COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2017, VOL. 0, NO. 0, 1–19
https://doi.org/10.1080/03610918.2017.1315730

Taylor & Francis
Taylor & Francis Group

Check for updates

# Bayesian hierarchical joint modeling using skew-normal/independent distributions

Geng Chen[a] and Sheng Luo[b]

[a]Clinical Statistics, GlaxoSmithKline, Collegeville, PA, USA; [b]Department of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

**ABSTRACT**

The multiple longitudinal outcomes collected in many clinical trials are often analyzed by multilevel item response theory (MLIRT) models. The normality assumption for the continuous outcomes in the MLIRT models can be violated due to skewness and/or outliers. Moreover, patients' follow-up may be stopped by some terminal events (e.g., death or dropout), which are dependent on the multiple longitudinal outcomes. We proposed a joint modeling framework based on the MLIRT model to account for three data features: skewness, outliers, and dependent censoring. Our method development was motivated by a clinical study for Parkinson's disease.

**ARTICLE HISTORY**
Received 28 May 2016
Accepted 31 March 2017

**KEYWORDS**
Clinical trial; Item-response theory; Latent variable; MCMC; Parkinson's disease

**MATHEMATICAL SUBJECT CLASSIFICATION**
62; 62F15

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

## 1. Introduction

Many clinical trials collect information on multiple longitudinal outcomes. Due to the characteristics of the disease and symptoms, the outcome measurements can be of mixed types, for example, binary, ordinal, and continuous. Clinical studies on Parkinson's disease (PD), for example, are a good representation of the case. PD is a chronic progressive neurodegenerative disease. Symptoms such as tremors, rigidity, slow movements, and loss of cognitive function can often be observed (Fahn et al., 2004). The multidimensional nature of the disease excludes the method of using a single outcome to summarize or represent the overall disease severity (Huang et al., 2005; Luo et al., 2016). Therefore, clinical trials searching for neuroprotective treatments to slow down the progression of PD symptoms 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), Tolvaptan Efficacy and Safety in Management of Autosomal Dominant Polycystic Kidney Disease and its Outcomes (TEMPO) study (Parkinson Study Group, 2002), 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).

This kind of multivariate longitudinal data structure has three sources of correlation within and between outcomes of the same patient, that is, inter-source (different measures at the same visit), longitudinal (same measure at different visits), and cross-correlation (different

**CONTACT** Sheng Luo Sheng.T.Luo@uth.tmc.edu Department of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler St., Houston, TX 77030, USA.

Supplemental data for this article can be accessed on the publisher's website.

© 2017 Taylor & Francis Group, LLC

measures at different visits). Many approaches has been developed to analyze data with multiple outcomes. For example, a linear combination of all outcomes, choosing one outcome as primary and other outcomes as secondary, and global statistical tests (GST). Among those methods, the implementation of the linear combination of all outcomes is relatively easy. However, it reduces the information substantially and the interpretation maybe 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 approach that assumes all outcomes are the clinical manifestation of the unobservable latent variables (Mungas and Reed, 2000). The multilevel item response theory (MLIRT) models, which is based on the latent variable approach, have been increasingly used (Douglas, 1999; Glas et al., 2009; He and Luo, 2016). The MLIRT model consists of two levels of models. The first level of the MLIRT model describes the outcome measurements as functions of the subject-specific disease severity (the latent variable) and measurement-specific parameters. In the second level model, the latent variable is regressed on the covariates of interest and the subject-specific random effects. Some of the advantages of the MLIRT model include: (1) it uses the full longitudinal information; and (2) it explicitly accounts for all three sources of correlations via the subject-specific random effects; (3) it has a better reflection of the multilevel data structure; and (4) it simultaneously estimates the measurement-specific parameters, the covariate effects, as well as the subject-specific disease progression characteristics (Kamata, 2001).

A normal distribution is usually assumed for the continuous outcomes in the MLIRT models. However, the parameter estimations may be biased due to the violation of normality assumptions. The departure from normality may be due to skewness, outliers, or both (Lachos et al., 2011). Common approaches of handling non-normal data include elimination of outliers (or influential data points) and data transformation. Elimination of outliers may not be appropriate in many efficacy assessments in clinical trial studies in order to follow the intent-to-treat (ITT) principle (Little and Yau, 1996). Data transformation may reduce information and the transformation may vary with different datasets (Bandyopadhyay et al., 2010; Lachos et al., 2011). Therefore, it is essential to replace the normality assumption with a more flexible distribution that accounts for the skewness and outliers in the continuous outcomes and produces robust parameter estimates. The skew-normal/independent (SNI) distribution is an attractive class of asymmetric heavy-tailed distributions. There has been a considerable size of literature using the SNI distributions in the mixed model framework (Azzalini and Capitanio, 2003; Bandyopadhyay et al., 2010, 2012; Lachos et al., 2009, 2010; Sahu et al., 2003). However, there is no literature discussing the SNI distribution in the MLIRT modeling framework.

The follow-up in longitudinal clinical studies may be stopped by terminal events, which are either noninformative (administrative censoring such as study termination) or informative (e.g., death or dropout due to disease progression). When the terminal events are related to patients' disease conditions, the unobserved outcomes are non-ignorable. The dependent terminal event time is commonly termed as informative censoring or dependent censoring. It has been shown that ignoring dependent censoring leads to biased estimates (Henderson et al., 2000). Joint modeling of the dependent terminal event time and the longitudinal outcomes provide consistent estimates (Henderson et al., 2000). In the current context,

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

the joint modeling is essential not only to correct the bias in the parameter estimates in the MLIRT model, but also to answer the research question such as "do patients with worse PD severity and/or faster disease progression rate tend to have a terminal event earlier?" Therefore, in this article, we propose a robust Bayesian parametric joint MLIRT model that accounts for both dependent censoring and the violation of the normality assumption due to skewness and outliers. The modeling framework consists of an MLIRT sub-model with an SNI distribution for the multivariate longitudinal data and a Cox proportional hazard sub-model for the dependent censoring event. Two sub-models are linked by a shared random effects vector.

The rest of the article proceeds as follow. In Section 2, we describe a clinical trial that motivated the method development. In Sections 3, we discuss the MLIRT models, the SNI distributions, the Bayesian inference, as well as various model selection criteria. In Section 4, we present an extensive simulation study comparing the performance of various models. In Section 5, we apply our proposed method to the DATATOP study dataset. Section 6 summarizes the main findings and discusses our future research.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

## 2. A motivating example

The method development was motivated by the Deprenyl And Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) study. The DATATOP study was a double-blind, placebo-controlled, multi-center clinical trial with $2 \times 2$ factorial design to determine if deprenyl 10 mg/d and/or tocopherol (vitamin E) 2000 IU/d administered to patients with early PD will delay the time until the initiation of levodopa therapy. Thus, the terminal event was the initiation of levodopa therapy. A total of 800 eligible patients were enrolled 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 to use the levodopa therapy (Parkinson Study Group, 1993). In this article, we investigate the effect of deprenyl and 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). The longitudinal outcomes are the Unified Parkinson's Disease Rating Scale (UPDRS), the Hoehn and Yahr scale (HY), and the Schwab and England activities of daily living (SEADL). Data were collected at baseline, month 1, every three months from month 3 to month 24. The UPDRS total score is approximated by a continuous variable with integer value from 0 (not affected) to 176 (most severely affected) that evaluates patients' mentation, behavior, and activities of daily life (Bushnell and Martin, 1999). The HY variable is an ordinal outcome that measures the disability level in daily activities, the score ranging from 1 to 5 with higher values indicating worse conditions (Müller et al., 2000). The SEADL variable assesses patients' daily activities and it is an ordinal variable with integer values ranging from 0 to 100 incrementing by 5, with larger values indicating better clinical conditions (McRae et al., 2000).

In the DATATOP study, there were 153 and 223 patients in treatment and placebo groups, respectively, initiated the levodopa therapy during the follow-up. The levodopa therapy provided temporary relief of PD symptoms and may significantly change the outcomes for a short period. For this reason, only the outcomes before the initiation of levodopa therapy can be used to evaluate the treatment efficacy. Figure 1 shows the mean UPDRS values for DATATOP patients with follow-up time less than 9 month (solid line), 9 to 15 month (dashed line), and more than 15 month (dotted line), respectively. Patients with shorter follow-up time have higher (worse) UPDRS values, suggesting strong association between the longitudinal outcomes and time to the initiation of levodopa therapy.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

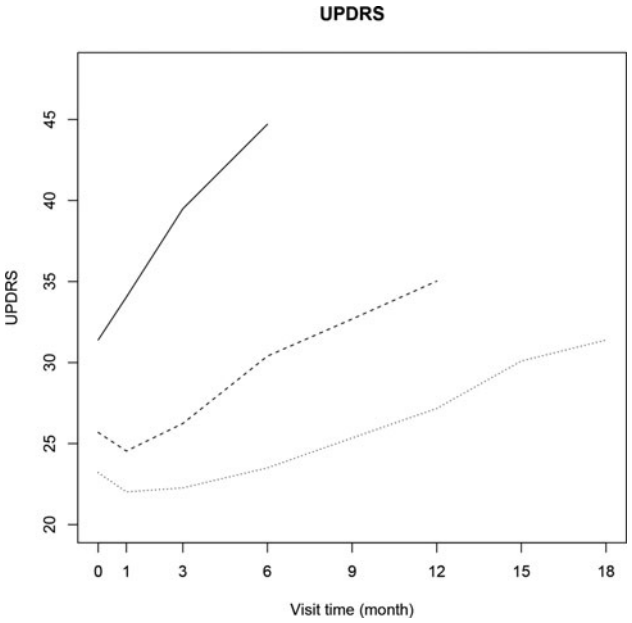


**Figure 1.** Mean UPDRS values for patients with follow-up time less than 9 months (solid line), 9 to 15 months (dashed line), and more than 15 months (dotted line).

Figure 2 displays the histograms and the Q-Q plots of the residuals of the continuous outcome UPDRS by fitting the MLIRT model (6) (with normal assumption for the continuous outcome UPDRS) to the DATATOP dataset. Figure 2 suggests the presence of skewness, heavy tails, outliers in the outcome UPDRS. So in the following sections, we will develop a joint modeling framework that accounts for all three data features, that is, skewness, outliers, and dependent censoring.

## 3. The model and estimation

### 3.1. The multilevel item response theory (MLIRT) model

Let $y_{ijk}$ be 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. Outcome $y_{ijk}$ can be binary, ordinal, or continuous and it is recoded

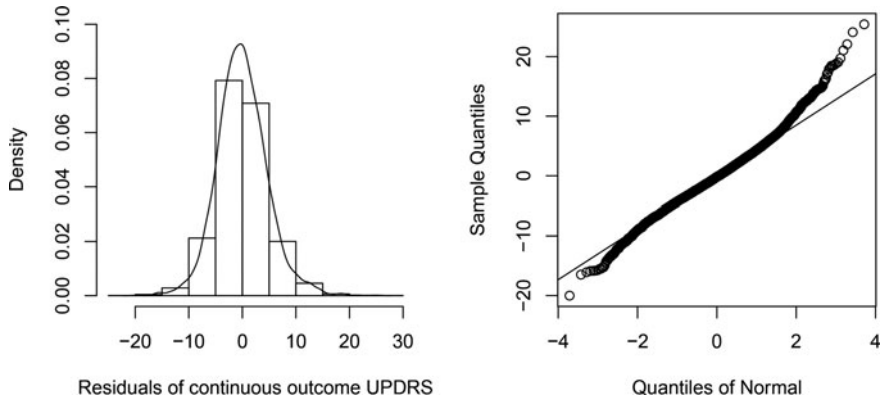


**Figure 2.** Histogram and Q-Q plot of residuals of the continuous outcome UPDRS obtained by fitting a joint MLIRT model (6) (with normal assumption for outcome UPDRS).

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

such that larger values indicate worse medical conditions. Let $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijk}, \ldots, y_{ijK})'$ be the vector of observations for patient $i$ at visit $j$ and let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{ij_i})'$ be the outcome matrix across visits. Let $\theta_{ij}$ be the continuous latent variable denoting the underlining disease severity for patient $i$ at visit $j$ with higher value indicating more severe status. The multilevel item response theory (MLIRT) model consists of two levels. In the first level measurement model, we model the outcomes as functions of $\theta_{ij}$ and subject-specific parameters. Specifically, we model the binary outcomes using a two-parameter sub-model (Fox, 2010), the cumulative probabilities of ordinal outcomes using graded response sub-model (Samejima, 1997), and the continuous outcomes using common factor sub-model (Lord et al., 1968) as follows.

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

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

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

where random error $\epsilon_{ijk} \sim N(0, \sigma_k^2)$ with $\sigma_k^2$ being variance of continuous outcome $k$, $a_k$ is the outcome-specific "difficulty" parameter, and $b_k$ is the "discriminating" parameter that is always positive and describes the intensity to which outcome $k$ discriminates between patients with latent disease severity $\theta_{ij}$. 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 $a_{k1} < \cdots < a_{kl} < \cdots < a_{kn_{k-1}}$. The probability that patient $i$ being in category $l$ on outcome $k$ at visit $j$ is defined as $p(Y_{ijk} = l|\theta_{ij}) = p(Y_{ijk} \le l|\theta_{ij}) - p(Y_{ijk} \le l - 1|\theta_{ij})$.

In the second level structural model, the latent disease severity $\theta_{ij}$ is regressed on covariates 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 and they may share all or part of the covariates, $u_{i0}$ is the random intercept that represents the subject-specific disease severity and $u_{i1}$ is the random slope that represents the subject-specific disease progression rate. Let $\boldsymbol{u}_i = (u_{i0}, u_{i1})'$ and we assume $u_{i0} \sim N(0, 1)$, $u_{i1} \sim N(0, \sigma_u^2)$, and $\text{corr}(u_{i0}, u_{i1}) = \rho$. It is widely known that the item-response models are over-parameterized (Samejima, 1997) and additional constraints need to be imposed to make the models identifiable. To this end, we impose the constraint $\text{Var}(u_{i0}) = 1$. Let $t_i$ denote the time to a terminal event for patient $i$, $\zeta_i$ (1 if the terminal event is observed and 0 if not) denote the censoring indicator for $t_i$. Let $\boldsymbol{X}_i$ denote the vector of potential risk factors. Then the Cox proportional hazard model for $t_i$, conditional on $\boldsymbol{u}_i$, can be written as

$$h(t_i|\boldsymbol{u}_i) = h_0(t_i)\exp(\boldsymbol{X}_i\boldsymbol{\gamma} + \eta_0 u_{i0} + \eta_1 u_{i1}), \tag{5}$$

where $\boldsymbol{\gamma}$ is the unknown parameter vector for the potential risk factors $\boldsymbol{X}_i$, $\eta_0$ and $\eta_1$ measure the strength of the association between the Cox proportional hazard model and the MLIRT model. Shared random effects $u_{i0}$ and $u_{i1}$ are used to account for the correlation between the survival time and longitudinal outcomes. When the parameters $\eta_0$ and $\eta_1$ are not equal to zero, the correlation between the survival time and longitudinal outcomes is incorporated. Specifically, positive parameters $\eta_0$ and $\eta_1$ indicate that the patients with more severe disease severity and faster disease progression rate tend to have higher risk of initiating levodopa therapy, and vice versa. We adopt a piecewise constant function to approximate the baseline hazard function $h_0(t)$ and assume the hazard rate is constant within a given time period (Lawless and Zhan, 1998). Given a set of time points $0 = \tau_0 < \tau_1 < \cdots < \tau_m$ and the baseline hazard vector $\boldsymbol{g} = (g_0, g_1, \cdots, g_{m-1})'$, the piecewise constant hazard function can be approximated by

$h_0(t) = \sum_{l=0}^{m-1} g_l I_l(t)$, where $I_l(t)$ is an indicator function with $I_l(t) = 1$ if $\tau_l \le t < \tau_{l+1}$ and 0 otherwise.

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

$$L(\boldsymbol{y}_i, \boldsymbol{u}_i) = \left[ \prod_{j=1}^{J_i} \prod_{k=1}^{K} p(y_{ijk}|\boldsymbol{u}_i) \right] \cdot h(t_i|\boldsymbol{u}_i)^{\xi_i} S(t_i|\boldsymbol{u}_i) \cdot p(\boldsymbol{u}_i), \tag{6}$$

where the survival function $S(t_i|\boldsymbol{u}_i) = \exp[-\int_0^{t_i} h(s|\boldsymbol{u}_i)ds]$ and $p(\boldsymbol{u}_i)$ is the density function for random effects vector $\boldsymbol{u}_i$. 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 let $\boldsymbol{\beta} = (\boldsymbol{\beta_0}', \boldsymbol{\beta_1}')'$. Because we assume normal random errors for the continuous outcomes in model (3), we refer to this joint modeling framework as model $\text{JM}_\text{N}$ with the parameter vector $\boldsymbol{\Phi} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k, \boldsymbol{\gamma}', \eta_0, \eta_1, \boldsymbol{g}')'$. We refer to as reduced model $\text{RM}_\text{N}$, the model assuming that the survival time is independent to the longitudinal outcomes (i.e., $\eta_0 = \eta_1 = 0$). The parameter vector of model $\text{RM}_\text{N}$ is $\boldsymbol{\Phi} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k, \boldsymbol{\gamma}', \boldsymbol{g}')'$.

### 3.2. The skew-normal/independent (SNI) distributions

The skew-normal/independent (SNI) distribution is an attractive class of asymmetric heavy-tailed distributions that includes skew-t (ST), skew-slash, and skew-contaminated normal distributions. In this article, we adopt the ST distribution, while the application of other SNI distributions is straightforward. It is of note that the ST distribution reduces to the skew-normal distribution (SN, asymmetric but not heavy-tailed) when degree of freedom is large. We consider the ST and SN distributions introduced by Sahu et al. (2003), which have the stochastic representation and are suitable for a Bayesian computation and briefly discussed below.

For simplicity, we illustrate the implementation of the univariate ST distribution in one continuous outcome $k$ in model (3). Let $\delta$ be the skewness parameter for continuous outcome $y_{ijk}$, and $\nu$ be the degree of freedom for the ST distribution, then following Sahu et al. (2003), the stochastic representation of the ST distribution for $y_{ijk}$ is given by

$$\begin{aligned} y_{ijk}|a_k, b_k, \theta_{ij}, \sigma_k^2, \omega_{ijk}, z_{ik}, \delta, \nu &\sim N(a_k + b_k\theta_{ij} + \delta z_{ik},\ \sigma_k^2/\omega_{ijk}), \\ z_{ik} &\sim N(0,1)I(z_{ik} > 0), \\ \delta &\sim N(0, \Gamma), \\ \omega_{ijk} &\sim \text{Gamma}(\nu/2, \nu/2), \end{aligned}$$

where the weight variable $\omega_{ijk}$ is a positive random variable with density $p(\omega_{ijk}|\nu)$, where the parameter $\nu > 0$ represents degree of freedom, $z_{ik}$ is a subject-specific variable for outcome $k$ that follows a truncated standard normal distribution. The skewness parameter $\delta$ indicates the skewness of outcome $k$, with positive $\delta$ representing a right skewed distribution and negative $\delta$ representing a left skewed distribution. The parameter $\Gamma$ determines the prior variance information for $\delta$. It is of note that when the degree of freedom $\nu \to \infty$, the distribution $\text{Gamma}(\nu/2, \nu/2)$ degenerates to 1, that is, $\omega_{ijk} \equiv 1$. In this case, the ST distribution reduces to the SN distribution. Moreover, when the skewness parameter $\delta = 0$, the ST distribution reduces to the symmetric and heavy-tailed student-$t$ distribution. In practice, the parameters

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

$\nu$ and $\delta$ can be estimated from the data and small $\nu$ and large $\delta$ in absolute value are indications of heavy tails (outliers) and skewness, respectively.

Incorporating the ST distribution, the full likelihood in (6) becomes

$$L(\boldsymbol{y}_i, \boldsymbol{u}_i, \omega_{ijk}, z_{ik}, \delta) = \prod_{j=1}^{J_i} \left[ \prod_{k=1}^{K} p(y_{ijk}|\boldsymbol{u}_i, \omega_{ijk}, z_{ik}, \delta) p(\omega_{ijk}) p(z_{ik}) \right] h(t_i|\boldsymbol{u}_i)^{\xi_i} S(t_i|\boldsymbol{u}_i) p(\boldsymbol{u}_i). \tag{7}$$

We denote it as model $\text{JM}_{\text{ST}}$ with the parameter vector $\boldsymbol{\Phi} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k, \boldsymbol{\gamma}', \eta_0, \eta_1, \boldsymbol{g}', \nu, \delta)'$. When we assume that the longitudinal outcomes and survival time are independent ($\eta_0 = \eta_1 = 0$), we denote the corresponding reduced model as model $\text{RM}_{\text{ST}}$ with the parameter vector $\boldsymbol{\Phi} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k, \boldsymbol{\gamma}', \boldsymbol{g}', \nu, \delta)'$. In addition, we denote the models using either $t$ or SN distributions under either joint or reduced model frameworks as models $\text{JM}_{\text{T}}$, $\text{JM}_{\text{SN}}$, and $\text{RM}_{\text{N}}$, respectively.

In summary, we consider the following models that account for various combinations of the three data features, that is, skewness, outliers, and dependent censoring.

$\text{JM}_{\text{ST}}$: accounts for all three features;
$\text{JM}_{\text{SN}}$: accounts for skewness and dependent censoring, but not outliers;
$\text{JM}_{\text{T}}$: accounts for outliers and dependent censoring, but not skewness;
$\text{JM}_{\text{N}}$: accounts for dependent censoring, but not skewness and outliers;
$\text{RM}_{\text{ST}}$: accounts for skewness and outliers, but not dependent censoring;
$\text{RM}_{\text{T}}$: accounts for outliers, but not skewness and dependent censoring;
$\text{RM}_{\text{N}}$: does not account for any of the three features.

### 3.3. Bayesian inference and model selection criteria

To infer the unknown parameter vector $\boldsymbol{\Phi}$, we use Bayesian inference based on Markov chain Monte Carlo (MCMC) posterior simulations. The model fitting is performed using the JAGS language (JAGS version 3.4.0). We assume vague priors on all parameters in $\boldsymbol{\Phi}$. Specifically, the prior distribution of all parameters in $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\eta_0$, $\eta_1$ are $N(0, 100)$. We use the prior distribution Gamma(0.001, 0.001) for all components in $\boldsymbol{b}$ and $\sigma_u$ to ensure positivity, and use Uniform$[-1, 1]$ for $\rho$. The prior distribution for the difficulty parameter $a_k$ of the continuous outcomes is $a_k \sim N(0, 100)$. For the ordinal outcomes, we let $a_{k1} \sim N(0, 100)$, and $a_{kl} = a_{k,l-1} + \delta_l$ for $l = 2, n_k - 1$ with $\delta_l \sim N(0, 100)I(0,)$ (normal distribution left truncated at 0). For the parameters related to the ST distributions, we use Gamma(0.001, 0.001) for $\nu$ and $\delta$. We have investigated other selections of vague prior distributions with different hyper-parameters and have obtained very similar results in both the simulation study and data application. Multiple chains with dispersed initial values are run. We use the history plots and view the absence of apparent trend in the plots as evidence of convergence. We also 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 JAGS code for fitting model $\text{JM}_{\text{ST}}$ has been posted in the web supplement.

Among the various model selection methods available in Bayesian inference, we select the following six: the log pseudo-marginal likelihood (LPML), deviance information criterion (DIC), expected Akaike information criterion (EAIC), expected Bayesian information criterion (EBIC), Watanabe-Akaike information criterion (WAIC), and Bayes factor (BF).

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

Conditional predictive ordinate (CPO) (Carlin and Louis, 2011) is a cross-validation predictive method that evaluates the predictive distribution of the model conditioning on the data but with one subject deleted. Let $\boldsymbol{y}$ be the full observed data and $\boldsymbol{y}_{(i)}$ be the data with subject $i$ deleted. Then the CPO for subject $i$ is defined as $\text{CPO}_i = p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}) = \int p(\boldsymbol{y_i}|\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\boldsymbol{y}_{(i)})d\boldsymbol{\Phi}$. Large CPO suggests that the data for subject $i$ can be well predicted by the model based on the data from all other subjects. Thus, lager CPO indicates a better model fit. Because there is no close form for $\text{CPO}_i$ for our proposed models, we adopt a Monte Carlo estimation obtained from the posterior distribution $p(\boldsymbol{\Phi}|\boldsymbol{y})$. Because the function of $\text{CPO}_i$ can be further derived as $\text{CPO}_i = p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}) = p(\boldsymbol{y})/p(\boldsymbol{y}_{(i)}) = 1/\int p(\boldsymbol{\Phi}|\boldsymbol{y})/p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}, \boldsymbol{\Phi})d\boldsymbol{\Phi}$, a harmonic-mean approximation of $\text{CPO}_i$ is $\widehat{\text{CPO}_i} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)},\boldsymbol{\Phi}^{(t)})})^{-1} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(\boldsymbol{y_i}|\boldsymbol{\Phi}^{(t)})})^{-1}$, where $M$ is the total number of post burn-in samples (Dey et al., 1997). A summary statistics of $\text{CPO}_i$ is log pseudo-marginal likelihood (LPML), defined as $\text{LPML} = \sum_{i=1}^{N}\log(\widehat{\text{CPO}_i})$. A larger value of LPML suggests a better model fit.

The deviance information criterion (DIC) assesses model fit based on the posterior mean of the deviance and a penalty on the model complexity (Spiegelhalter et al., 2002). A smaller value of DIC indicates better model fit. In addition, the expected Akaike information criterion (EAIC) and the expected Bayesian information criterion (EBIC) (Carlin and Louis, 2011), which add penalty on the total number of parameters in the model and the sample size, respectively, were also used to evaluate the model performance. Smaller values of EAIC and EBIC suggest better model fit. Moreover, WAIC is an approximation of cross-validation using full Bayesian method. WAIC is computed based on the pointwise posterior predictive density and adjusted for the effective number of parameters (Gelman et al., 2014). Smaller values of WAIC suggest better model fit.

Furthermore, Bayes factor (BF) is a standard Bayesian solution (an alternative to $p$ value in the frequentist framework) to the hypothesis testing for competing models. The BF quantifies the degree to which the observed data support a hypothesis (Lewis and Raftery, 1997):

$$\text{BF}(M_1; M_2) = \frac{f(\boldsymbol{y}|M_1)}{f(\boldsymbol{y}|M_2)} = \frac{\int f(\boldsymbol{y}|\boldsymbol{\Phi_1}, M_1)f(\boldsymbol{\Phi_1}|M_1)d\boldsymbol{\Phi_1}}{\int f(\boldsymbol{y}|\boldsymbol{\Phi_2}, M_2)f(\boldsymbol{\Phi_2}|M_2)d\boldsymbol{\Phi_2}},$$

where $\boldsymbol{\Phi}_l$ is the parameter vectors for model $M_l$ for $l = 1, 2$; $f(\boldsymbol{y}|\boldsymbol{\Phi}_l, M_l)$ is the likelihood of model $M_l$; and $f(\boldsymbol{\Phi}_l|M_l)$ is the posterior density of $\boldsymbol{\Phi}_l$ for model $M_l$. The direct computation of the integration involved in the BF is not straightforward, so the Laplace–Metropolis estimator based on normal distribution is used to approximate the marginal likelihood $f(\boldsymbol{y}|M_l)$. Specifically, the $f(\boldsymbol{y}|M_l) \approx (2\pi)^{d_l/2}|\boldsymbol{\Sigma}_l|^{1/2}f(\boldsymbol{y}|\bar{\boldsymbol{\Phi}}_l, M_l) \cdot f(\bar{\boldsymbol{\Phi}}_l|M_l)$, where $d_l$ is the number of parameters in $\boldsymbol{\Phi}_l$, $\boldsymbol{\Sigma}_l$ is the posterior covariance matrix of $\boldsymbol{\Phi}_l$, $\bar{\boldsymbol{\Phi}}_l$ is the posterior mean of $\boldsymbol{\Phi}_l$, $f(\bar{\boldsymbol{\Phi}}_l|M_l)$ is the prior probability of parameters evaluated at $\bar{\Phi}_l$, and $f(\boldsymbol{y}|\bar{\boldsymbol{\Phi}}_l, M_l)$ is the likelihood evaluated at the posterior mean $\bar{\Phi}_l$. The interpretation of the BF is summarized by Kass and Raftery (1995). In particular, when BF is greater than 100, decisive evidence is shown in favor of Model $M_1$ over $M_2$.

## 4. Simulation studies

In this section, we conducted an extensive simulation study with two settings to compare the performance of the following models: $\text{JM}_{\text{ST}}$, $\text{JM}_{\text{SN}}$, $\text{JM}_{\text{T}}$, $\text{JM}_{\text{N}}$, $\text{RM}_{\text{ST}}$, $\text{RM}_{\text{T}}$, and $\text{RM}_{\text{N}}$. Note that the method proposed by Chen and Luo (2016) considers outliers problem but ignored two data features, that is, skewness and dependent censoring, and it is essentially model $\text{RM}_{\text{T}}$ as illustrated in Section 3.2 In both settings, we generated 400 datasets with a sample size of

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

400 patients (200 in both treatment and placebo groups). The data structure was similar to the motivating dataset, and it had one continuous outcomes and two ordinal outcomes (both with seven categories) at five visits (months 0, 1, 3, 9, 15).

We generated data based on the following model $\theta_{ij} = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0 + u_{i0} + (\boldsymbol{X}_{i1}\boldsymbol{\beta}_1 + u_{i1})t_{ij}$, and $h(t_i|\boldsymbol{u}_i) = h_0(t_i)\exp(\boldsymbol{X}_i\boldsymbol{\gamma} + \eta_0 u_{i0} + \eta_1 u_{i1})$. We set $\boldsymbol{X}_{i0} = 0$ and $\boldsymbol{X}_{i1} = x_i$, where the covariate $x_i$ took value 0 or 1 each with probability 0.5 to mimic treatment assignment. We set the coefficients to be $\boldsymbol{\beta} = (\beta_{10}, \beta_{11})' = (0.4, -0.5)'$ and $\gamma = -1$. The parameters for the continuous outcomes were set to be $a_1 = 25$, $b_1 = 10$, and $\sigma_1 = 5$. The parameters for the ordinal outcomes were set to be $\boldsymbol{a}_2 = (-2.7, -0.6, 2, 2.8, 5, 6)$, $b_2 = 2$, $\boldsymbol{a}_3 = (-0.1, 1, 1.8, 2.6, 3.3, 4)$, $b_3 = 0.4$. We assumed that the subject-specific random effects vector $\boldsymbol{u}_i = (u_{i0}, u_{i1})' \sim N_2(0, \boldsymbol{\Sigma}_u)$, where $\boldsymbol{\Sigma}_u = \{(1, \rho\sigma_u), (\rho\sigma_u, \sigma_u^2)\}$ with $\rho = 0.4$ and $\sigma_u = 1.3$. The baseline hazard $h_0$ was set to 0.5 and $\eta_0 = 0.4$ and $\eta_1 = 1$. The independent censoring time was sampled from uniform(10, 20). We applied the Bayesian framework in Section 3.2 and we ran two MCMC chains with dispersed initial values. Each MCMC chain was run for 20,000 iterations with the first 10,000 iterations discarded as burn-in. We computed the average of the posterior mean minus the true values (Bias), the square root of the average of the posterior variance (SE), the standard deviation of the posterior means (SD), and coverage probabilities of the 95% equal-tail credible intervals (CP).

In setting I, we evaluated the model performance when both skewness and outliers exist in the continuous outcome. To generate the skewness, we first generated one continuous outcome from a skew-normal distribution with mean $\mu_{ij} = a_1 + b_1\theta_{ij}$, $\sigma_1 = 5$, and skewness parameter $\delta = 7$. Then, to generate outliers, we randomly selected 5% of the data points from the continuous outcome and replaced them with noise from either uniform$[-\mu - 4\sigma_1, \mu - 2\sigma_1]$ or uniform$[\mu + 2\sigma_1, \mu + 4\sigma_1]$ with equal probabilities.

Due to the space restriction, in Table 1, we only presented the results from the four joint models, that is, $JM_N$, $JM_T$, $JM_{SN}$, and $JM_{ST}$, while the results from the reduced models $RM_N$, $RM_T$, and $RM_{ST}$ can be found in the Web Table 1. Among all the models, $JM_{ST}$ provided the best parameter estimates with smallest bias, SE close to SD, and the CPs reasonably close to nominal level of 0.95. The models that ignore skewness and/or outliers (models $JM_N$, $JM_T$, and $JM_{SN}$) overestimated the difficulty parameter $a_1$ and the discriminating parameter $b_1$, and the CPs were markedly away from the nominal value of 0.95. However, model $RM_{ST}$ provided more reasonable estimates (smaller bias and CPs closer to 95%) to the parameters $a_1$ and $b_1$. All models except model $JM_{ST}$ gave biased estimates and poor CPs to the parameters $\beta_{10}$ (disease progression rate for the patients in the placebo group), $\beta_{11}$ (the change in disease progression rate by the treatment). Comparing to their joint model counterparts (models $JM_N$, $JM_T$, and $JM_{ST}$), the reduced models $RM_N$, $RM_T$, and $RM_{ST}$ provided larger bias on the parameters $\beta_{10}$, $\beta_{11}$, and $\gamma$ (treatment effect on the survival time). This is because that by ignoring the dependent censoring (the subjects with more severe disease severity and faster disease progression tend to have a terminal event earlier), the reduced models tend to reduce the difference between two groups and therefore underestimate the treatment effects in both the longitudinal and survival models. The poor estimates of $\sigma_1$ from all models are not surprising because the variance of the continuous outcome was artificially inflated by replacing 5% of data by large noise. The results in this simulation setting indicated that it was essential to account for the skewness and outliers in the continuous outcome as well as dependent censoring within the MLIRT joint modeling framework to obtain correct estimates in disease progression and treatment effects, in addition to the parameters specific to the continuous outcome.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017



**Table 1.** Setting I: simulation results from models $JM_N$, $JM_T$, $JM_{SN}$, and $JM_{ST}$ when there were skewness and 5% outliers in the continuous outcome.

| | | $JM_N$ | | | | $JM_T$ | | | | $JM_{SN}$ | | | | $JM_{ST}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias | SD | SE | CP | Bias | SD | SE | CP | Bias | SD | SE | CP | Bias | SD | SE | CP |
| $a_1$ | 25.000 | **4.203** | 0.756 | 0.708 | **0.000** | **5.829** | 0.599 | 0.591 | **0.000** | **1.469** | 0.984 | 1.581 | **0.928** | 0.329 | 0.984 | 1.004 | 0.935 |
| $b_1$ | 10.000 | **0.298** | 0.762 | 0.656 | **0.871** | **0.708** | 0.458 | 0.458 | **0.628** | **0.240** | 0.742 | 0.669 | **0.897** | −0.008 | 0.509 | 0.503 | 0.930 |
| $\sigma_1$ | 5.000 | 7.512 | 1.319 | 0.373 | 0.000 | −1.245 | 0.254 | 0.262 | 0.008 | 7.393 | 1.325 | 0.376 | 0.000 | −1.411 | 0.236 | 0.249 | 0.003 |
| $a_{21}$ | −2.700 | 0.081 | 0.197 | 0.182 | 0.894 | 0.198 | 0.170 | 0.162 | 0.761 | 0.064 | 0.195 | 0.183 | 0.917 | −0.031 | 0.191 | 0.184 | 0.935 |
| $a_{22}$ | −0.600 | 0.029 | 0.148 | 0.141 | 0.937 | 0.070 | 0.133 | 0.131 | 0.893 | 0.027 | 0.150 | 0.142 | 0.934 | −0.015 | 0.146 | 0.143 | 0.941 |
| $a_{23}$ | 2.000 | −0.013 | 0.163 | 0.164 | 0.950 | −0.067 | 0.153 | 0.150 | 0.915 | −0.005 | 0.168 | 0.165 | 0.943 | 0.008 | 0.166 | 0.164 | 0.949 |
| $a_{24}$ | 2.800 | −0.012 | 0.191 | 0.188 | 0.939 | −0.092 | 0.176 | 0.170 | 0.913 | −0.004 | 0.197 | 0.189 | 0.920 | 0.025 | 0.190 | 0.187 | 0.941 |
| $a_{25}$ | 5.000 | 0.012 | 0.349 | 0.321 | 0.942 | −0.124 | 0.331 | 0.297 | 0.876 | 0.026 | 0.357 | 0.322 | 0.940 | 0.085 | 0.352 | 0.318 | 0.933 |
| $a_{26}$ | 6.000 | 0.171 | 0.518 | 0.472 | 0.942 | 0.004 | 0.497 | 0.449 | 0.935 | 0.174 | 0.518 | 0.471 | 0.937 | 0.241 | 0.504 | 0.466 | 0.935 |
| $b_2$ | 2.000 | −0.039 | 0.149 | 0.148 | 0.939 | −0.167 | 0.126 | 0.121 | 0.679 | −0.023 | 0.151 | 0.149 | 0.945 | 0.020 | 0.145 | 0.142 | 0.952 |
| $a_{31}$ | −0.100 | 0.014 | 0.074 | 0.075 | 0.945 | 0.020 | 0.074 | 0.074 | 0.938 | 0.013 | 0.073 | 0.075 | 0.943 | 0.008 | 0.072 | 0.075 | 0.955 |
| $a_{32}$ | 1.000 | 0.017 | 0.083 | 0.083 | 0.945 | 0.021 | 0.083 | 0.082 | 0.941 | 0.016 | 0.081 | 0.083 | 0.943 | 0.014 | 0.081 | 0.082 | 0.947 |
| $a_{33}$ | 1.800 | 0.017 | 0.096 | 0.103 | 0.953 | 0.016 | 0.093 | 0.102 | 0.966 | 0.016 | 0.095 | 0.103 | 0.960 | 0.012 | 0.095 | 0.102 | 0.952 |
| $a_{34}$ | 2.600 | 0.038 | 0.142 | 0.140 | 0.953 | 0.040 | 0.144 | 0.139 | 0.949 | 0.040 | 0.142 | 0.140 | 0.957 | 0.037 | 0.143 | 0.140 | 0.949 |
| $a_{35}$ | 3.300 | 0.076 | 0.200 | 0.193 | 0.937 | 0.081 | 0.197 | 0.193 | 0.932 | 0.075 | 0.198 | 0.193 | 0.931 | 0.072 | 0.196 | 0.193 | 0.944 |
| $a_{36}$ | 4.000 | 0.159 | 0.293 | 0.279 | 0.923 | 0.167 | 0.294 | 0.279 | 0.913 | 0.156 | 0.288 | 0.279 | 0.934 | 0.155 | 0.288 | 0.279 | 0.930 |
| $b_3$ | 0.400 | 0.008 | 0.060 | 0.060 | 0.960 | −0.005 | 0.056 | 0.056 | 0.963 | 0.009 | 0.061 | 0.060 | 0.954 | −0.003 | 0.057 | 0.056 | 0.955 |
| $\beta_{10}$ | 0.400 | **−0.150** | 0.226 | 0.220 | **0.897** | **−0.056** | 0.170 | 0.176 | **−0.930** | **−0.142** | 0.223 | 0.222 | **0.902** | −0.025 | 0.189 | 0.186 | 0.938 |
| $\beta_{11}$ | −0.500 | **0.064** | 0.227 | 0.221 | **0.929** | **0.034** | 0.190 | 0.190 | **0.941** | **0.054** | 0.230 | 0.222 | **0.934** | 0.003 | 0.207 | 0.203 | 0.933 |
| $\rho$ | 0.400 | −0.002 | 0.110 | 0.111 | 0.953 | −0.020 | 0.077 | 0.078 | 0.952 | −0.005 | 0.110 | 0.111 | 0.951 | −0.011 | 0.079 | 0.080 | 0.952 |
| $\sigma_u$ | 1.300 | −0.179 | 0.162 | 0.154 | 0.755 | −0.111 | 0.114 | 0.114 | 0.828 | −0.171 | 0.156 | 0.157 | 0.802 | −0.012 | 0.128 | 0.126 | 0.944 |
| $\gamma$ | −1.000 | −0.004 | 0.177 | 0.181 | 0.950 | −0.006 | 0.179 | 0.178 | 0.958 | −0.009 | 0.180 | 0.181 | 0.951 | −0.002 | 0.177 | 0.177 | 0.947 |
| $\eta_1$ | 0.400 | −0.029 | 0.166 | 0.175 | 0.953 | −0.023 | 0.114 | 0.112 | 0.941 | −0.019 | 0.162 | 0.171 | 0.957 | 0.011 | 0.114 | 0.116 | 0.963 |
| $\eta_2$ | 1.000 | 0.229 | 0.238 | 0.231 | 0.815 | 0.130 | 0.126 | 0.132 | 0.845 | 0.212 | 0.197 | 0.217 | 0.816 | 0.022 | 0.117 | 0.122 | 0.975 |

Note: Large bias and poor CP are highlighted in bold.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

In setting II, we generated data from model $JM_N$ with no skewness and outliers in the continuous outcome and presented the results from models $RM_T$, $RM_{ST}$, $JM_N$, and $JM_{ST}$ in Table 2. Due to space constraint, we did not present the outcome-specific parameters $\boldsymbol{a}$ and $\boldsymbol{b}$, but the estimation of all these parameters were reasonably good. Table 2 suggests that in comparison to the results from the true model $JM_N$, the proposed model $JM_{ST}$ produced comparable and satisfactory results, that is, the bias was negligible, SE was close to SD, and the overage probabilities were reasonably close to nominal level of 0.95. The over-parameterized model $JM_{ST}$ is still a reasonable choice. However, models $RM_T$ and $RM_{ST}$ produced misleading results with large bias on all parameters due to the ignorance of dependent censoring.

From the simulation study, when both skewness and outliers existed in the continuous outcome, model $JM_{ST}$ produced accurate parameter estimates while all other models provided severely biased estimates for some parameters. In the absence of skewness and outliers in the continuous outcome, the overparameterized model $JM_{ST}$ provided results comparable to model $JM_N$. Due to the ignorance to skewness and dependent censoring, model $RM_T$, that is, the method proposed by Chen and Luo (2016), produced larger bias on the treatment effect related parameters $\beta_{10}$, $\beta_{11}$, and $\gamma$ under both simulation settings I and II.

## 5. Application to the DATATOP study

In this section, we applied our proposed method to the motivating DATATOP study. We compared the results of four joint models $JM_N$, $JM_T$, $JM_{SN}$, and $JM_{ST}$, as well as the reduced models $RM_N$, $RM_T$, and $RM_{ST}$. The longitudinal outcomes under consideration are UPDRS (continuous), HY (ordinal with five categories), and SEADL (ordinal with six categories). We also reverse-coded the SEADL variable so that all three outcomes have higher value indicating worse clinical condition. We removed one patient that has no UPDRS measurements across all visits. Therefore, there were 398 and 401 patients in the treatment and placebo groups, respectively. And there were 153 and 222 patients in treatment and placebo groups, respectively, experienced the terminal event, that is, initiation of levodopa therapy. We let $\boldsymbol{X}_{i0} = 0$ and considered the treatment assignment as the only covariate in $\boldsymbol{X}_{i1}$. So the second level model (4) became $\theta_{ij} = u_{i0} + (\beta_{10} + \beta_{11}x_i + u_{i1})t_{ij}$. The Cox model for the time to levodopa therapy became $h(t_i|\boldsymbol{u}_i) = h_0(t_i)\exp(\gamma x_i + \eta_0 u_{i0} + \eta_1 u_{i1})$. To specify the baseline hazard $h_0(t_i)$, we divided the survival time into $K = 3$ intervals by every $1/K$th quantiles. Other specification of $K$ were also explored, the results were very similar. We used two parallel MCMC chains with over-dispersed initial values, and ran each chain for 20,000 iterations. The first 10,000 iterations were discarded as burn-in and the inference was based on the remaining 10,000 iterations. Good mixing properties of the MCMC chains for all model parameters were observed in the trace plots. The scale reduction $\widehat{R}$ of all parameter were smaller than 1.1.

Table 3 displays various model comparison criteria obtained from the models under consideration. The proposed model $JM_{ST}$ performed significantly better than all other models with largest LPML value and smallest DIC, EAIC, EBIC, and WAIC values. Hence, we selected model $JM_{ST}$ as the final model. As introduced in Section 3.3, BF is a Bayesian alternative to $p$ value for hypothesis testing for competing models. In our model settings, when $\eta_0 = \eta_1 = 0$, model $JM_{ST}$ reduces to model $RM_{ST}$; when $\nu \to \infty$ or equivalently $\omega_{ijk} = 1$, model $JM_{ST}$ reduces to model $JM_{SN}$; when $\delta = 0$, model $JM_{ST}$ reduces to model $JM_T$; when both $\delta = 0$ and $\nu \to \infty$, model $JM_{ST}$ reduces to model $JM_N$; when $\delta = \eta_0 = \eta_1 = 0$, model $JM_{ST}$ reduces to model $RM_T$; and when $\delta = \eta_0 = \eta_1 = 0$ and $\nu \to \infty$, model $JM_{ST}$ reduces to model $RM_N$. We computed the BFs to compare all the models stated above to our proposed model $JM_{ST}$. The BF results were summarized in Table 3. Decisive evidence (BF $\gg$ 100) was shown in

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017



**Table 2.** Setting II: simulation results from models $RM_T$, $RM_{ST}$, $JM_N$, and $JM_{ST}$ when there were no skewness and outliers in the continuous outcome.

| | | $RM_T$ | | | | $RM_{ST}$ | | | | $JM_N$ | | | | $JM_{ST}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True | Bias | SD | SE | CP | Bias | SD | SE | CP | Bias | SD | SE | CP | Bias | SD | SE | CP |
| $\beta_{10}$ | 0.400 | **−1.226** | 0.160 | 0.167 | **0.000** | **−1.229** | 0.163 | 0.170 | **0.000** | −0.009 | 0.183 | 0.179 | 0.921 | −0.000 | 0.190 | 0.182 | 0.924 |
| $\beta_{11}$ | −0.500 | **0.365** | 0.173 | 0.192 | **0.496** | **0.367** | 0.188 | 0.196 | **0.532** | −0.003 | 0.190 | 0.197 | 0.958 | −0.010 | 0.206 | 0.199 | 0.965 |
| $\gamma$ | −1.000 | **0.293** | 0.172 | 0.109 | **0.338** | **0.300** | 0.169 | 0.109 | **0.327** | −0.006 | 0.179 | 0.175 | 0.929 | −0.014 | 0.175 | 0.175 | 0.935 |
| $\eta_1$ | 0.400 | — | — | — | — | — | — | — | — | 0.009 | 0.104 | 0.106 | 0.955 | 0.021 | 0.105 | 0.107 | 0.953 |
| $\eta_2$ | 1.000 | — | — | — | — | — | — | — | — | 0.015 | 0.110 | 0.110 | 0.950 | 0.003 | 0.112 | 0.110 | 0.959 |

Note: Large bias and poor CP are highlighted in bold.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

**Table 3.** Model comparison statistics from various models for the DATATOP dataset. LPML: log pseudo-marginal likelihood; DIC: deviance information criterion; EAIC: expected Akaike information criterion; EBIC: expected Bayesian information criterion; BF: Bayes factor. The best fitting model is highlighted in bold.

| | LPML | DIC | EAIC | EBIC | WAIC | BF |
|---|---|---|---|---|---|---|
| $JM_{ST}$ | **−24832.28** | **51079.95** | **48319.02** | **48431.42** | **56359.83** | Ref |
| $RM_{ST}$ | −25103.57 | 51514.16 | 48897.54 | 49000.57 | 56796.36 | ≫100 |
| $JM_{SN}$ | −25416.64 | 52187.91 | 50163.75 | 50271.47 | 56982.33 | ≫100 |
| $JM_{T}$ | −26154.57 | 53863.16 | 51772.63 | 51880.35 | 58422.35 | ≫100 |
| $RM_{T}$ | −26404.45 | 54273.81 | 52360.01 | 52467.73 | 58832.19 | ≫100 |
| $JM_{N}$ | −26571.84 | 54689.14 | 53295.50 | 53398.54 | 58913.93 | ≫100 |
| $RM_{N}$ | −26820.56 | 55150.33 | 53919.96 | 54013.62 | 59367.91 | ≫100 |

favor of $JM_{ST}$ over all other models, suggesting that it was essential to account for all three data features (skewness, outliers, and dependent censoring) in the DATATOP study.

Table 4 displays the posterior means and standard deviation (SD) of all parameters from all six models. The skewness parameter $\delta$ was significantly positive in both models $JM_{SN}$ and $JM_{ST}$, suggesting the existence of right skewness in the outcome UPDRS. To visualize this, Fig. 3 displays the posterior density distribution of $\delta$ from models $JM_{SN}$ (left panel) and $JM_{ST}$ (right panel). It suggests that the UPDRS variable followed a right skewed distribution with a statistically significant skewness parameter ($\widehat{\delta} = 11.958$ with 95% CI [11.179, 12.805] in model $JM_{SN}$; $\widehat{\delta} = 12.159$ with 95% CI [11.424, 12.982] in model $JM_{ST}$). Figure 4 displays the posterior density distributions of the degree of freedom $\nu$ from models $JM_{T}$ (left panel) and $JM_{ST}$ (right panel). For both models, the densities of $\nu$ concentrated around small value (mean: 4.250, 95% CI [3.594, 5.010] in model $JM_{T}$; mean: 3.668, 95% CI [3.200, 4.232] in model $JM_{ST}$), providing some evidence for the existence of heavy tails (outliers) in the UPDRS variable.

**Table 4.** Results of fitting various models in the DATATOP dataset. Parameters $a_1$ and $b_1$ are for the outcome UPDRS. Parameters $a_{21}, \ldots, a_{24}$ and $b_2$ are for the outcome HY. Parameters $a_{31}, \ldots, a_{35}$ and $b_3$ are for the outcome SEADL. Parameters $\nu$ and $\delta$ are the tuning parameter and skewness parameter, respectively, for the outcome UPDRS.

| | $RM_N$ Mean(SD) | $RM_T$ Mean(SD) | $RM_{ST}$ Mean(SD) | $JM_N$ Mean(SD) | $JM_T$ Mean(SD) | $JM_{SN}$ Mean(SD) | $JM_{ST}$ Mean(SD) |
|---|---|---|---|---|---|---|---|
| $a_1$ | 23.998(0.448) | 23.894(0.410) | 14.123(0.431) | 23.935(0.421) | 23.777(0.347) | 14.368(0.475) | 14.134(0.427) |
| $b_1$ | 10.863(0.308) | 10.797(0.265) | 9.691(0.355) | 10.731(0.290) | 10.661(0.257) | 9.804(0.288) | 9.568(0.294) |
| $\sigma_1$ | 5.231(0.075) | 3.807(0.105) | 3.419(0.085) | 5.224(0.075) | 3.849(0.108) | 4.861(0.062) | 3.449(0.087) |
| $\nu$ | — | 4.096(0.348) | 3.552(0.240) | — | 4.250(0.368) | — | 3.668(0.262) |
| $\delta$ | — | — | 12.258(0.430) | — | — | 11.958(0.410) | 12.159(0.411) |
| $a_{21}$ | −0.871(0.068) | −0.884(0.063) | −0.804(0.057) | −0.863(0.065) | −0.869(0.056) | −0.788(0.059) | −0.798(0.054) |
| $a_{22}$ | 0.099(0.066) | 0.078(0.061) | 0.118(0.056) | 0.108(0.065) | 0.095(0.055) | 0.138(0.058) | 0.128(0.054) |
| $a_{23}$ | 3.218(0.086) | 3.173(0.081) | 3.047(0.077) | 3.230(0.084) | 3.197(0.078) | 3.082(0.079) | 3.071(0.078) |
| $a_{24}$ | 5.471(0.135) | 5.415(0.126) | 5.296(0.128) | 5.480(0.132) | 5.439(0.129) | 5.327(0.132) | 5.307(0.129) |
| $b_2$ | 1.400(0.053) | 1.378(0.048) | 1.323(0.059) | 1.385(0.050) | 1.368(0.046) | 1.317(0.051) | 1.324(0.052) |
| $a_{31}$ | −2.555(0.094) | −2.514(0.086) | −3.502(0.132) | −2.557(0.092) | −2.510(0.078) | −3.479(0.132) | −3.485(0.127) |
| $a_{32}$ | −0.506(0.083) | −0.520(0.076) | −0.563(0.112) | −0.501(0.080) | −0.503(0.067) | −0.531(0.115) | −0.548(0.106) |
| $a_{33}$ | 1.932(0.090) | 1.865(0.081) | 2.890(0.126) | 1.954(0.086) | 1.902(0.073) | 2.952(0.131) | 2.922(0.118) |
| $a_{34}$ | 2.774(0.096) | 2.688(0.087) | 4.065(0.138) | 2.800(0.092) | 2.733(0.080) | 4.145(0.143) | 4.104(0.131) |
| $a_{35}$ | 4.955(0.123) | 4.821(0.115) | 7.099(0.190) | 4.989(0.118) | 4.878(0.111) | 7.211(0.195) | 7.129(0.185) |
| $b_3$ | 1.827(0.067) | 1.762(0.058) | 3.019(0.123) | 1.814(0.062) | 1.759(0.058) | 2.999(0.099) | 2.993(0.104) |
| $\beta_{10}$ | 1.286(0.068) | 1.208(0.061) | 1.272(0.070) | 1.433(0.068) | 1.367(0.063) | 1.504(0.063) | 1.415(0.063) |
| $\beta_{11}$ | −0.609(0.071) | −0.540(0.067) | −0.579(0.072) | −0.681(0.073) | −0.630(0.069) | −0.730(0.080) | −0.655(0.064) |
| $\rho$ | 0.365(0.043) | 0.374(0.042) | 0.329(0.044) | 0.389(0.042) | 0.388(0.040) | 0.338(0.040) | 0.350(0.040) |
| $\sigma_u$ | 0.817(0.042) | 0.767(0.040) | 0.827(0.046) | 0.897(0.045) | 0.861(0.043) | 0.968(0.045) | 0.905(0.044) |
| $\gamma$ | −0.680(0.105) | −0.683(0.106) | −0.685(0.106) | −1.046(0.143) | −1.059(0.148) | −1.059(0.150) | −1.066(0.141) |
| $\eta_1$ | — | — | — | 0.333(0.068) | 0.324(0.070) | 0.428(0.070) | 0.448(0.072) |
| $\eta_2$ | — | — | — | 1.466(0.120) | 1.534(0.130) | 1.309(0.104) | 1.418(0.115) |

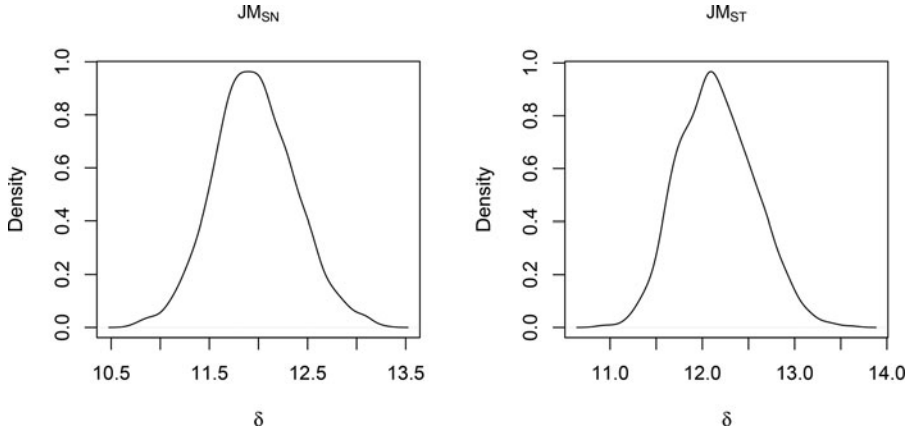


**Figure 3.** Posterior density functions of the skewness parameter $\delta$ from models $JM_{SN}$ (left panel) and $JM_{ST}$ (right panel).

Comparing to model $JM_{ST}$, the models that ignore skewness and/or outliers (models $RM_N$, $RM_T$, $JM_N$, $JM_T$, and $JM_{SN}$) provided larger estimates to outcome UPDRS' difficulty parameter $a_1$ and discriminating parameter $b_1$, a phenomenon also observed in simulation setting I. Moreover, comparing to their joint model counterparts (models $JM_N$, $JM_T$, and $JM_{ST}$), the reduced models (models $RM_N$, $RM_T$, and $RM_{ST}$) provided markedly different estimates to the parameters $\beta_{10}$, $\beta_{11}$, and $\gamma$, which was also reported in the simulation setting I.

Table 4 suggests that all six models demonstrated significant disease progression rate for placebo patients ($\beta_{10}$) and significant deprenyl effects in delaying the progression of PD symptoms ($\beta_{11}$). The results were consistent with the findings from the DATATOP study analysis (Parkinson Study Group, 1993). In model $JM_{ST}$, the disease progression rate for placebo patients ($\beta_{10}$) was 1.415 units per month (95% Cl: [1.302, 1.536]) and the changes in disease progression rate by deprenyl ($\beta_{11}$) was $-0.655$ units per month (95% Cl: $[-0.792, -0.534]$). Furthermore, deprenyl significantly decreased the risk of initiating levodopa therapy by 65.6% ($\gamma = -1.066$, $1 - \exp(-1.066) = 0.656$, 95% CI: [0.549, 0.739]). In addition, the significant positive estimates of parameters $\eta_0$ (mean: 0.448, 95% CI: [0.302, 0.586]) and $\eta_1$ (mean: 1.418,

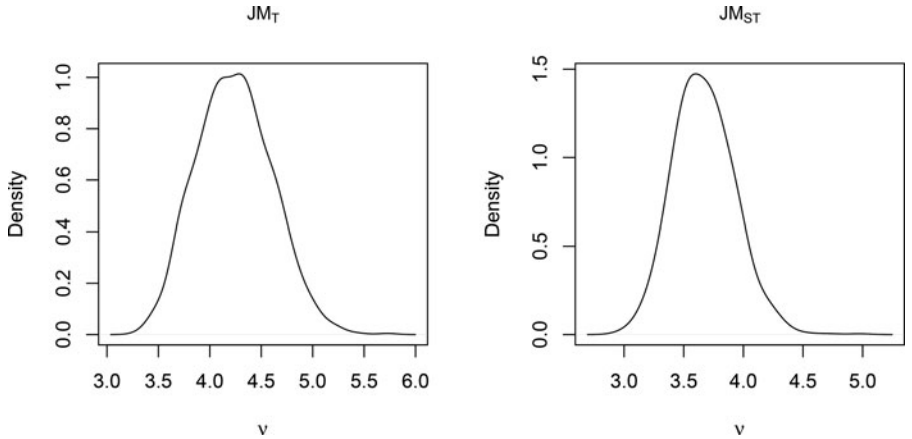


**Figure 4.** Posterior density functions of the degree of freedom $\nu$ from models $JM_T$ (left panel) and $JM_{ST}$ (right panel).

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

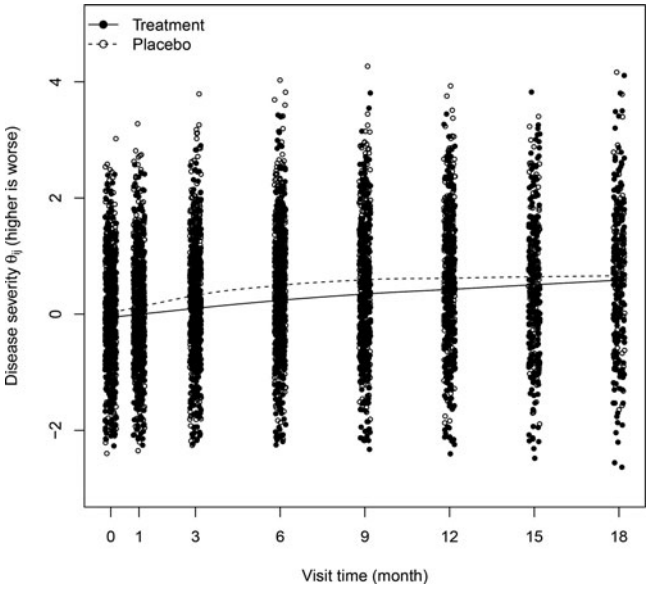


**Figure 5.** Bayesian posterior estimates of the subject-specific disease severity $\theta_{ij}$ at each visit and the lowess smooth curves for treatment and placebo groups from model $JM_{ST}$.

95% CI: [1.195, 1.646]) suggested that patients with worse disease severity (larger $u_{i0}$) and faster disease progression rate (larger $u_{i1}$) had higher risk of initiating levodopa therapy.

To provide insight into the latent disease progression at each visit, Fig. 5 displays the posterior estimates of the subject-specific latent disease severity $\theta_{ij}$ for all patients at all visits. The estimates of $\theta_{ij}$ of the patients in the treatment and placebo groups were denoted by black dots and circles, respectively. The solid and dashed lines were the lowess smooth curves for

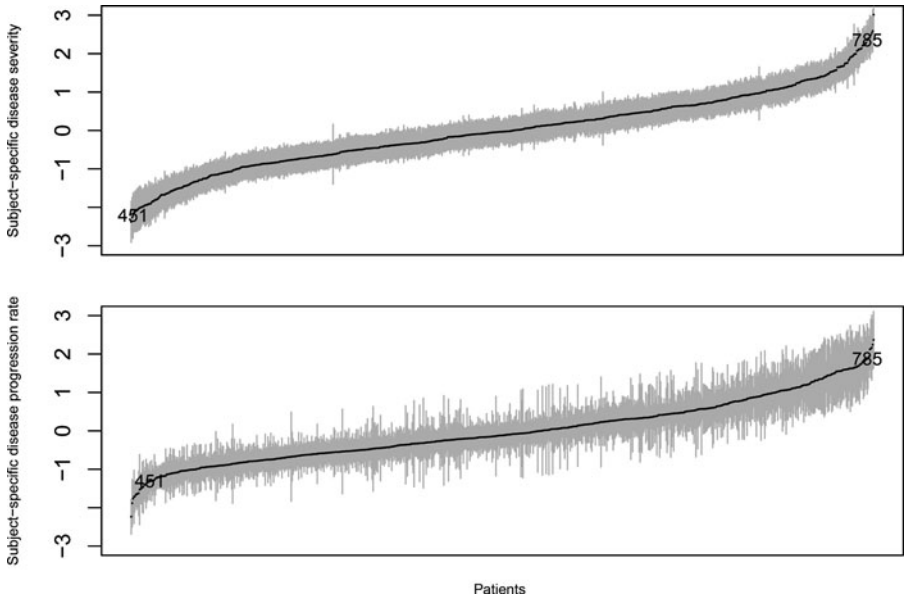


**Figure 6.** The ranking of the Bayesian posterior estimates of the subject-specific disease severity $u_{i0}$ (upper panel) and disease progression rate $u_{i1}$ (lower panel) with 95% CI, obtained from model $JM_{ST}$. The numbers in the figures are patient numbers.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

the treatment and placebo groups, respectively. Figure 5 suggests that the placebo patients had a faster disease progression rate before 9 month and slowed down thereafter, while the treatment patients maintained a relative stable slower progression rate overtime. Overall, the treatment patients had lower disease severity across all visits than the placebo patients. In addition, from Table 4, we also observed a significant positive correlation $\rho = 0.350$ (95% CI: [0.268, 0.424]) between the subject-specific disease severity $u_{i0}$ and disease progression rate $u_{i1}$. This suggests that patients with worse disease severity (higher $u_{i0}$) tend to have a faster disease progression rate (higher $u_{i1}$) and vice versa. To gain further insight into $u_{i0}$, $u_{i1}$, and $\rho$, we ranked the patients so that the patients with mild disease severity and slow disease progression rates had low ranks; while patients with more severe disease and faster disease progression rate had high ranks. Figure 6 displays the ranked $u_{i0}$ (upper panel) and ranked $u_{i1}$ (lower panel) for each patient. To visualize the effect of positive correlation coefficient $\rho$, we have selected two patients. Patient 785 who ranked No. 792 for disease severity (upper panel) had the same rank No. 792 in disease progression rate (lower panel) while patient 451 who ranked No. 3 in disease severity ranked No. 21 in disease progression rate.

## 6. Conclusions

In this article, we developed a joint modeling framework that consisted of an MLIRT sub-model for multivariate longitudinal outcomes and a Cox proportional hazard sub-model for the dependent terminal event while accounting for the skewness and outliers in the continuous outcomes by using the asymmetric and heavy-tailed skew-normal/independent (SNI) distributions. Simulation studies demonstrated that when both skewness and outliers existed in the continuous outcome, the proposed model $JM_{ST}$ produced accurate parameter estimates with small bias and CPs close to the nominal level. Additionally, in the absence of outliers and skewness in the continuous outcomes, our proposed model $JM_{ST}$ model still produced satisfactory results under model overparameterization. We applied our method to the motivating DATATOP study and discovered positive skewness and heavy tails (outliers) in the continuous outcome UPDRS. And we demonstrated that model $JM_{ST}$ was the best-fitting model to the DATATOP dataset based on six model selection criteria. Specifically, the Bayes factor results indicated decisive evidences in favor of model $JM_{ST}$ over all other models. 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 subject-specific disease progression rate provides visualization of their correlations. The Bayesian hierarchical implementation of SNI distributions to the MLIRT model is relatively straightforward. The easy access of publicly available software `JAGS` or `BUGS` 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. In addition to the DATATOP patients who were stopped due to the initiation of levodopa therapy, there were 24 additional patients (12 in each group) who dropped out due to adverse clinical effects of drugs, adverse lab tests due to drugs, and worsening of disease. In this article, these dropouts have been treated as independent censoring. However, there are two important reasons for carefully modeling these dropout events. First, they can be viewed as dependent censoring for the initiation of levodopa because both events are related to the overall patient disease condition. Second, disease-related dropout events generate non-ignorable missing values in the outcomes. To handle them, we may treat disease-related dropout events as a competing risk for the initiation of levodopa and generalize the proposed modeling framework

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

to account for multiple competing dependent censoring events. Moreover, we have modeled the covariate effects as linear. Although this assumption simplifies the models, it may not be supported by the data. In our future research, we would like investigate a class of varying-coefficient models (Sun and Wu, 2005) that incorporate the time-dependent covariate effects via penalized splines with a truncated polynomial basis and a fixed number of knots (Ruppert, 2002). Moreover, some researchers (Song et al., 2002; Zeng and Cai, 2005) have reported that the statistical inference of joint models is generally robust to the departure from the normality assumption of the random effects. It is of interest to investigate our joint model's performance when the underlying random effects distribution is symmetric non-normal or even asymmetric.

## Acknowledgments

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) for providing high-performing computing resources. The authors express appreciation to The University of Texas School of Public Health information technology staff for their technical support.

## Funding

Sheng Luo's research was supported by the National Institute of Neurological Disorders and Stroke under Award Numbers R01NS091307 and 5U01NS043127.

## References

Azzalini, A., Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B* 65(2):367–389.

Bandyopadhyay, D., Lachos, V. H., Abanto-Valle, C. A., Ghosh, P. (2010). Linear mixed models for skew-normal/independent bivariate responses with an application to periodontal disease. *Statistics in Medicine* 29(25):2643–2655.

Bandyopadhyay, D., Lachos, V. H., Castro, L. M., Dey, D. K. (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal* 54(3):405–425.

Bandyopadhyay, S., Ganguli, B., Chatterjee, A. (2011). A review of multivariate longitudinal data analysis. *Statistical Methods in Medical Research* 20(4):299–330.

Bushnell, D. M., 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(4):345–350.

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

Chen, G., Luo, S. (2016). Robust Bayesian hierarchical model using normal/independent distributions. *Biometrical Journal* 58(4):831–851.

Dey, D. K., Chen, M. H., Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* 53(4):1239–1252.

Douglas, J. A. (1999). Item response models for longitudinal quality of life data in clinical trials. *Statistics in Medicine* 18(21):2917–2931.

Elm, J. J., 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(12):1513–1521.

Fahn, S., Oakes, D., Shoulson, I., Kieburtz, K., Rudolph, A., Lang, A., Olanow, C. W., Tanner, C., Marek, K. (2004). Levodopa and the progression of Parkinson's disease. *The New England Journal of Medicine* 351(24):2498–2508.

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

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). *Bayesian Data Analysis*. Boca Raton, FL: CRC Press.
Gelman, A., Hwang, J., Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24(6):997–1016.
Glas, C. A. W., Geerlings, H., van de Laar, M. A. F. J., Taal, E. (2009). Analysis of longitudinal randomized clinical trials using item response models. *Contemporary Clinical Trials* 30(2):158–170.
He, B., Luo, S. (2016). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research* 25(4):1346–1358.
Henderson, R., Diggle, P., Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1(4):465–480.
Huang, P., Goetz, C. G., Woolson, R. F., Tilley, B., Kerr, D., Palesch, Y., Elm, J., Ravina, B., Bergmann, K. J., Kieburtz, K. (2009). Using global statistical tests in long-term Parkinson's disease clinical trials. *Movement Disorders* 24(12):1732–1739.
Huang, P., Tilley, B. C., Woolson, R. F., Lipsitz, S. (2005). Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens–Fisher problem. *Biometrics* 61(2):532–539.
Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement* 38(1):79–93.
Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* 90(430):773–795.
Lachos, V. H., Bandyopadhyay, D., Dey, D. K. (2011). Linear and nonlinear mixed-effects models for censored HIV viral loads using Normal/Independent distributions. *Biometrics* 67(4):1594–1604.
Lachos, V. H., Dey, D. K., Cancho, V. G. (2009). Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *Journal of Statistical Planning and Inference* 139(12):4098–4110.
Lachos, V. H., Ghosh, P., Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* 20(1):303–322.
Lawless, J. F., Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics* 26(4):549–565.
Lewis, S. M., Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace–Metropolis estimator. *Journal of the American Statistical Association* 92(438):648–655.
Little, R., Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics* 52(4):1324–1333.
Lord, F. M., Novick, M. R., Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Boston, MA: Addison-Wesley.
Luo, S., Lawson, A., He, B., Elm, J. J., Tilley, B. C. (2016). Bayesian multiple imputation for missing multivariate longitudinal data from a Parkinson's disease clinical trial. *Statistical Methods in Medical Research* 25(2):821–837.
McRae, C., Diem, G., Vo, A., O'Brien, C., Seeberger, L. (2000). Schwab & England: Standardization of administration. *Movement Disorders* 15(2):335–336.
Müller, J., Wenning, G. K., Jellinger, K., McKee, A., Poewe, W., Litvan, I. (2000). Progression of Hoehn and Yahr stages in Parkinsonian disorders: A clinicopathologic study. *Neurology* 55(6):888–891.
Mungas, D., 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(11–12):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(3):176–183.
Parkinson Study Group. (2002). A controlled trial of rasagiline in early Parkinson disease: The TEMPO study. *Archives of Neurology* 59(12):1937.
Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* 11(4):735–757.
Sahu, S. K., Dey, D. K., Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* 31(2):129–150.
Samejima, F. (1997). *Graded Response Model*. New York: Springer.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017

Song, X., Davidian, M., Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 58(4):742–753.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* 64(4):583–639.

Sun, Y., Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics* 32(1):21–47.

Zeng, D., Cai, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics* 33(5):2132–2163.

Downloaded by [Duke University Medical Center] at 15:05 07 September 2017