterations were discarded as burn-in and the inference was based on the remaining 15,000 iterations. Table 4 compares models Indep-N, Dep-CN, and Indep-CN using the model comparison criteria discussed in Section 2.4. The proposed models Dep-CN and Indep-CN

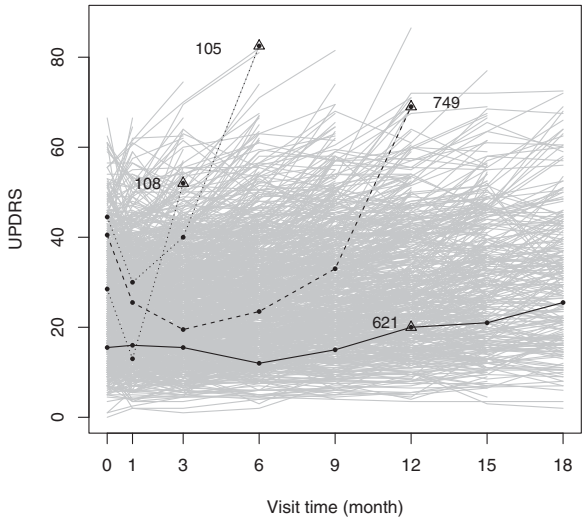

Figure 2 Longitudinal profile of the outcome UPDRS. Numbers 105, 108, 621, and 749 denote four patients.

Table 4 Model comparison statistics for the DATATOP dataset from models Indep-N, Dep-CN, and Indep-CN. The best fitting model is highlighted in bold.

	LPML	BF
Indep-N	−27312.50	$\gg$100
Dep-CN	−26930.44	$\gg$100
Indep-CN	**−26873.84**	Ref

performed significantly better than model Indep-N with larger LPML value, suggesting the advantage of accounting for outliers and heavy tails in the outcome UPDRS and random effects. The BFs in favor of model Indep-CN over models Dep-CN and Indep-N were much larger than 100, suggesting decisive evidence in favor of model Indep-CN. The Indep-CN model had the best fit in terms of LPML and BF values and hence it was selected as the final model.

Table 5 displays the posterior means, standard deviation (SD), and 95% equal-tail credible intervals (CI) from various models. The results from all models suggested that placebo patients experienced significant deterioration in PD symptoms overtime (β_{10}) and deprenyl effectively delayed the progression of PD symptoms (β_{11}), which was consistent with the findings in the original DATATOP study analysis (Parkinson Study Group, 1993). Specifically, the disease progression rates for the placebo patients (β_{10}) were 1.285 (95% CI: [1.163, 1.414]), 1.334 (95% CI: [1.200, 1.471]), and 1.226 (95% CI: [1.107, 1.351]) units per year from models Indep-N, Dep-CN, and Indep-CN, respectively. The changes in disease progression rate introduced by deprenyl (β_{11}) were −0.606 (95% CI: [−0.751, −0.457]), −0.609 (95% CI: [−0.762, −0.461]), and −0.572 (95% CI: [−0.707, −0.443]) units per year from models Indep-N, Dep-CN, and Indep-CN, respectively. Another interesting observation from Table 5 is that the results from model Indep-CN suggested markedly different proportions (parameters ν_1 vs. ν_2) and scales (parameters γ_1 vs. γ_2) of contamination in the random effects vector and outcome UPDRS. This is also the reason why model Indep-CN was preferred based on the model selection criteria.

To obtain further insight into how the NI distributions control the influence of the outliers, in Fig. 3 we plotted the standardized residuals (SRs) of UPDRS measurements for all patients at each visit

Table 5 Results of fitting various models in the DATATOP dataset. Parameters a_k and b_k for $k = 1, 2, 3$ are the outcome-specific parameters for the outcomes UPDRS, HY, and SEADL. Parameters ν and γ are from model Dep-CN. Parameters ν_1 and γ_1 are for the random effects vector and parameters ν_2 and γ_2 are for outcome UPDRS from model Indep-CN.

	Indep-N		Dep-CN		Indep-CN	
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
a_1	23.985 (0.391)	23.185, 24.670	22.280 (0.400)	21.510, 23.085	23.595 (0.408)	22.695, 24.365
b_1	10.855 (0.274)	10.325, 11.400	8.748 (0.301)	8.129, 9.334	10.220 (0.316)	9.596, 10.855
σ_1	5.230 (0.0756)	5.0840, 5.383	4.082 (0.102)	3.881, 4.280	3.592 (0.179)	3.241, 3.931
a_{21}	−0.870 (0.061)	−0.984, −0.744	−0.668 (0.065)	−0.796, −0.543	−0.849 (0.063)	−0.970, −0.716
a_{22}	0.101 (0.060)	−0.012, 0.223	0.308 (0.064)	0.183, 0.430	0.113 (0.062)	−0.004, 0.242
a_{23}	3.218 (0.082)	3.062, 3.384	3.458 (0.088)	3.288, 3.632	3.220 (0.082)	3.061, 3.384
a_{24}	5.471 (0.129)	5.224, 5.729	5.768 (0.138)	5.502, 6.043	5.494 (0.132)	5.234, 5.756
b_2	1.398 (0.050)	1.304, 1.498	1.170 (0.050)	1.070, 1.266	1.312 (0.052)	1.211, 1.415
a_{31}	−2.552 (0.086)	−2.716, −2.385	−2.324 (0.091)	−2.506, −2.148	−2.473 (0.087)	−2.641, −2.298
a_{32}	−0.504 (0.074)	−0.642, −0.354	−0.243 (0.080)	−0.402, −0.088	−0.475 (0.076)	−0.621, −0.319
a_{33}	1.935 (0.082)	1.784, 2.103	2.277 (0.089)	2.104, 2.451	1.913 (0.082)	1.757, 2.083
a_{34}	2.777 (0.089)	2.611, 2.957	3.154 (0.097)	2.966, 3.349	2.740 (0.088)	2.572, 2.922
a_{35}	4.957 (0.120)	4.728, 5.200	5.434 (0.132)	5.181, 5.696	4.894 (0.117)	4.672, 5.131
b_3	1.825 (0.062)	1.708, 1.953	1.570 (0.066)	1.442, 1.700	1.676 (0.064)	1.553, 1.805
β_{10}	1.285 (0.062)	1.163, 1.414	1.334 (0.069)	1.200, 1.471	1.226 (0.063)	1.107, 1.351
β_{11}	−0.606 (0.074)	−0.751, −0.457	−0.609 (0.078)	−0.762, −0.461	−0.572 (0.067)	−0.707, −0.443
ρ	0.362 (0.042)	0.277, 0.440	0.373 (0.044)	0.286, 0.458	0.354 (0.047)	0.263, 0.447
σ_u	0.818 (0.041)	0.740, 0.898	0.755 (0.039)	0.683, 0.835	0.739 (0.039)	0.666, 0.819
ν	—	—	0.247 (0.029)	0.194, 0.309	—	—
γ	—	—	0.204 (0.015)	0.176, 0.233	—	—
ν_1	—	—	—	—	0.038 (0.013)	0.018, 0.067
γ_1	—	—	—	—	0.058 (0.023)	0.025, 0.113
ν_2	—	—	—	—	0.217 (0.042)	0.141, 0.304
γ_2	—	—	—	—	0.167 (0.015)	0.139, 0.199

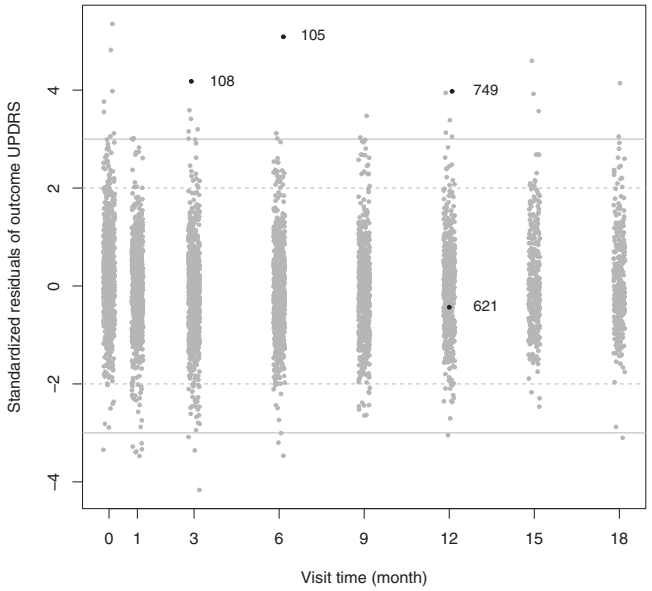

Figure 3 Standardized residuals of the UPDRS measurements for all patients at each visit when fitting model Indep-N. The dashed lines are horizontal lines at -2 and 2 and the solid lines are horizontal lines at -3 and 3. Numbers 105, 108, 621, and 749 denote four patients.

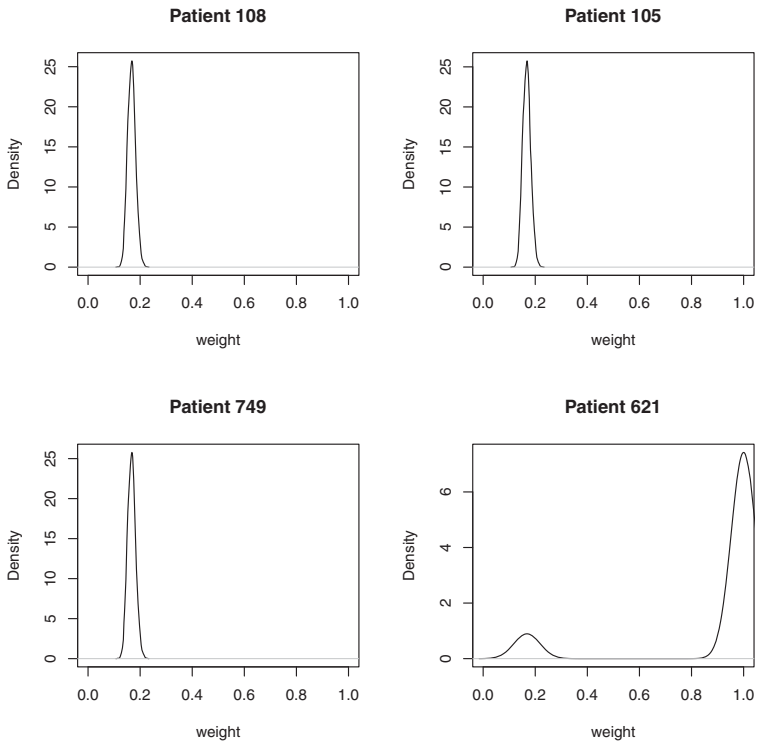

Figure 4 Estimates of the weight variable ω_{ijk} for patients 105, 108, 621, and 749 at certain visits from model Indep-CN.

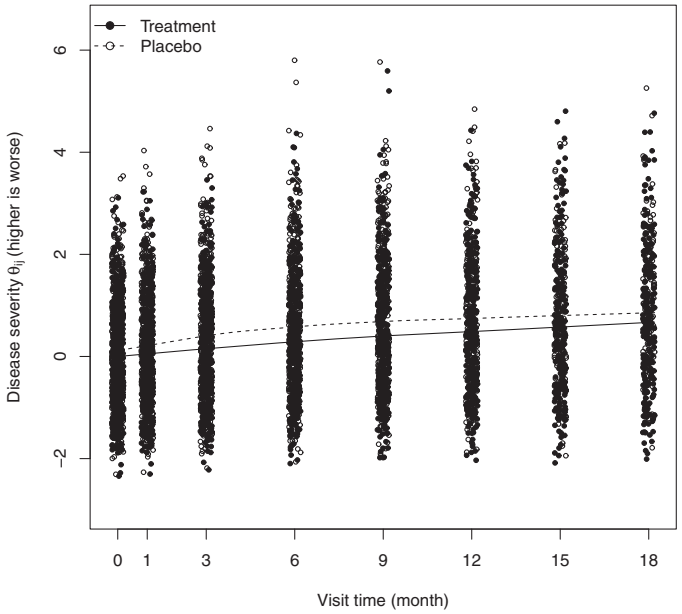

Figure 5 Bayesian posterior estimates of the subject-specific disease severity θ_{ij} at each visit and the lowess smooth curves for treatment and placebo groups from model Indep-CN.

after fitting model Indep-N. A few data points had SRs with absolute value larger than 3 (e.g., 5.12 for patient 105 at 6-month visit, 4.18 for patient 108 at 3-month visit, and 4.00 for patient 749 at 12-month visit), indicating potential outliers. In contrast, the SRs for patient 621 at 12-month visit was −0.40, indicating a nonoutlier. Without proper adjustment for the outliers, they may affect the accuracy of the model estimation due to the violation of the normality assumption. As pointed out by Rosa et al. (2003), the weight variable ω_{ijk} in the NI distributions can be estimated and used for outlier detections. Figure 4 displays the posterior distributions of the weight variable ω_{ijk} for patients 105, 108, 749, and 621 at certain visits after fitting the final model Indep-CN. As indicated in Figs. 2 and 3, patient 105 at 6-month visit, patient 108 at 3-month visit, and patient 749 at 12-month visit were potential outliers, their posterior distributions of the weight variable ω_{ijk} in Fig. 4 were sharp with majority of the density close to zero and their posterior means of the weight variable ω_{ijk} were 0.18, 0.17, and 0.18, respectively. In comparison, for patient 621 at 12-month visit, the posterior distribution of the weight had majority of the density at large value with posterior mean of 0.91, which indicated that this observation was not an outlier.

Figure 5 displays the visualization of disease progression for all patients. Each black dot (for treatment group) or circle (for placebo group) was a Bayesian posterior estimate of the subject-specific latent disease severity θ_{ij} for patient i at visit j. The solid line and dashed line were the the lowess smooth curves for treatment and placebo groups, respectively. Figure 5 suggests that the placebo patients had a faster disease progression rate than the treatment patients as manifested by the large gaps between two lowess smooth curves. In Table 5, the standard error (σ_u) of the random slope (u_{i1}) from model Indep-CN was 0.739 (95% CI: [0.666, 0.819]), while the estimate of the correlation coefficient ρ between u_{i0} and u_{i1} was 0.354 (95% CI: [0.263, 0.447]). The statistically significant positive correlation coefficient indicates that individuals with worse disease severity tends to have faster disease progression rate and vice versa. To gain further insight into u_{i0}, u_{i1}, and ρ, we plotted in Fig. 6 the rankings of individuals' subject-specific disease severity u_{i0} (upper panel) and disease progression rate u_{i1} (lower panel). We

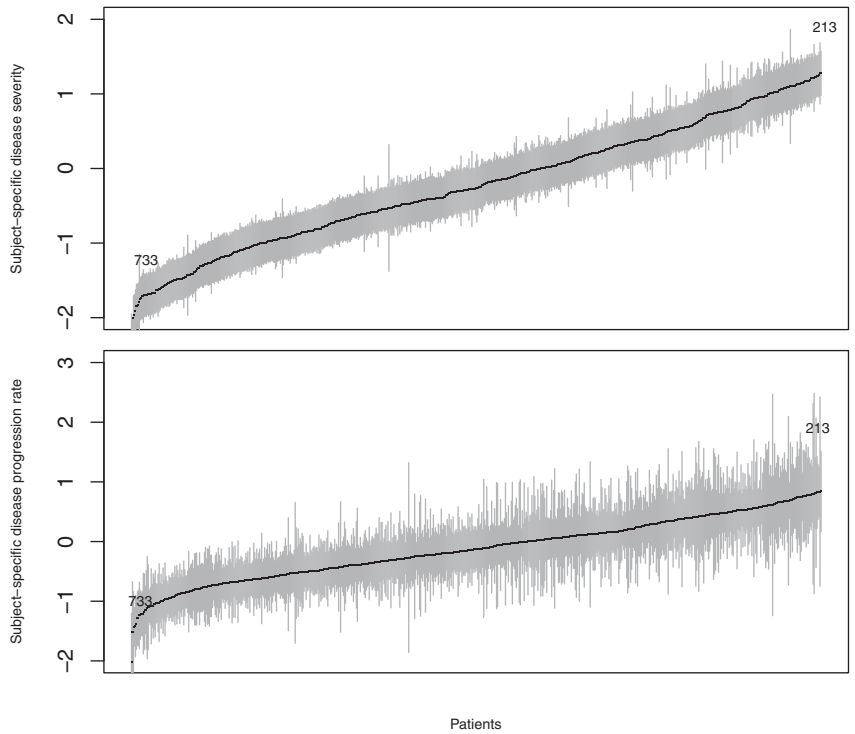

Figure 6 The ranking of subject-specific disease severity u_{i0} (upper panel) and disease progression rate u_{i1} (lower panel) with 95% CI from model Indep-CN. The numbers in the figures are patient numbers.

ranked the patients so that patients with mild disease severity and slow disease progression rate had low ranks; while patients with severe disease and disease progression rate had high ranks. To assist with the interpretation of the positive correlation ρ, we selected two patients in Fig. 6. Patient 213 who had the worst disease severity (ranked 799, upper panel) had the 9th fastest disease progression rate (ranked 791, lower panel) and patient 733 who ranked 16 in the disease severity (upper panel) had the 10th disease progression rate (lower panel).

5 Conclusions and Discussions

In this article, we provided a robust statistical analysis framework for the multivariate longitudinal data while accounting for the outliers and heavy tails in the continuous outcomes and random effects by using the symmetric heave-tailed normal/independent (NI) distributions. Our extensive simulation results demonstrated that when both the continuous outcomes and random effects followed normal distributions, our proposed models Dep-NI and Indep-NI provided satisfactory results comparable to those from the regular model Indep-N. However, when outliers existed in both the continuous outcome and random effects, the proposed models Dep-NI and Indep-NI provided more accurate results than model Indep-N. We applied our models to the motivating DATATOP study and discovered outliers and heavy tails in the continuous outcome UPDRS and the random effects. We have demonstrated that model Indep-NI was the best-fitting model based on the model selection criteria. We displayed the longitudinal profile of outcome UPDRS and provided visual illustration of how the NI distribution

controls the influence of the outliers. We provided visualization of the subject-specific disease severity for each visit to gain insight into the different disease progression rates for the treatment and placebo groups. The figure on the subject-specific disease severity and disease progression rate provides visualization of their correlations. The hierarchical implementation of the NI distributions to the MLIRT model under Bayesian framework is relatively straightforward. The easy access of publicly available software, such as `WinBUGS` and `OpenBUGS`, provides a practical and feasible platform for practitioner and researchers to perform analysis using our proposed method.

There are some limitations in our proposed model that we will address in our future study. The violation of nonnormality may be due to heavy tails or skewness or both. In this article, we only considered the influence of the violation of the normal assumptions due to outliers and heavy tails. For future work, we will investigate the influence of skewness under the MLIRT framework using skewed normal (SN) distribution (Azzalini and Capitanio, 1999) and skewed normal/independent (SNI) distribution (Lachos et al., 2010). Alternatively, we can relax the parametric assumption on the continuous variable and random effects by considering Bayesian nonparametric (BNP) framework based on Dirichlet process mixture (Escobar, 1994). Moreover, the informative dropout (i.e., sicker patients are more likely to drop out earlier) is another common issue in longitudinal studies. Ignoring the "missing values" due to informative dropout leads to biased parameter estimations (Henderson et al., 2000). Thus, a joint model approach that considers both the longitudinal outcome and survival outcome in the presence of outcome outliers and skewness is also part of our future research.

Inference from the proposed Bayesian MLIRT modeling framework is valid when the model fits the data and the model assumptions are met. In this article, we have assumed item parameter invariance (or measurement invariance as in Fox (2010), that is the difficulty and discriminating parameters are assumed to be equal across populations and over time). When the invariance assumption does not hold, outcomes behave differently across populations and time and they show differential item functioning (DIF, i.e. people with the same latent disease severity have different outcome measurements). To address this issue, De Jong et al. (2007) and De Jong et al. (2008) introduced a random item modeling approach which models the item or outcome characteristics as random item effects parameters, where the random part captures random error due to DIF. How to detect and account for DIF in the proposed hierarchical model is an interesting future research direction. Furthermore, the assumption of a single latent variable θ_{ij} (unidimensional assumption) may be unrealistic in the study of a chronic and progressive disease such as PD, because PD is a heterogeneous disorder characterized by multiple impaired domains (e.g. motor, cognitive, and behavioral) with variable clinical symptoms and disease progression (Thenganatt and Jankovic, 2014). We would like to investigate ways to relax the unidimensional assumption to allow multidimensional latent variables in the future research. As illustrated in Fox (2010), it is very difficult to accurately compute the Bayes factor for complex nonlinear models such as the proposed MLIRT models because the computation involves high-dimensional and intractable integrals. Moreover, the Bayes factor for model selection depends on choices of prior distributions and detailed sensitivity analysis should be investigated in the future research.

Acknowledgment The project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2 TR000370. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors are grateful to Drs. Barbara C. Tilley and Adriana Perez for helpful discussion and comments. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performing computing resources that have contributed to the research results reported within this article. URL: http://www.tacc.utexas.edu. The authors express appreciation to The University of Texas School of Public Health information technology staff for their technical support.

Conflict of interest

The authors have declared no conflict of interest.

Appendix

Table A1 Simulation results from a MLIRT model (which ignored the outliers in the random effects but accounted for the outliers on the first continuous outcome using CN distribution) in setting II in which there were 5% outliers in both the continuous outcome and random effects.

		Model ignoring outliers on REs			
	True	Bias	SD	SE	CP
a_1	25.000	2.190	0.617	0.812	0.157
b_1	10.000	7.167	0.284	0.607	0.000
a_2	80.000	3.746	1.139	1.512	0.267
b_2	18.000	12.863	0.554	1.105	0.000
a_{31}	−2.700	−0.433	0.155	0.190	0.340
a_{32}	−0.600	−0.411	0.138	0.175	0.288
a_{33}	2.000	−0.391	0.146	0.179	0.393
a_{34}	2.800	−0.388	0.155	0.185	0.424
a_{35}	5.000	−0.398	0.186	0.220	0.607
a_{36}	6.000	−0.398	0.208	0.243	0.670
b_3	2.000	1.410	0.122	0.158	0.000
a_{41}	−0.100	−0.080	0.056	0.058	0.728
a_{42}	1.000	−0.079	0.058	0.062	0.728
a_{43}	1.800	−0.079	0.066	0.072	0.832
a_{44}	2.600	−0.072	0.087	0.089	0.874
a_{45}	3.300	−0.058	0.105	0.113	0.958
a_{46}	4.000	−0.042	0.139	0.145	0.958
b_4	0.400	0.282	0.035	0.042	0.000
β_{10}	0.400	0.004	0.060	0.069	0.974
β_{11}	−0.500	0.139	0.077	0.080	0.613
ρ	0.400	0.359	0.017	0.023	0.000
σ_u	1.300	−0.130	0.026	0.041	0.052

References

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**, 579–602.

Baghfalaki, T., Ganjali, M. and Berridge, D. (2013). Robust joint modeling of longitudinal measurements and time to event data using normal/independent distributions: a Bayesian approach. *Biometrical Journal* **55**, 844–865.

Bandyopadhyay, S., Ganguli, B. and Chatterjee, A. (2011). A review of multivariate longitudinal data analysis. *Statistical Methods in Medical Research* **20**, 299–330.

Bushnell, D. M. and Martin, M. L. (1999). Quality of life and Parkinson's disease: translation and validation of the US Parkinson's disease questionnaire (PDQ-39). *Quality of Life Research* **8**, 345–350.

Carlin, B. P. and Louis, T. A. (2011). *Bayesian Methods for Data Analysis*. Chapman & Hall, Boca Raton, FL.

Chen, M. H., Shao, Q. M. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer Series in Statistics, New York, NY.

Cummings, J. L. (1992). Depression and Parkinson's disease: a review. *The American Journal of Psychiatry* **149**, 443–454.

De Jong, M. G., Steenkamp, J.-B. E. and Fox, J.-P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research* **34**, 260–278.

De Jong, M. G., Steenkamp, J.-B. E., Fox, J.-P. and Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: a global investigation. *Journal of Marketing Research* **45**, 104–115.

Douglas, J. A. (1999). Item response models for longitudinal quality of life data in clinical trials. *Statistics in Medicine* **18**, 2917–2931.

Dunson, D. D. (2007). Bayesian methods for latent trait modelling of longitudinal data. *Statistical Methods in Medical Research* **16**, 399–415.

Elm, J. J. and The NINDS NET-PD Investigators (2012). Design innovations and baseline findings in a long-term Parkinson's trial: The National Institute of Neurological Disorders and Stroke exploratory trials in Parkinson's Disease Long-Term study–1. *Movement Disorders* **27**, 1513–1521.

Escobar, M. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277.

Fahn, S., Oakes, D., Shoulson, I., Kieburtz, K., Rudolph, A., Lang, A., Olanow, C., Tanner, C. and Marek, K. (2004). Levodopa and the progression of Parkinson's disease. *The New England Journal of Medicine* **351**, 2498–2508.

Fox, J. P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. Springer, New York, New York.

Geisser, S. (1993). *Predictive Inference: An Introduction*, volume 55. CRC Press, Boca Raton, FL.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman Hall, London, UK.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–534.

Glas, C. A., Geerlings, H., van de Laar, M. A. and Taal, E. (2009). Analysis of longitudinal randomized clinical trials using item response models. *Contemporary Clinical Trials* **30**, 158–170.

He, B. and Luo, S. (2013). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research*, doi: 10.1177/0962280213480877

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

Huang, P., Goetz, C. G., Woolson, R. F., Tilley, B., Kerr, D., Palesch, Y., Elm, J., Ravina, B., Bergmann, K. J. and Kieburtz, K. (2009). Using global statistical tests in long-term Parkinson's disease clinical trials. *Movement Disorders* **24**, 1732–1739.

Huang, P., Tilley, B. C., Woolson, R. F. and Lipsitz, S. (2005). Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens–Fisher problem. *Biometrics* **61**, 532–539.

Jasra, A., Holmes, C. and Stephens, D. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science* **20**, 50–67.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement* **38**, 79–93.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

Lachin, J. M. (2000). Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials* **21**, 167–189.

Lachos, V. H., Bandyopadhyay, D. and Dey, D. K. (2011). Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. *Biometrics* **67**, 1594–1604.

Lachos, V. H., Castro, L. M. and Dey, D. K. (2013). Bayesian inference in nonlinear mixed-effects models using normal independent distributions. *Computational Statistics and Data Analysis* **64**, 237–252.

Lachos, V. H., Ghosh, P. and Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* **20**, 303–322.

Lange, K. and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* **2**, 175–198.

Lee, S.-Y. and Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research* **39**, 653–686.

Lewis, S. M. and Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association* **92**, 648–655.

Lin, T. I. and Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference* **137**, 484–495.

Little, R. and Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics* **52**, 1324–1333.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association* **91**, 1219–1227.

Lord, F. M., Novick, M. R. and Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Boston, MA.

Luo, S., Lawson, A. B., He, B., Elm, J. J. and Tilley, B. C. (2012). Bayesian multiple imputation for missing multivariate longitudinal data from a Parkinson's disease clinical trial. *Statistical Methods in Medical Research*, doi: 10.1177/0962280212469358

Luo, S., Ma, J. and Kieburtz, K. D. (2013). Robust Bayesian inference for multivariate longitudinal data by using normal/independent distributions. *Statistics in Medicine* **32**, 3812–3828.

Maier, K. S. (2001). A rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics* **26**(3), 307–330.

McCulloch, C. E. and Neuhaus, J. M. (2011). Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics* **67**(1), 270–279.

McRae, C., Diem, G., Vo, A., O'Brien, C. and Seeberger, L. (2000). Schwab & England: standardization of administration. *Movement Disorders* **15**, 335–336.

Miller, T. M., Balsis, S., Lowe, D. A., Benge, J. F. and Doody, R. S. (2012). Item response theory reveals variability of functional impairment within clinical dementia rating scale stages. *Dementia and Geriatric Cognitive Disorders* **32**, 362–366.

Müller, J., Wenning, G., Jellinger, K., McKee, A., Poewe, W. and Litvan, I. (2000). Progression of Hoehn and Yahr stages in Parkinsonian disorders: a clinicopathologic study. *Neurology* **55**, 888–891.

Mungas, D. and Reed, B. R. (2000). Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine* **19**, 1631–1644.

Parkinson Study Group (1989). DATATOP: a multicenter controlled clinical trial in early Parkinson's disease. *Archives of Neurology* **46**(10), 1052–1060.

Parkinson Study Group (1993). Effects of tocopherol and deprenyl on the progression of disability in early Parkinson's disease. *The New England Journal of Medicine* **328**, 176–183.

Reise, S. P. and Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology* **5**, 27–48.

Rosa, G., Padovani, C. R. and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal* **45**, 573–590.

Samejima, F. (1997). *Graded Response Model*. Springer, New York, NY.

Snitz, B. E., Yu, L., Crane, P. K., Chang, C.-C. H., Hughes, T. F. and Ganguli, M. (2012). Subjective cognitive complaints of older adults at the population level: an item response theory analysis. *Alzheimer Disease and Associated Disorders* **26**, 344–351.

Thenganatt, M. A. and Jankovic, J. (2014). Parkinson disease subtypes. *JAMA Neurology* **71**, 499–504.

Vaccarino, A. L., Anderson, K., Borowsky, B., Duff, K., Giuliano, J., Guttman, M., Ho, A. K., Orth, M., Paulsen, J. S., Sills, T., van Kammen, D. P., Evans, K. R. and PREDICT–HD and REGISTRY Investigators Coordinators (2011). An item response analysis of the motor and behavioral subscales of the unified Huntington's disease rating scale in Huntington disease gene expansion carriers. *Movement Disorders* **26**, 877–884.

Wang, C., Douglas, J. and Anderson, S. (2002). Item response models for joint analysis of quality of life and survival. *Statistics in Medicine* **21**, 129–142.

Weisscher, N., Glas, C. A., Vermeulen, M. and De Haan, R. J. (2010). The use of an item response theory–based disability item bank across diseases: accounting for differential item functioning. *Journal of Clinical Epidemiology* **63**, 543–549.

