BAYESIAN INFERENCE FOR MULTIVARIATELONGITUDINAL DATA ANALYSIS USING ROBUST DISTRIBUTIONS

by

GENG CHEN, MS

APPROVED:

SHENG LUO, PhD

LUISA FRANZINI, PhD

XIAOMING LIU, PhD

ADRIANA PÉREZ, PhD

BARBARA TILLEY, PhD

DEAN, THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH

DEDICATION

This work is dedicated to my father Liwen Chen, mother Luning Zhao and my wife Min Tian, who have been giving me their selfless support for my study and research during all these years.

BAYESIAN INFERENCE FOR MULTIVARIATE LONGITUDINAL DATA ANALYSIS USING ROBUST DISTRIBUTIONS

by

GENG CHEN, BS, MS

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
December, 2014

PREFACE

The dissertation is submitted for the degree of Doctor of Philosophy at the University of Texas School of Public Health. The research was inspired by the early work in the Neuroprotection Exploratory Trials in Parkinson's Disease (NET-PD) study as a graduate research assistant. The purpose of the research was to try to find statistical methods that can better handle multivariate longitudinal data in clinical trial studies.

ACKNOWLEDGEMENTS

Working on the dissertation research has been challenging. It would not have been possible to write this dissertation without the support, patience and guidance from the professors who devote their precious time to serve on my committee. I am grateful for all the faculty members of my committee for their helpful discussion and comments.

I would like to express sincere appreciation to my committee chair, my major advisor, Dr. Sheng Luo, who has devoted himself on mentoring and nourishing me throughout my study and research. I cannot have gone far without his guidance and support. His profound positive impact will continue to benefit my future professional work.

I would like to thank Dr. Barbara Tilley, my committee member and the director of Biostatistics, who provided me the opportunity to work on the NET-PD Parkinson's disease study. Working with Dr. Tilley on the NET-PD study not only inspired me on my research topics, but also taught me how to handle various real-life problems, deal with people with different backgrounds and more importantly, how to perform like a real statistician.

I would like to thank Dr. Adriana Pérez, who guided me into this program and provided selfless support to my life and study. I am very grateful to have you in my committee, your hard-working and positive attitude will continue to guide my life and future professional work.

I would like to thank my minor advisor Dr. Luisa Franzini and breadth advisor Dr. Xiaoming Liu for their support, ideas and input to my dissertation.

At last, I would like to thank my family for their support for my study and research during all these years.

BAYESIAN INFERENCE FOR MULTIVARIATE LONGITUDINAL DATA ANALYSIS USING ROBUST DISTRIBUTIONS

GENG CHEN, MS, PhD
The University of Texas
School of Public Health, 2014

Dissertation Chair: Sheng Luo, PhD

Many clinical trials collect information on multiple longitudinal outcomes such as Parkinson's disease. As a method that account for all the information from multiple longitudinal outcomes, the multilevel item response theory (MLIRT) models have been increasingly used in clinical studies. 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 be violated due to skewness or outliers and thus may produce misleading results. In addition, patients' follow-up in the longitudinal studies may be stopped by terminal events such as death or dropout due to disease progression. The normal/independent (NI) distribution and the skew-normal/independent (SNI) distribution has been increasingly used to handle the outlier and skewness problem to produce robust inference. In the first paper, we developed approaches that implement the NI distributions on both continuous outcomes and random effects in the MLIRT models and discussed the model performance on different strategies of implementing the NI distribution. In the second paper, we developed approaches that implement the SNI distributions to the joint MLIRT models framework and evaluated the performance of the models on all the three data features

skewness, outliers and dependent censoring. Extensive simulation studies were conducted to evaluate the performance of various models. Specifically, we considered two continuous outcomes and two ordinal outcomes with the first outcome have outliers in the first paper; and we considered one continuous outcome and two ordinal outcomes with the continuous outcome has outliers and skewness in the second paper. Our proposed model Indep-CN-MLIRT in paper 1 and JM_{ST} model in paper 2 performed significantly better than their corresponding competing models. Our proposed methods were applied to a motivating Parkinson's disease study, the DATATOP study, to investigate the effect of deprenyl in slowing down the disease progression.

TABLE OF CONTENTS

List of Tables

List of Figures

1 BACKGROUND

1.1 Literature Review

1.1.1 Parkinson's disease as an example

Many clinical trials collect information on multiple longitudinal outcomes. Due to the characteristics of the disease and symptoms, the outcomes can be of mixed types, e.g., binary, ordinal, and continuous. Clinical studies on Parkinson's disease (PD), for example, are good representations of this case.

PD is a chronic progressive neurodegenerative disease. In the United States, the estimated prevalence is more than half a million people, and fifty thousand new cases are reported each year. The healthcare cost due to the morbidity and mortality of PD is around six billion dollars annually (Luo, 2014). The impairments caused by PD are multidimensional. Symptoms such as tremors, stiffness, slowness of movements, and loss of cognitive function can often be observed from PD patients (Cummings, 1992; Fahn et al., 2004). The multidimensional nature of the disease excludes the method that use only single outcome to summarize or represent the overall disease severity (Huang et al., 2005; Luo et al., 2012). Therefore, clinical trials that search for a neuroprotective treatment 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 (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).

To obtain more insight on the PD study, we introduce the motivating DATATOP trial as an example. The DATATOP study was a double-blind, placebo-controlled, multi-center clinical trial. A 2×2 factorial design 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 application of levodopa therapy. Eight hundred eligible patients were enrolled in the DATATOP study and randomized to one of the four treatment arms: active deprenyl alone, active tocopherol alone, both active deprenyl and tocopherol, and double placebo. Some of the longitudinal outcomes involved in the study were Unified Parkinson's Disease Rating Scale (UPDRS), Hoehn and Yahr scale (HY), and Schwab and England activities of daily living (SEADL), which were collected at baseline and months 1, 3, 6, 9, 12, 15, 18, 21, and 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). HY 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). SEADL is an ordinal variable that assesses patients' daily activities, the score range with integer values from 0 to 100 incrementing by 5 with larger values indicating better clinical conditions (McRae et al.,

2000). In the DATATOP study, only deprenyl was found to be effective in delaying the time until the need of levodopa therapy (Parkinson Study Group, 1989, 1993). There were 153 and 223 patients in treatment and placebo groups, respectively, reached the terminal events which were the application of levodopa therapy due to PD progression. The levodopa therapy provided temporary relief of PD symptoms and may significantly change the outcomes for a short period.

Thus, to investigate the effect of deprenyl, (1) we defined the treatment group as the patients who received deprenyl (active deprenyl alone and both active deprenyl and tocopherol), and defined the placebo group as the patients who did not receive deprenyl (active tocopherol alone and double placebo) and (2) since the introduction of levodopa therapy may significantly change the outcomes, only the outcomes results before the application of levodopa therapy were used to evaluate the efficacy of deprenyl.

1.1.2 Methods on analyzing multivariate outcomes

Among the multiple outcome measurements, there are three sources of correlation, i.e. inter-source (different measures at the same visit), intra-source (same measure at different visits), and cross correlation (different measures at different visits). Many approaches have been proposed 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) (Huang et al., 2005, 2009). Among those methods, the implementation of the linear combination of all outcomes is relatively easy, however, it reduces the information substantially and the interpretation may be difficult

(Huang et al., 2009; Bandyopadhyay et al., 2011). The method choosing one outcome as primary and other outcomes as secondary may encounter problems when the conclusion 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 adjusts for covariates of interest nor utilizes the full longitudinal data information.

An alternative approach is the latent variable approach that assumes all the outcomes are due to the unobservable latent variables (Mungas and Reed, 2000; Wang et al., 2002; Reise and Waller, 2009). The multilevel item response theory (MLIRT) models, which are based on the latent variable approach, have been increasingly used (Douglas, 1999; Glas et al., 2009; Luo et al., 2013; He and Luo, 2013). The MLIRT model assumes that the multivariate outcomes are clinical manifestations of a unobserved univariate latent variable that measures disease severity(Luo et al., 2013; He and Luo, 2013). The MLIRT model consists of two levels of models. The first level of the MLIRT models describes the relationship between outcome measurements and the subject-specific disease severity (the latent variable) and measurement-specific parameters. In the second level models, the latent variable is then 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 all three sources of correlations for each subject 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 (He and Luo, 2013; Maier, 2001; Kamata, 2001).

1.1.3 Multilevel item response theory (MLIRT) models

Let y_{ijk} be the observed outcome k ($k = 1, \ldots, K$) for subject 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 subject 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 underlining disease severity for subject i at visit j with higher value denoting more severe status. The multilevel item response theory (MLIRT) model consists of two levels. 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 measurement-specific parameters (e.g., the measurements ability to distinguish PD patients in disease severity). In the second level structural model, the latent disease severity is regressed on predictors (e.g., treatment, disease duration, and time) and subject-specific random effects (describing between-subject differences) to study the overall treatment effects via a single hypothesis test. Specifically, we model the cumulative probabilities of ordinal outcomes using graded response sub-model (Samejima, 1997), and the continues outcomes using common factor sub-model (Lord et al., 1968) as follows.

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

$$y_{ijk} = a_k + b_k\theta_{ij} + \epsilon_{ijk}, \tag{2}$$

where random error $\epsilon_{ijk} \sim \mathrm{N}(0, \sigma_k^2)$ with σ_k^2 being variance of continuous outcome k, a_k is the outcome-specific "difficulty" parameter and b_k is the "discriminating" parameter that is always positive and describes the intensity to which outcome k discriminates between patients with latent disease severity θ_{ij}. Suppose the ordinal outcome k in model (1) 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 that subject i being in category l on outcome k at visit j is defined as $\mathrm{P}(Y_{ijk} = l|\theta_{ij}) = \mathrm{P}(Y_{ijk} \leq l|\theta_{ij}) - \mathrm{P}(Y_{ijk} \leq l-1|\theta_{ij})$.

Model (3) is the level two model where the latent variables θ_{ij} is regressed on covariates of interest and the subject-specific random effects.

$$\theta_{ij} = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0 + u_{i0} + (\boldsymbol{X}_{i1}\boldsymbol{\beta}_1 + u_{i1})t_{ij}, \tag{3}$$

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 which determines the subject-specific disease severity and u_{i1} is the random slope which determines the subject-specific disease progression rate. We let $\boldsymbol{u}_i = (u_{i0}, u_{i1})'$ 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 over-parameterized (Lord et al., 1968; Samejima, 1997) and some constraints have to be imposed to make the models identifiable. To this end, we set $\mathrm{Var}(u_{i0}) = 1$ to make the model identifiable. 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 subject i across all visits is

$$\mathrm{L}(\boldsymbol{y}_i, \boldsymbol{u}_i) = \left[\prod_{j=1}^{J_i}\prod_{k=1}^{K} \mathrm{P}(y_{ijk}|\boldsymbol{u}_i)\right] \mathrm{P}(\boldsymbol{u}_i). \tag{4}$$

1.1.4 Joint modeling frame work of Multilevel item response theory (MLIRT) models

In addition to the multivariate longitudinal outcomes, the follow-up of patients in longitudinal clinical studies may be stopped by terminal events. Joint modeling of the terminal event time and the longitudinal outcomes have been used to provide consistent estimates (Henderson et al., 2000; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997).

Let t_i be the time to a terminal event ζ_i for subject i. Let $\mathbf{X}_i$ denote the vector of potential risk factors. Then the Cox proportional hazard model can be written as

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

where γ is the unknown parameter for the potential risk factors $\mathbf{X}_i$, η_0 and η_1 measure the association between the Cox proportional hazard model and the MLIRT model. We address the dependent censoring issue by jointly modeling the MLIRT model and the survival model. The shared random effects u_{i0} and u_{i1} account for the correlation between the survival time and longitudinal outcomes. To allow different baseline hazard rates at different time periods, we adopt the piecewise constant function to approximate the baseline hazard function $h_0(t)$ and assume that the hazard rate is constant within each given time period Lawless and Zhan (1998). Given a set of time points $0 = \tau_0 < \tau_1 < \ldots < \tau_m$, the baseline hazard $h_0(t)$ and the baseline hazard vector $\mathbf{g} = g_0, g_1, \ldots, g_{m-1}$, the piesewise constant hazard function can be defined as $h_0(t) = \sum_{l=0}^{m-1} g_l I_l(t)$, where $I_l(t)$ is a indicator function with $I_l(t) = 1$ if $\tau_l \leq t < \tau_{l+1}$ 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 subject i across all visits is

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

where the survival function $S(t_i) = \exp[\int_0^{t_i} h(s)ds]$ and $p(\boldsymbol{u}_i)$ is the density function for random effects $\boldsymbol{u}_i$.

1.1.5 The normal/independent (NI) distribution family

In the MLIRT models, normal distributions are usually assumed for the continuous outcomes and the random effects. However, the parameter estimation may be biased in the presence of heavy tails and outliers in the outcomes and/or random effects. To address the issue of non-normality due to heavy-tails and outliers, one solution is to replace the normality assumption by normal/independent (NI) distributions.

The normal/independent (NI) distribution is a family of symmetric distributions with heavier tails. Extensive discussion about the NI distributions can be found in (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, e is normally distributed with mean zero and variance σ^2, ω is a positive weight variable with density function P$(\omega|\nu)$, independent of e with tuning parameter ν. The NI distribution stochastically assigns different weights ω to each y, i.e. lower weights for potential outliers or influencing points, and thus controls the impact of the outliers on the overall inference

(Lachos et al., 2011). Given ω, y follows a normal distribution $\mathrm{N}(\mu,\ \omega^{-1}\sigma^2)$ with the marginal pdf of y given by $NI(y|\mu,\sigma^2,\nu) = \int \mathrm{P}(y|\mu,\Sigma^2,\omega)\mathrm{P}(\omega|\nu)d\omega$. When $\omega = 1$ (e.g., when $\nu \to \infty$), $NI(y|\mu,\sigma^2,\nu)$ becomes a normal distribution (Lange and Sinsheimer, 1993; Rosa et al., 2003; Lachos et al., 2011).

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 (2) as an example. When a univariate NI distribution is applied to model (2), 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 follows density function $\mathrm{P}(\omega_i|\nu)$ with positive tuning parameter ν. When $\omega_i \sim \mathrm{Gamma}(\nu/2,\ \nu/2)$, $\epsilon^{'}_{ijk}$ follows student's t distribution and when $\omega_i \sim \mathrm{Beta}(\nu,1)$, $\epsilon^{'}_{ijk}$ follows slash distribution. In addition, $\epsilon^{'}_{ijk}$ follows a contaminated normal distribution when ω_i takes one of the two discrete values with the pdf $\mathrm{P}(\omega_i|\nu) = \nu I_{(\omega_i=\gamma)} + (1-\nu)I_{(\omega_i=1)}$, where ν $(0 < \mu \leq 1)$ is the proportion of contamination (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). When $\mathrm{P}(\omega_i|\nu) = 1-\nu$, $\omega_i = 1$, indicating that $\epsilon^{'}_{ijk} \sim \mathrm{N}(0,\sigma_k^2)$ with the probability of $1-\nu$; when $\mathrm{P}(\omega_i|\nu) = \nu$, $\omega_i = \gamma$, indicating $\epsilon^{'}_{ijk}$ is contaminated with the probability of ν and $\mathrm{Var}(\epsilon^{'}_{ijk}) = \sigma_k^2/\omega_i = \sigma_k^2/\gamma$ (Lange and Sinsheimer, 1993; Rosa et al., 2003).

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).

1.1.6 The skew-normal/independent (SNI) distributions

The violation of normal distribution assumptions or the "departure from normality" may be due to skewness, heavy tails, or both (Lachos et al., 2011). Different methods of specifying the skew-normal (SN) and skew-t (ST) distribution have been discussed and used in literatures (Azzalini and Capitanio, 2003; Sahu et al., 2003; Arellano-Valle et al., 2007; Ghosh et al., 2007; Huang and Dagne, 2012). Here we consider the class of distribution introduced by Sahu which includes multivariate t, SN and ST distributions (Sahu et al., 2003).

Let D be a diagonal skewness matrix with elements $\delta_1 \dots \delta_m$. Let $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ν be the location vector, scale matrix, and degree of freedom, respectively. Then an m-dimensional skew-t distribution is given by: $f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, D, \nu) = 2^m t_{m,\nu}(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + D^2) P(\boldsymbol{V} > 0)$, where $\boldsymbol{V}$ follows a m-dimensional t distribution $t_{m,\nu+m}$ (Sahu et al., 2003). We denote the distribution as $ST_{m,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, D)$. Specifically, when $D = \delta I$ and $\boldsymbol{\Sigma} = \sigma^2 I$, where I is a $m \times m$ identity matrix, the density simplifies to

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\mu}, \sigma^2, \delta, \nu) \; &= 2^m (\sigma^2 + \delta^2)^{-m/2} \frac{\Gamma((\nu+m)/2)}{\Gamma(\nu/2)(\nu\pi)^{m/2}} \left\{ 1 + \frac{(\boldsymbol{y}-\boldsymbol{\mu})'(\boldsymbol{y}-\boldsymbol{\mu})}{\nu(\sigma^2+\delta^2)} \right\}^{-(\nu+m)/2} \\ &\times T_{m,\nu+m} \left[\left\{ \frac{\nu + (\sigma^2+\delta^2)^(-1)(\boldsymbol{y}-\boldsymbol{\mu})'(\boldsymbol{y}-\boldsymbol{\mu})}{\nu+m} \right\} \frac{\delta(\boldsymbol{y}-\boldsymbol{\mu})}{\sigma\sqrt{\sigma^2+\delta^2}} \right], \quad (7) \end{aligned}$$

where $T_{m,\nu+m}(\cdot)$ denotes the cumulative density function of $t_{m,\nu+m}(0, I)$.

The implementation of the simplified ST density (7) to the MLIRT model is not straightforward. But the ST distribution has two types of stochastic representations that are much easier to implement under Bayesian framework.

(i) As discussed by Sahu et al. (2003) and Huang and Dagne (2012), $\mathbf{Y} = \boldsymbol{\mu} + D|\mathbf{X}_0| +$

$\boldsymbol{\Sigma}^{1/2}\mathbf{X}_1$, where $\mathbf{X}_0$ and $\mathbf{X}_1$ are two independent random vectors with $|\mathbf{X}_0|$ following an m-dimensional t distribution $t_{m,v}(\mathbf{0}, \mathbf{I}_m)$ truncated in the space $|\mathbf{X}_0| > \mathbf{0}$ and $\mathbf{X}_1$ following $t_{m,v}(\mathbf{0}, \mathbf{I}_m)$. Let $\mathbf{z} = |\mathbf{X}_0|$, then a hierarchical representation is given by

$\mathbf{Y}|\mathbf{z} \sim t_{m,\nu+m}(\boldsymbol{\mu} + D\mathbf{z}, z\boldsymbol{\Sigma})$, $\mathbf{z} \sim t_{m,v}(\mathbf{0}, \mathbf{I}_m)\mathbf{I}(\mathbf{z} > \mathbf{0})$, where $z = (\nu + \mathbf{z}'\mathbf{z})/(\nu + m)$.

(ii) Sahu et al. (2003) also provided an alternative approach to specify the ST distribution using hierarchical specification method. Chen (2012) discussed the principle of the method based on Proposition 1 of Arellano-Valle et al. (2007)Arellano-Valle et al. (2007); Chen (2012). $\mathbf{Y} = \boldsymbol{\mu} + D|\mathbf{X}_0| + \boldsymbol{\omega}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{X}_1$, where $\mathbf{X}_0$ and $\mathbf{X}_1$ are two independent random vectors with $|\mathbf{X}_0|$ following $N_m(\mathbf{0}, \mathbf{I}_m)$ truncated in the space $|\mathbf{X}_0| > \mathbf{0}$ and $\mathbf{X}_1$ following $N_m(\mathbf{0}, \mathbf{I}_m)$; $\boldsymbol{\omega}$ is a weight variable following Gamma$(\nu/2, \nu/2)$ with ν be the degree of freedom for t distribution. Let $\mathbf{z} = |\mathbf{X}_0|$, then a hierarchical representation is given by

$\mathbf{Y}|\mathbf{z}, \boldsymbol{\omega} \sim N_m(\boldsymbol{\mu} + D\mathbf{z}, \boldsymbol{\omega}^{-1}\boldsymbol{\Sigma})$, $\mathbf{z} \sim N_m(\mathbf{0}, \mathbf{I}_m)\mathbf{I}(\mathbf{z} > \mathbf{0})$, $\boldsymbol{\omega} \sim$ Gamma$(\nu/2, \nu/2)$.

1.1.7 Bayesian inference and model selection criteria

There are various model selection methods available in Bayesian inference, for example, the log pseudo-marginal likelihood (LPML), the deviance information criterion (DIC), the expected Akaike information criterion (EAIC), the expected Bayesian information criterion (EBIC) and Bayes factor (BF) to assess model performance.

Conditional predictive ordinate (CPO) (Geisser, 1993; Carlin and Louis, 2011) is a cross-validation predictive method that evaluates the predictive distribution of the model conditioning on the data but with single data point deleted (Lachos et al., 2009; Chen

et al., 2000). 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 $CPO_i = p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}) = \int p(\boldsymbol{y_i}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{(i)})d\boldsymbol{\theta}$. Large CPO implies that the data for subject i can be well predicted by the model based on the data from all the other subjects. Thus lager CPO means a better fit for the model. Because there is no close form for CPO_i in our proposed model, a Monte Carlo estimation can be obtained from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. Since the function of CPO_i can be further derived as $CPO_i = p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}) = p(\boldsymbol{y})/p(\boldsymbol{y}_{(i)}) = 1/\int p(\boldsymbol{\theta}|\boldsymbol{y})/p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}, \boldsymbol{\theta})d\boldsymbol{\theta}$, let M be the total number of post burn-in samples, then a harmonic-mean approximation of CPO_i is $\widehat{CPO_i} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)},\boldsymbol{\theta}^{(t)})})^{-1} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(\boldsymbol{y_i}|\boldsymbol{\theta}^{(t)})})^{-1}$ (Luo et al., 2013; Chen et al., 2000). A summary statistics of CPO_i is log pseudo-marginal likelihood (LPML), defined as $LPML = \sum_{i=1}^{N}\log(\widehat{CPO_i})$. A larger value of LPML implies a better model fitting.

The deviance information criterion (DIC) assesses model fittings based on the posterior mean of the deviance and a penalty on the model complexity (Spiegelhalter et al., 2002). The deviance statistics is defined as $D(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y}|\boldsymbol{\theta}) + 2\log h(\boldsymbol{y})$, where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood of the observed data $\boldsymbol{y}$ given parameter vector $\boldsymbol{\theta}$, $h(\boldsymbol{y})$ is a standardized function of the data alone and have no impact on the assessment of the model fitting. Let $\bar{D}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}|\boldsymbol{y}}[D(\boldsymbol{\theta})]$ be the posterior mean of the deviance and let $D(\bar{\boldsymbol{\theta}}) = D(E_{\boldsymbol{\theta}|\boldsymbol{y}}[\boldsymbol{\theta}])$ be the deviance evaluated at the posterior mean of the parameter vector $\boldsymbol{\theta}$. The DIC is defined as $DIC = \bar{D}(\boldsymbol{\theta}) + p_D$, where $p_D = \bar{D}(\boldsymbol{\theta}) - D(\bar{\boldsymbol{\theta}})$ is the number of effective parameters and it captures the complexity of the model. A smaller value of DIC implies better fit of the model. Moreover, the expected Akaike information criterion (EAIC) and

the expected Bayesian information criterion (EBIC) (Carlin and Louis, 2011) are defined as $EAIC = \bar{D}(\boldsymbol{\theta}) + 2p$ and $EBIC = \bar{D}(\boldsymbol{\theta}) + p \log N$, respectively, where p is the total number of parameters used in the model and N is the sample size. Smaller values of EAIC and EBIC imply better fit of the model.

Furthermore, Bayes factor (BF) is a standard Bayesian solution (an alternative to p value) to the hypothesis testing for competing models. The BF quantifies the the degree to which whether the observed data support a hypothesis (Lavine and Schervish, 1999; Lewis and Raftery, 1997). Let two competing models be M_1 and M_2. Then for observed data $\boldsymbol{y}$, BF in favor of model M_1 over M_2 is defined as

$$\mathrm{BF}(M_1; M_2) = \frac{f(\boldsymbol{y}|M_1)}{f(\boldsymbol{y}|M_2)} = \frac{\int f(\boldsymbol{y}|\boldsymbol{\theta_1}, M_1) f(\boldsymbol{\theta_1}|M_1) d\boldsymbol{\theta_1}}{\int f(\boldsymbol{y}|\boldsymbol{\theta_2}, M_2) f(\boldsymbol{\theta_2}|M_2) d\boldsymbol{\theta_2}},$$

where $\boldsymbol{\theta_i}$ is the parameter vectors for model M_i for $i = 1, 2$; $f(\boldsymbol{y}|\boldsymbol{\theta_i}, M_i)$ is the likelihood of model M_i; and $f(\boldsymbol{\theta_i}|M_i)$ is the posterior density of $\boldsymbol{\theta_i}$ for model M_i (Lewis and Raftery, 1997; Gelman et al., 2013). 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 $f(\boldsymbol{y}|M_i)$ (Lewis and Raftery, 1997). Specifically, the $f(\boldsymbol{y}|M_i) \approx (2\pi)^{d_i/2} |\boldsymbol{\Sigma_i}|^{1/2} f(\boldsymbol{y}|\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}, M_i) f(\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}|M_i)$, where d_i is the number of parameters in $\boldsymbol{\theta_i}$, $\boldsymbol{\Sigma_i}$ is the posterior covariance matrix of $\boldsymbol{\theta_i}$, $\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}$ is the posterior mean of $\boldsymbol{\theta_i}$, $f(\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}|M_i)$ is the prior probability of parameters evaluated at $\bar{\theta}_i$, and $f(\boldsymbol{y}|\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}, M_i)$ is the likelihood evaluated at the posterior mean $\bar{\theta}_i$ (He and Luo, 2013; Lewis and Raftery, 1997). The interpretation of the BF is summarized by Kass and Raftery (1995)(Kass and Raftery, 1995). In paticular, when BF is greater than 100, decisive evidence is shown in

favor of Model M_1 over M_2.

1.2 Public Health Significance

The common goal of public health is to prevent diseases, promote health, and ensure patients who have disease receive the most effective intervention for such disease. Drug development and assessment play an important role in finding the best interventions for various diseases. In clinical trials, an accurate estimation of treatment effects may efficiently eliminate drugs that are not effective and maintain drugs that possess a positive treatment effect. Developing more accurate statistical tools has always been one of the most important goals of performing statistical research.

The MLIRT model uses the full longitudinal information and accounts for all three sources of correlations (inter-source, intra-source and cross correlation) for each subject via the subject-specific random effects and simultaneously estimates the measurement-specific parameters, the covariate effects, as well as the subject-specific disease progression characteristics (He and Luo, 2013; Maier, 2001; Kamata, 2001). The MLIRT models specifically reflect the multilevel data structure in PD clinical study data. However, the application of the MLIRT model is not only limited to PD, but also to any other diseases that have the multilevel data structure. In addition, in the field of bioinformatics, the MLIRT model could also be used to study the relationship between certain genetic marks and phenotypes. In this case, the phenotypes would be the outcome measures in the first level of the model and the the genetic marks are the covariates of interests in the second level of model, the latent variable would then be the overall effect of the genetic markers.

In real life data, it is not uncommon to have data that violate the usual normality assumption. The departure from normality may be due to outliers or skewness or both. Our goal is to provide a framework that takes into account the flexibility in distribution assumptions and produces robust statistical inference. The NI family of distributions downweights the potential outliers without having to delete them. Furthermore, the SNI distributions adjust the skewness in addition to the outliers. We construct a MLIRT model framework that relax the normal assumption with more flexible NI and SNI distributions for the multivariate longitudinal data and provide robust parameter estimations and more accurate statistical inference on the drug effects.

1.3 Hypothesis, Research Question, Specific aims

Common approaches handling non-normal data include elimination the outliers (or influential data points) and data transformation. Following the intent-to-treat (ITT) principle, elimination of the outliers may not be appropriate in many efficacy assessments in clinical trial studies (Little and Yau, 1996; Lachin, 2000). Data transformation methods (e.g., log, square-root, Box-Cox) might generate distributions close to normality, but the disadvantages from this approach 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 subject to "departure from normality" (Lachos et al., 2011). While the transformation for the outcomes might be feasible, the transformation for random effects may not be straightforward.

The patients' follow-up in longitudinal clinical studies may be stopped by terminal events. The terminal events could be noninformative (e.g. moving to other states or countries or traffic accidents) or informative (e.g. death or dropout due to disease progression). When the terminal events are related to patients' disease conditions, they are non-ignorable. The dependent terminal event time is commonly referred as informative censoring or dependent censoring. It has been shown that the ignorance of the dependent terminal time leads to biased estimates (Henderson et al., 2000). Joint modeling of the terminal event time and the longitudinal outcomes has been used to provide consistent estimates (Henderson et al., 2000; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997). Under the MLIRT framework, the Cox proportional hazard model or accelerated failure time (AFT) models has been used to jointly model the longitudinal outcomes and the dependent terminal events (He and Luo, 2013; Luo, 2014). He and Luo's paper discussed how the ignorance of the informative survival time may affect the model performance. Luo's paper further discussed the use of AFT model when the proportionality assumption is violated in Cox model.

To this end, we developed a framework for multivariate longitudinal data using a more generalized distribution family that (i) has high flexibility in shapes with a wide range of skewness; (ii) universally handles the skewness or outliers problems without the dataset-specific transformation to satisfy the normality assumption; (iii) accommodates the three data features (skewness, outliers and dependent censoring) simultaneously; and (iv) allows straightforward implementation using existing software.

1.3.1 Specific aims for paper 1

We explored the effects of using normal assumptions on random effects and outcomes under the MLIRT model framework when there are certain percentage of outliers on both random effects and outcomes through simulation. To solve the outlier problems on both random effects and outcomes, we imposed the NI distribution assumptions, and determined if the NI distribution assumption handles the outlier problems. Our hypothesis was that the replacement of the normal distribution with NI distribution may reduce the bias and improve model performance.

1.3.2 Specific aims for paper 2

The violation of the normal distribution assumption may be due to skewness or outliers or both. In addition, the follow-up in longitudinal clinical studies may be stopped by terminal events such as death or dropout. In paper 2, we explored the effect of the three data features, (i) skewness, (ii) outliers and (iii) dependent censoring, on our model performance. The effect of each of the data features were evaluated via simulation studies when the data are generated with skewness, outliers, and dependent censoring. We explored how the joint modeling framework of the MLIRT model with the SNI model assumptions and survival model handles the influence from the three data features. Our hypothesis was that the joint MLIRT model with SNI distribution assumptions provides accurate parameter estimates and robust statistical inference in terms of small bias and large coverage probabilities.

2 Article 1: Robust Bayesian hierarchical model using normal/independent distributions

(proposed for article submission to Biometrical Journal)

Robust Bayesian hierarchical model using normal/independent distributions

Abstract

The Multilevel item response theory (MLIRT) models have been increasingly used in 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) distribution has been increasingly used to handle the outlier and heavy tail problems to produce robust inference. In this article, we develop approaches that implement the NI distributions on both continuous outcomes and random effects in the MLIRT models and discuss the model performance

on different strategies of implementing the NI distribution. Extensive simulation studies are conducted to evaluate the performance of various models. Specifically, in the simulation study, we considered two continuous outcomes and two ordinal outcomes with the first outcome have outliers. Our proposed methods are 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; Normal/independent distribution; Random effects

2.1 Introduction

Many clinical trials collect information on multiple outcomes at different visits. For example Parkinson's disease (PD). PD is a chronic progressive disease with multidimensional impairments. Symptoms such as tremors, stiffness, slowness of movements, and loss of cognitive function can often be 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; Luo et al., 2012). 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). Due to the nature of the symptoms, the outcome measurements can be of mixed types, e.g., 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, i.e. inter-source (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.

To address the challenge, many approaches have been developed to analyze data with multiple outcomes, such as a linear combination of all outcomes, choosing one outcome as primary and other outcomes as secondary, multiple tests with adjustment to the overall significance level, and global statistical tests (GST) (Huang et al., 2005, 2009). Among those methods, 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 consist of two levels. The first level of the MLIRT models describes the relationship between outcome measurements and the 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 subject 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).

In the MLIRT models, normal distributions are usually assumed for the continuous outcomes and the random effects. However, the parameter estimation may be biased in the presence of heavy tails and outliers in the 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 of this approach 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 the outcomes might be feasible, the transformation for random effects may not be straightforward. To illustrate this, Figure 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 (18) (with normal assumptions) to the motivating DATATOP study. The plots suggests the presence of outliers in the continuous outcome UPDRS, the random intercept, and the random slope through the residual Q-Q plots. While McCulloch and Neuhaus (2011) concluded that the misspecification of the distribution of random effects does not severely affect the model estimation in 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.

To address the issue of non-normality due to heavy-tails and outliers, one solution is to replace the normality assumption by normal/independent (NI) distributions. The NI is an attractive class of symmetric heavy-tailed densities that includes the normal, Students-t, slash, and the contaminated normal distributions as special cases. The NI distributions have been applied to multivariate liner 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). Moreover, Lachos et al. (2011, 2013) used a NI distribution with a shared weight on both

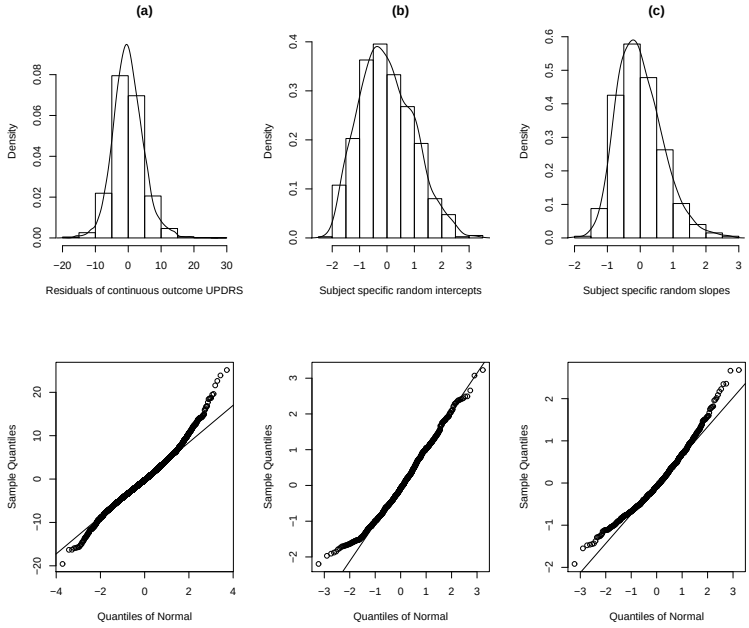

Figure 1: Histogram and normal Q-Q plot of residuals of UPDRS (column a), subject-specific random intercepts (column b) and subject-specific random slopes (column c) obtained by fitting a MLIRT model using normal assumptions.

continuous outcomes and random effects when both of them are shifted from normality due to heavy tails and outliers. In their methods, when conditional on the shared weight, the distribution of continuous outcomes and the random effects are independent and uncorrelated (Lachos et al., 2011). However, this method makes a relatively strong assumption that the continuous outcome and random effects share the same weight variable ω and NI distribution on the outliers or heavy tails. Alternatively, a more flexible approach is to assume different NI distributions for the continuous outcome and random effects. To the best of our knowledge, there are no studies on Bayesian MLIRT models using the NI distributions for both the continuous outcome and random effects. In this article, we propose a robust Bayesian parametric MLIRT model 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 3.3, we discuss the MLIRT models, the NI distributions, likelihood formulation, Bayesian inference, and model selection criterion. Section 3.7 presents an extensive simulation study comparing the performance of various models. In section 3.8, we apply the proposed models to the DATATOP study dataset. Section 3.9 summarizes the main findings and discusses the possible directions in our future research.

2.2 The model and estimation

2.2.1 The multilevel item response theory (MLIRT) model

Let y_{ijk} be the observed outcome k ($k = 1, \ldots, K$) for subject 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 subject 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 underlining disease severity for subject i at visit j with higher value denoting more severe status. The multilevel item response theory (MLIRT) model consists of two levels. 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 measurement-specific parameters (e.g., the measurements ability to distinguish PD patients in disease severity). In the second level structural model, the latent disease severity is regressed on predictors (e.g., treatment, disease duration,

and time) and subject-specific random effects (describing between-subject differences) to study the overall treatment effects via a single hypothesis test. Specifically, we model the cumulative probabilities of ordinal outcomes using graded response sub-model (Samejima, 1997), and the continues outcomes using common factor sub-model (Lord et al., 1968) as follows.

$$\text{logit}\big\{\text{P}(y_{ijk} \leq l|\theta_{ij})\big\} = a_{kl} - b_k\theta_{ij}, \text{ with } l = 1, 2, \ldots, n_k - 1, \tag{8}$$

$$y_{ijk} = a_k + b_k\theta_{ij} + \epsilon_{ijk}, \tag{9}$$

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 "discriminating" parameter that is always positive and describes the intensity to which outcome k discriminates between patients with latent disease severity θ_{ij}. Suppose the ordinal outcome k in model (14) 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 that subject i being in category l on outcome k at visit j is defined as $\text{P}(Y_{ijk} = l|\theta_{ij}) = \text{P}(Y_{ijk} \leq l|\theta_{ij}) - \text{P}(Y_{ijk} \leq l - 1|\theta_{ij})$.

Model (10) is the level two model where the latent variables θ_{ij} is regressed on covariates of interest and the subject-specific random effects.

$$\theta_{ij} = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0 + u_{i0} + (\boldsymbol{X}_{i1}\boldsymbol{\beta}_1 + u_{i1})t_{ij}, \tag{10}$$

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 which determines the subject-specific disease severity and u_{i1} is the random slope which determines the subject-specific disease progression rate. We let $\boldsymbol{u}_i = (u_{i0}, u_{i1})'$ and assume $u_{i0} \sim \text{N}(0, 1)$, $u_{i1} \sim \text{N}(0, \sigma_u^2)$, and $\text{corr}(u_{i0}, u_{i1}) =$

ρ. It is well-known that the item-response models are over-parameterized (Lord et al., 1968; Samejima, 1997) and some constraints have to be imposed to make the models identifiable. To this end, we set $\mathrm{Var}(u_{i0}) = 1$ to make the model identifiable. 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 subject i across all visits is

$$\mathrm{L}(\boldsymbol{y}_i, \boldsymbol{u}_i) = \left[\prod_{j=1}^{J_i}\prod_{k=1}^{K} \mathrm{P}(y_{ijk}|\boldsymbol{u}_i)\right] \mathrm{P}(\boldsymbol{u}_i). \tag{11}$$

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 $\boldsymbol{b} = (b_1, \ldots, b_K)^{'}$ for outcome k, and $\boldsymbol{\beta} = (\boldsymbol{\beta_0}^{'},\ \boldsymbol{\beta_1}^{'})^{'}$. Since we assume normal random error for the continuous outcome in model (9) and normal distribution for random effects, we refer this model as N-MLIRT model with the parameter vector $\boldsymbol{\Phi}_{\mathrm{N}} = (\boldsymbol{a}^{'}, \boldsymbol{b}^{'}, \boldsymbol{\beta}^{'},\ \rho,\ \sigma_u,\ \sigma_k)^{'}$.

2.2.2 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 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, e is normally distributed with mean zero and variance σ^2, ω is a positive weight variable with density function $\mathrm{P}(\omega|\nu)$, independent of e with tuning parameter ν. The NI

distribution stochastically assigns different weights ω to each outcome observation y, i.e. lower weights for potential outliers or influencing points, and thus control the impact of the outliers on the overall inference (Lachos et al., 2011). Given ω, y follows a normal distribution $\mathrm{N}(\mu,\ \omega^{-1}\sigma^2)$ with the marginal pdf of y given by $NI(y|\mu,\sigma^2,\nu) = \int \mathrm{P}(y|\mu,\sigma^2,\omega)\mathrm{P}(\omega|\nu)d\omega$. When $\omega = 1$ (or equivalently when $\nu \to \infty$), $NI(y|\mu,\sigma^2,\nu)$ becomes a normal distribution (Lange and Sinsheimer, 1993; Rosa et al., 2003; Lachos et al., 2011).

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 (9) as an example. When a univariate NI distribution is applied to model (9), 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 follows density function $\mathrm{P}(\omega_i|\nu)$ with positive tuning parameter ν. When $\omega_i \sim \mathrm{Gamma}(\nu/2,\ \nu/2)$, $\epsilon^{'}_{ijk}$ follows student's t distribution and when $\omega_i \sim \mathrm{Beta}(\nu,1)$, $\epsilon^{'}_{ijk}$ follows slash distribution. In addition, $\epsilon^{'}_{ijk}$ follows a contaminated normal (CN) distribution when ω_i takes one of the two discrete values with the pdf $\mathrm{P}(\omega_i|\nu) = \nu I_{(\omega_i=\gamma)} + (1-\nu)I_{(\omega_i=1)}$, where ν $(0 < \mu \leq 1)$ 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). When $\mathrm{P}(\omega_i|\nu) = 1-\nu$, $\omega_i = 1$, $\epsilon^{'}_{ijk} \sim \mathrm{N}(0,\sigma_k^2)$ with the probability of $1-\nu$; when $\mathrm{P}(\omega_i|\nu) = \nu$, $\omega_i = \gamma$, $\epsilon^{'}_{ijk}$ is contaminated with the probability of ν and $\mathrm{Var}(\epsilon^{'}_{ijk}) = \sigma_k^2/\omega_i = \sigma_k^2/\gamma$ (Lange and Sinsheimer, 1993; Rosa et al., 2003).

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 (9) and the random effect vector $\boldsymbol{u}_i = (u_{i0}, u_{i1})'$ in model (10). We first discuss the method in Lachos et al. (2011, 2013) in linear and nonlinear mixed models. Lachos et al. (2011, 2013) assume that $(\boldsymbol{u}_i, \epsilon_i) \sim NI(\mathbf{0}, \text{Diag}(\boldsymbol{\Sigma}_u, \sigma^2), \omega_i)$, where ϵ_i is the random error of a continuous outcome for subject 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 (Lachos et al., 2011). To use their methods in the MLIRT model, we specify $\boldsymbol{u}_i$, ϵ_i and ω_i hierarchically as

$$\begin{aligned} \boldsymbol{u}_i|\omega_i &\sim \text{N}(\mathbf{0},\ \omega_i^{-1}\Sigma_u), \\ \epsilon_i|\omega_i &\sim \text{N}(\mathbf{0},\ \omega_i^{-1}\sigma_k^2), \\ \omega_i &\sim \text{P}(\omega_i|\nu). \end{aligned}$$

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

$$L(\boldsymbol{y}_i, \omega_i, \boldsymbol{u}_i) = \left[\prod_{j=1}^{J}\prod_{k=1}^{K} \text{P}(y_{ijk}|\boldsymbol{u}_i, \omega_i)\right] \text{P}(\omega_i)\text{P}(\boldsymbol{u}_i). \tag{12}$$

We denote the model as Dep-NI-MLIRT because $\boldsymbol{u}_i$ and ϵ_i are marginally dependent. The corresponding parameter vector is $\boldsymbol{\Phi}_{\text{D}} = (\boldsymbol{a}',\ \boldsymbol{b}',\ \boldsymbol{\beta}',\ \rho,\ \sigma_u,\ \sigma_k,\ \nu)'$. We also refer to the Dep-NI-MLIRT model using the student's t, slash, and contaminated normal distributions as Dep-T-MLIRT, Dep-SL-MLIRT, and Dep-CN-MLIRT models, respectively.

The Dep-NI-MLIRT model assumes that the continuous variable and the random effects vector maintain the same scale of heavy tails or outliers, which may not necessarily be true for all data. This assumption may affect the accuracy of model estimation. Alternatively, we assume that $\boldsymbol{u}_i$ and ϵ_i are scaled by different weight variables, as $\boldsymbol{u}_i \sim NI(\mathbf{0},\ \Sigma_u,\ \omega_{1i})$, where ω_{1i} is a subject level weight variable for $\boldsymbol{u}_i$, and $\epsilon_{ijk} \sim NI(0,\ \sigma_k^2,\ \omega_{2ijk})$, where ω_{2ijk} is a weight variable (specific to outcome k from subject 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

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

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 subject i is

$$L(\boldsymbol{y}_i, \boldsymbol{\omega}_i, \boldsymbol{u}_i) = \prod_{j=1}^{J_i}\left[\prod_{k=1}^{K}\mathrm{P}(y_{ijk}|\boldsymbol{u}_i, \omega_{2ijk})\mathrm{P}(\omega_{2ijk})\right]\mathrm{P}(\omega_{1i})\mathrm{P}(\boldsymbol{u}_i). \tag{13}$$

We denote this model as Indep-NI-MLIRT model because $\boldsymbol{u}_i$ and ϵ_i are marginally independent. The corresponding parameter vector is $\boldsymbol{\Phi}_{\mathrm{I}} = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_k, \nu_1, \nu_2)'$. We also refer to the Indep-NI-MLIRT model using the student's t, slash, and contaminated normal distributions as Indep-T-MLIRT, Indep-SL-MLIRT, and Indep-CN-MLIRT models, respectively.

2.2.4 Bayesian inference and model selection criteria

We develop a Bayesian approach using MCMC techniques to analyze the multivariate longitudinal data by applying the NI distribution to the MLIRT models. The model fitting is conducted using the `BUGS` language. We assume vague priors on all elements in the parameter vectors $\boldsymbol{\Phi}_{\mathrm{N}}$, $\boldsymbol{\Phi}_{\mathrm{D}}$, and $\boldsymbol{\Phi}_{\mathrm{I}}$. Specifically, the prior distribution of all parameters in $\boldsymbol{\beta}$ is $\mathrm{N}(0, 100)$. We use the prior distribution Gamma(0.001, 0.001) for σ_u and all components in $\boldsymbol{b}$ to ensure positivity, and use 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 Gamma$(0.001, 0.001)$ for parameter ν in the student's t and slash distributions and use Beta$(1, 1)$ for γ and ν in the contaminated normal distribution. All model fitting is performed in `OpenBUGS` (`OpenBUGS` version 3.2.2). Multiple chains with dispersed initial values are run. To assess convergence, we use the history plots embedded in `OpenBUGS` to make sure there are no appearance of trend for all parameters. In addition, we used Gelman-Rubin diagnostic statistics to ensure the scale reduction $\hat{R}$ of all parameters are smaller than 1.1 (Gelman et al., 2004).

Many model selection criteria have been proposed for Bayesian inference. We select five of the criteria, the log pseudo-marginal likelihood (LPML), the deviance information criterion (DIC), the expected Akaike information criterion (EAIC), the expected Bayesian information criterion (EBIC) 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 condition on the data but with single data point deleted (Chen et al., 2000; Lachos et al., 2009). 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 $\text{CPO}_i = \text{P}(y_i|\boldsymbol{y}_{(i)}) = \int \text{P}(y_i|\boldsymbol{\theta})\text{P}(\boldsymbol{\theta}|\boldsymbol{y}_{(i)})d\boldsymbol{\theta}$. Large CPO implies that the data for subject i can be well predicted by the model using posterior density of $\boldsymbol{\theta}$ based on $\boldsymbol{y}_{(i)}$. For our proposed model, there is no close form for CPO_i, thus a Monte Carlo estimation method is used to obtain $\widehat{\text{CPO}}_i$. The methods of computing $\widehat{\text{CPO}}_i$ using Monte Carlo estimation are detailed by Chen et al. (2000) and Luo et al. (2013). A summary statistics of CPO_i is log pseudo-marginal likelihood (LPML), defined as $LPML = \sum_{i=1}^{N} \log(\widehat{\text{CPO}}_i)$. A larger value of LPML implies a better model fitting.

The deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) assesses model fittings based on the posterior mean of the deviance and a penalty on the model complexity. The deviance statistics is defined as $\text{D}(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y}|\boldsymbol{\theta}) + 2\log h(\boldsymbol{y})$, where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood of the observed data $\boldsymbol{y}$ given parameter vector $\boldsymbol{\theta}$, $h(\boldsymbol{y})$ is a standardized function of the data alone and have no impact on the assessment of the model fitting. Let $\bar{\text{D}}(\boldsymbol{\theta}) = \text{E}_{\boldsymbol{\theta}|\boldsymbol{y}}[\text{D}(\boldsymbol{\theta})]$ be the posterior mean of the deviance and let $\text{D}(\bar{\boldsymbol{\theta}}) = \text{D}(\text{E}_{\boldsymbol{\theta}|\boldsymbol{y}}[\boldsymbol{\theta}])$ be the deviance evaluated at the posterior mean of the parameter vector $\boldsymbol{\theta}$. The DIC is defined as $\text{DIC} = \bar{\text{D}}(\boldsymbol{\theta}) + p_D$, where $p_D = \bar{\text{D}}(\boldsymbol{\theta}) - \text{D}(\bar{\boldsymbol{\theta}})$ is the number of effective parameters and it captures the complexity of the model. A smaller value of DIC implies better fit of the model. Moreover, the expected Akaike information criterion (EAIC) and the expected Bayesian information criterion (EBIC) (Carlin and

Louis, 2011) are defined as EAIC $= \bar{\mathrm{D}}(\boldsymbol{\theta}) + 2p$ and $EBIC = \bar{\mathrm{D}}(\boldsymbol{\theta}) + p \log \mathrm{N}$, respectively, where p is the total number of parameters used in the model and N is the sample size. Smaller values of EAIC and EBIC imply better fit of the model.

Furthermore, Bayes factor (BF) is an 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 (Lavine and Schervish, 1999; 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 $\mathrm{BF}(M_1; M_2) = \frac{f(\boldsymbol{y}|M_1)}{f(\boldsymbol{y}|M_2)} = \frac{\int f(\boldsymbol{y}|\boldsymbol{\theta_1}, M_1) f(\boldsymbol{\theta_1}|M_1) d\boldsymbol{\theta_1}}{\int f(\boldsymbol{y}|\boldsymbol{\theta_2}, M_2) f(\boldsymbol{\theta_2}|M_2) d\boldsymbol{\theta_2}}$, where $\boldsymbol{\theta_i}$ is the parameter vectors for model M_i for $i = 1, 2$; $f(\boldsymbol{y}|\boldsymbol{\theta_i}, M_i)$ is the likelihood of model M_i; and $f(\boldsymbol{\theta_i}|M_i)$ is the posterior density of $\boldsymbol{\theta_i}$ for model M_i (Lewis and Raftery, 1997; Gelman et al., 2013). The Laplace-Metropolis estimator based on normal distribution is used to approximate the marginal likelihood $f(\boldsymbol{y}|M_i)$ (Lewis and Raftery, 1997). In particular, the $f(\boldsymbol{y}|M_i) \approx (2\pi)^{d_i/2} |\boldsymbol{\Sigma_i}|^{1/2} f(\boldsymbol{y}|\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}, M_i) f(\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}|M_i)$, where d_i is the number of parameters in $\boldsymbol{\theta_i}$, $\boldsymbol{\Sigma_i}$ is the posterior covariance matrix of $\boldsymbol{\theta_i}$, $\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}$ is the posterior mean of $\boldsymbol{\theta_i}$, $f(\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}|M_i)$ is the prior probability of parameters evaluated at $\bar{\theta}_i$, and $f(\boldsymbol{y}|\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}, M_i)$ is the likelihood evaluated at the posterior mean $\bar{\theta}_i$ (Lewis and Raftery, 1997). The interpretation of the BF is summarized by Kass and Raftery (1995) (Kass and Raftery, 1995). When BF is greater than 100, decisive evidence is shown in favor of Model M_1 over M_2.

2.3 Simulation studies

In this section, we conducted an extensive simulation study under three settings to compare the performance of the models N-MLIRT, Dep-NI-MLIRT, and Indep-NI-MLIRT in different scenarios. For all settings, we generated 100 datasets with a sample size of 400 patients (200 in both treatment and placebo groups). The data structure is similar to the motivating DATATOP study, and it has 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 (10) 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 $\boldsymbol{u}_i = (u_{i0}, u_{i1})' \sim N_2(0, \boldsymbol{\Sigma}_u)$, where $\boldsymbol{\Sigma}_u = \{(1, \rho\sigma_u), (\rho\sigma_u, \sigma_u^2)\}$ with $\rho = 0.4$ and $\sigma_u = 1.3$. 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 square root of the average of the posterior variance (SE), the standard deviation of the posterior means (SD), and the coverage probabilities (CP) of the 95% equal-tail credible intervals.

In setting I, model N-MLIRT was the true model, and both the continuous outcomes

and the random effects followed normal distributions without outliers. The results were summarized in Table 1. Due to the space constraint, we only presented the MLIRT models which use CN to specify the NI distribution. The results suggested that all three models (N-MLIRT, Dep-CN-MLIRT, and Indep-CN-MLIRT) generate comparable results, i.e. the bias was negligible, SE was close to SD, and the CP's were reasonably close to nominal level of 95%.

In setting II, we evaluated the model performance in the presence of heavy tails and outliers. The model settings were similar to setting I, but with the first continuous outcome and the random effects 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 from the first continuous outcome and add the noise generated from either Uniform$[3\sigma_1, 6\sigma_1]$ or Uniform$[-6\sigma_1, -3\sigma_1]$ with equal probabilities. Similarly, we randomly selected 5% of the generated random effects $\boldsymbol{u}_i$ and replace by data generated from either Uniform$[5, 15]$ or Uniform$[-15, -5]$ with equal probabilities.

The results of setting II were summarized in Table 2. The results from models Dep-CN-MLIRT and Indep-CN-MLIRT indicated that the estimates of all parameters had negligible bias and SE being close to SD. The CP's were all reasonably around the nominal value. In contrast, model N-MLIRT provided severely biased estimates and low coverage probabilities for all parameters. The misspecification of the random effects distribution affected the estimation of $\boldsymbol{u}_i$ and thus of θ_{ij}. As shown in models (14) and (9), the imprecise estimation on θ_{ij} affected the estimation of the outcome-specific

parameter vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ for all outcomes. In sum, simulation setting II suggested that the misspecification of the distribution assumptions on the continuous outcome and random effects severely impacted the inference in the MLIRT modeling framework.

In simulation setting III, we compared the performance of Dep-NI-MLIRT and Indep-NI-MLIRT models with all the three NI distributions specificatoins (slash, t and CN) when the datasets have 5% outliers as in setting II. In our simulation model, β_{10} and $\beta_{10} + \beta_{11}$ indicate the disease progression rate for placebo and treatment, respectively, with β_{11} indicating the change in disease progression rate introduced by treatment. Thus, due to space constraint, only the estimations on the treatment effect parameter vector $\boldsymbol{\beta}$ were summarized in Table 3. The results suggested that both models Dep-CN-MLIRT and Indep-CN-MLIRT perform better than their counterparts with other NI distributions, as indicated by smaller bias, smaller SD and SE, and the CP's closer to the nominal levels. In addition, the contaminated normal (CN) distribution performed better than slash and t distribution in both Dep-CN-MLIRT and Indep-CN-MLIRT models.

From the simulation study, we concluded that when there are not outliers, the N-MLIRT, Dep-NI-MLIRT and Indep-NI-MLIRT models all provided comparable and satisfactory results. However, when outliers are presented in both the continuous outcome and the random effects, both Dep-NI-MLIRT and Indep-NI-MLIRT models provided less bias results than the N-MLIRT model, which provides severely biased estimates to all parameters.

2.4 Application to the DATATOP study

In this section, we fit the proposed Dep-CN-MLIRT and Indep-CN-MLIRT models to the motivating Deprenyl And Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) study, which was a double-blind, placebo-conrolled, multi-center clinical trial. A factorial design 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 application 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 need of levodopa therapy (Parkinson Study Group, 1989, 1993). In our analysis, we defined the treatment group as the patients who received deprenyl (active deprenyl alone and both active deprenyl and tocopherol), and defined the placebo group as the patients who did not receive deprenyl (active tocopherol alone and double placebo). Three longitudinal outcomes were used. The Unified Parkinson's Disease Rating Scale (UPDRS) total score, the Hoehn and Yahr scale (HY), and the Schwab and England activities of daily living (SEADL), which were collected at baseline and months 1, 3, 6, 9, 12, 15, 18, 21, and 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). HY measures the disability level in daily activities and it is an ordinal outcome ranging from 1 to 5 with higher values indicating worse

conditions (Müller et al., 2000). 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. In addition, the validity of the asymptotic distribution of the GST test is based on the equal variance and covariance assumption on the longitudinal outcomes, however, this assumption remain to be checked in the DATATOP study.

Figure 2 displayed 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 change in UPDRS measurement. However, patients 105, 108 and 749 (dashed lines) have some potential outlying measurements indicated by their dramatic value change in their UPDRS profiles. These four patients would be used for further discussion.

To fit our proposed model 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 (10) became $\theta_{ij} = u_{i0} + (\beta_{10} + \beta_{11} x_i + u_{i1}) t_{ij}$. We used two parallel MCMC chains with over-dispersed initial values, and ran each chain for 30,000 iterations. The first 15,000 iterations are discarded as burn-in and the inference is based on the remaining 15,000 iterations.

Table 4 compared the N-MLIRT, Dep-CN-MLIRT and Indep-CN-MLIRT models

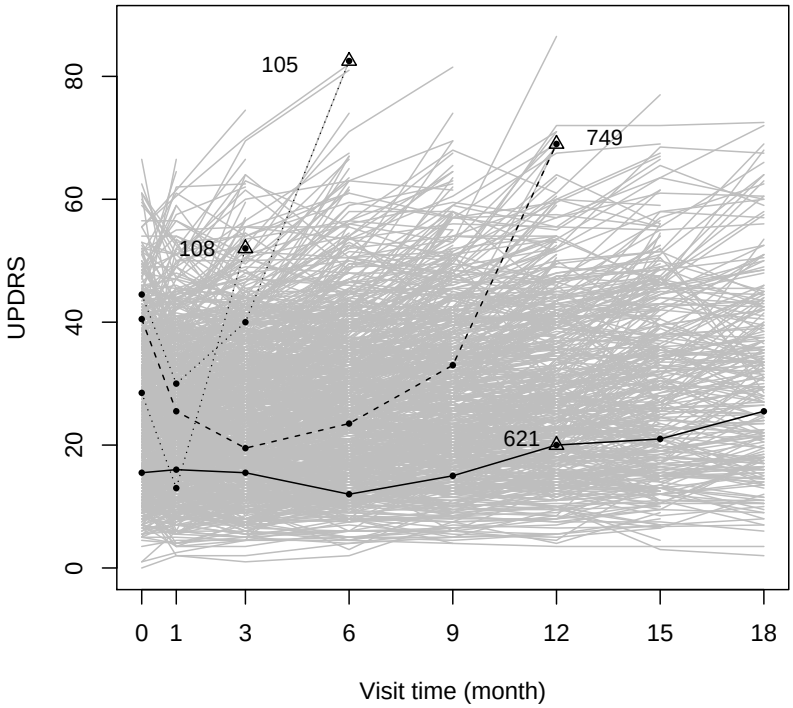

Figure 2: Longitudinal profile plots of observed outcome UPDRS. Numbers 105, 108, 621, 749 denotes four patients to be used for further discussion.

using the four model comparison criteria discussed in Section 3.6 on the overall model performance. The proposed models Dep-CN-MLIRT and Indep-CN-MLIRT performed significantly better than model N-MLIRT with larger LPML value and smaller DIC, EAIC, EBIC values, suggesting the advantage of accounting for outliers in the outcome UPDRS and random effects. The BFs in favor of the Indep-CN-MLIRT model over the Dep-CN-MLIRT model and N-MLIRT model are both much larger than 100, indicating decisive evidence in favor of the Indep-CN-MLIRT model. The Indep-CN-MLIRT model has the best fit with the highest LPML value and the lowest DIC, EAIC, EBIC and BF

values and it was selected as the final model.

In Table 5, the estimates of the standard deviations for UPDRS (σ_1) and random effects (σ_u) were smaller in Dep-CN-MLIRT and Indep-CN-MLIRT models than those in model N-MLIRT. It was interesting to note that Indep-CN-MLIRT model provided smaller estimates on both σ_1 and σ_u than model Dep-CN-MLIRT.

Parameters β_{10} and $\beta_{10} + \beta_{11}$ represent the disease progression rates for the placebo and treatment patients, respectively, with parameter β_{11} indicating the change in disease progression rate introduced by target treatment deprenyl. As shown in Table 5, 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% Cl: [1.163, 1.414]), 1.334 (95% Cl: [1.200, 1.471]), and 1.226 (95% Cl: [1.107, 1.351]) from models N-MLIRT, Dep-CN-MLIRT and Indep-CN-MLIRT, respectively. The changes in disease progression rate introduced by deprenyl (β_{11}) were -0.606 (95% Cl: [-0.751, -0.457]), -0.609 (95% Cl: [-0.762, -0.461]), and -0.572 (95% Cl: [-0.707, -0.443]) from models N-MLIRT, Dep-CN-MLIRT and Indep-CN-MLIRT, respectively.

To obtain further insight into how the NI distributions control the influence of the outliers, in Figure 3 we plotted the standardized residuals (SRs) of UPDRS for all patients at each visit after fitting model N-MLIRT. A few data points had SRs with absolute value larger than 3 (for example, 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 was -0.40 at 12-month visit, indicating a non-outlier. Without proper adjustment for the outliers, they may effect the accuracy of the model estimation due to the violation of the model assumptions and therefore a heavy-tail distribution assumption for the UPDRS is essential. The weight variable ω_{ijk} in the NI distributions can be used for outlier detections (Rosa et al., 2003). Figure 4 demonstrated the posterior distribution of the weight variable ω_{ijk} for patient 105, 108, 749 and 621 after fitting the Indep-CN-MLIRT model. The NI distribution stochastically attributes lower weights for potential outliers and thus control the impact of the outliers on the overall inference (Lachos et al., 2011). As indicated in Figure 2 and Figure 3, patient 105 at 6-month visit, patient 108 at 3-month visit and patient 749 at 12-month visit were potential outliers, and their posterior mean of the weight were 0.18, 0.17 and 0.18, respectively. On the contrary, the posterior mean of the weight for patient 621 was 0.91, which indicated that this observation may not be an outlier.

The visualization of the disease progression for all patients was displayed in Figure 5. 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 subject i at visit j. The solid line and dashed line were the the lowess smooth curves for treatment and placebo groups, respectively. Figure 5 suggested 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. Moreover, we observed in Table 5 a positive correlation, $\widehat{\rho} = 0.354$ (95% CI: [0.263, 0.447]), between the subject-specific disease severity u_{i0} and disease progression rate u_{i1}. To visualize the correlation between u_{i0}

and u_{i1}, we ranked the patients so that the patients who have mild disease severity and slow disease progression rate have low ranks; while patients who have severe disease and disease progression rate have high ranks. Figure 6 showed the ranked u_{i0} (upper panel) and ranked u_{i1} (lower panel) for each patient. To assist with the interpretation of the positive correlation ρ, we selected two patients in Figure 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 rank 16 in the disease severity (upper panel) had the 10th disease progression rate (lower panel).

2.5 Conclusions

In this article we provided a new statistical analysis framework for the multivariate longitudinal data when two continuous outcomes and two ordinal outcomes were used and the first continuous outcome and random effects have heavy tails and outliers. We performed an extensive simulation study to illustrate how our proposed methods handle the heavy tails and outlier problems and provide robust parameter estimations. We compared the simulation results and selected the Indep-CN-MLIRT model, which maintains the flexibility of specifying different tuning parameters for continuous outcome and random effects, as our best model in handling outliers and providing accurate parameter estimates. We also applied our method to the DATATOP study on Parkinson's disease and demonstrated how the implementation of the NI distribution to the MLIRT model obtains robust inference. We plotted the observed longitudinal profile of 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 of 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 offered the visualization to their correlations.

The hierarchical implementation of 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.

In our simulation study, 5% outliers on the first continuous outcome and random effects were considered and discussed. The percent of outliers could be increased to 10% or 20%, however, the computational time will need to be doubled or tripled to accommodate the additional outlier scales. The current computing time for one dataset and one chain took approximately eight hours to finish the job on SPH cluster. In our simulation study, we had 100 datasets and two chains for each dataset. In addition, we compared the slash, t, and contaminated normal distributions with the normal distribution under both Indep-NI-MLIRT and Dep-NI-MLIRT model frameworks, respectively. So approximately a total of 12800 computing hours were used. It would be a significant burden if we double or triple the computing hours to the SPH cluster. Furthermore, as the proportion of outliers become larger, it is expected that the parameter estimates of our proposed Indep-NI-MLIRT model will become worse. However, comparing to the normal distribution, the NI distribution can still account for some scale of outliers by change the turning parameter ν, for example, lower the degree of freedom in t distribution. Therefore, our

conclusion may maintain the same, that is, the use of the NI distribution assumption in the MLIRT model will have a better performance than the normal distribution when outliers exist. But it is interesting to study the robustness of the NI distributions under various degree of outliers. Further more, the violation of non-normality may be due to heavy tails or skewness or both. In this article, we considered only the influence of the violation of the normal assumptions due to heavy tails or outliers. For future work, we may investigate the influence of skewness under MLIRT framework using skewed normal distribution (Azzalini and Capitanio, 1999) and skewed normal independent distribution (Lachos et al., 2010). The informative dropout is another common issue involved in longitudinal studies. The ignorance of the "missing values" due to informative dropout may produce biased 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 may also be part of our future research.

Acknowledgements

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 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`.

Table 1: Simulation results of parameter estimation for N-MLIRT, Dep-CN-MLIRT, Indep-CN-MLIRT models with no outliers.

		N-MLIRT				Dep-CN-MLIRT				Indep-CN-MLIRT			
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
a_1	25.000	−0.055	0.552	0.512	0.920	0.037	0.519	0.515	0.950	−0.011	0.537	0.509	0.920
b_1	10.000	−0.054	0.403	0.371	0.930	−0.024	0.409	0.375	0.940	−0.040	0.406	0.373	0.930
a_2	80.000	−0.064	1.013	1.006	0.920	0.044	1.052	1.009	0.920	−0.063	1.005	1.000	0.950
b_2	18.000	−0.101	0.736	0.691	0.930	−0.054	0.753	0.699	0.940	−0.074	0.754	0.694	0.940
a_{31}	−2.700	−0.009	0.138	0.139	0.980	−0.021	0.127	0.140	0.980	−0.011	0.132	0.138	0.980
a_{32}	−0.600	0.004	0.132	0.121	0.900	−0.017	0.124	0.122	0.940	−0.006	0.130	0.120	0.920
a_{33}	2.000	0.019	0.146	0.131	0.920	−0.004	0.136	0.132	0.940	0.011	0.140	0.130	0.950
a_{34}	2.800	0.027	0.157	0.141	0.900	−0.006	0.143	0.141	0.950	0.021	0.147	0.140	0.950
a_{35}	5.000	0.018	0.206	0.186	0.950	0.023	0.186	0.187	0.930	0.011	0.194	0.186	0.950
a_{36}	6.000	0.046	0.227	0.214	0.950	0.039	0.209	0.214	0.950	0.038	0.218	0.213	0.930
b_3	2.000	−0.001	0.103	0.093	0.940	0.010	0.096	0.094	0.950	0.001	0.099	0.094	0.950
a_{41}	−0.100	−0.009	0.050	0.052	0.940	−0.000	0.053	0.052	0.950	−0.009	0.049	0.052	0.950
a_{42}	1.000	0.001	0.054	0.057	0.970	0.000	0.055	0.057	0.950	−0.003	0.051	0.057	0.990
a_{43}	1.800	0.006	0.068	0.068	0.950	0.009	0.068	0.068	0.940	0.003	0.066	0.068	0.950
a_{44}	2.600	0.010	0.083	0.087	0.950	0.003	0.092	0.087	0.930	0.008	0.082	0.087	0.960
a_{45}	3.300	0.030	0.111	0.112	0.950	0.021	0.110	0.111	0.940	0.034	0.112	0.112	0.950
a_{46}	4.000	0.063	0.152	0.149	0.920	0.040	0.134	0.147	0.970	0.059	0.150	0.148	0.930
b_4	0.400	−0.001	0.027	0.027	0.970	−0.000	0.028	0.027	0.930	−0.002	0.026	0.027	0.960
β_{10}	0.400	−0.002	0.088	0.091	0.960	0.003	0.087	0.091	0.960	−0.001	0.090	0.090	0.950
β_{11}	−0.500	0.003	0.109	0.126	0.960	0.002	0.132	0.124	0.950	−0.007	0.124	0.124	0.920
ρ	0.400	−0.008	0.042	0.045	0.990	−0.003	0.044	0.045	0.940	−0.009	0.044	0.045	0.980
σ_u	1.300	0.014	0.066	0.064	0.910	0.005	0.073	0.064	0.930	0.012	0.065	0.064	0.920

Table 2: Simulation results of parameter estimations for N-MLIRT, Dep-CN-MLIRT, Indep-CN-MLIRT models with 5% outliers.

		N-MLIRT				Dep-CN-MLIRT				Indep-CN-MLIRT			
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
a_1	25.000	**1.345**	0.605	0.891	**0.750**	0.147	0.558	0.531	0.920	0.151	0.580	0.526	0.930
b_1	10.000	**6.396**	**0.742**	**0.632**	**0.000**	−0.159	0.396	0.385	0.920	−0.030	0.422	0.390	0.960
a_2	80.000	**3.843**	1.091	1.597	**0.210**	−0.014	1.034	1.036	0.950	0.079	1.120	1.026	0.920
b_2	18.000	**12.576**	**0.960**	**1.162**	**0.000**	−0.260	0.733	0.709	0.920	−0.070	0.760	0.708	0.940
a_{31}	−2.700	−0.444	0.149	0.199	0.340	−0.013	0.137	0.142	0.960	−0.018	0.143	0.142	0.960
a_{32}	−0.600	−0.412	0.124	0.181	0.280	0.007	0.117	0.124	0.950	−0.007	0.126	0.123	0.940
a_{33}	2.000	−0.374	0.144	0.185	0.470	0.027	0.136	0.135	0.970	−0.002	0.149	0.133	0.920
a_{34}	2.800	−0.370	0.152	0.192	0.510	0.026	0.147	0.145	0.930	0.001	0.158	0.143	0.940
a_{35}	5.000	−0.378	0.193	0.229	0.660	0.036	0.186	0.192	0.950	0.007	0.199	0.190	0.960
a_{36}	6.000	−0.372	0.223	0.253	0.710	0.053	0.216	0.220	0.950	0.034	0.238	0.218	0.960
b_3	2.000	1.327	0.150	0.165	0.000	−0.024	0.107	0.096	0.920	0.003	0.112	0.097	0.910
a_{41}	−0.100	−0.083	0.054	0.059	0.790	0.005	0.052	0.053	0.930	−0.001	0.056	0.052	0.940
a_{42}	1.000	−0.080	0.054	0.063	0.800	0.000	0.054	0.057	0.970	−0.000	0.053	0.057	0.970
a_{43}	1.800	−0.078	0.064	0.073	0.860	0.005	0.062	0.069	0.970	0.001	0.066	0.069	0.960
a_{44}	2.600	−0.074	0.082	0.090	0.880	0.010	0.086	0.087	0.930	0.004	0.084	0.087	0.940
a_{45}	3.300	−0.069	0.102	0.113	0.900	0.025	0.106	0.112	0.950	0.014	0.099	0.111	0.960
a_{46}	4.000	−0.067	0.149	0.145	0.910	0.033	0.156	0.145	0.920	0.018	0.147	0.144	0.940
b_4	0.400	0.275	0.035	0.042	0.000	−0.004	0.026	0.025	0.940	0.001	0.026	0.025	0.940
β_{10}	0.400	0.019	0.062	0.076	0.970	−0.021	0.113	0.098	0.890	0.008	0.095	0.092	0.940
β_{11}	−0.500	**0.111**	0.083	0.089	**0.750**	0.026	0.138	0.127	0.940	−0.006	0.119	0.126	0.960
ρ	0.400	0.353	0.021	0.025	0.000	0.036	0.043	0.045	0.880	0.028	0.039	0.046	0.940
σ_u	1.300	−0.109	0.055	0.047	0.370	−0.005	0.060	0.063	0.940	−0.000	0.063	0.064	0.940

Large bias, large SE and SD, and poor CP are highlighted in bold.

Table 3: Simulation results of parameter estimation on $\boldsymbol{\beta}$ for model Indep-SL-MLIRT, Indep-T-MLIRT, Indep-CN-MLIRT with 5% outliers. The best fitting models are highlighted in bold.

Models	Parameters	True	Bias	SD	SE	CP
Dep-SL-MLIRT	β_{10}	0.400	0.176	0.154	0.139	0.720
	β_{11}	−0.500	−0.201	0.201	0.176	0.750
Dep-T-MLIRT	β_{10}	0.400	0.039	0.123	0.105	0.880
	β_{11}	−0.500	−0.036	0.153	0.135	0.890
Dep-CN-MLIRT	β_{10}	0.400	**-0.021**	**0.113**	**0.098**	**0.890**
	β_{11}	−0.500	**0.026**	**0.138**	**0.127**	**0.940**
Indep-SL-MLIRT	β_{10}	0.400	0.164	0.141	0.132	0.780
	β_{11}	−0.500	−0.200	0.176	0.175	0.810
Indep-T-MLIRT	β_{10}	0.400	0.077	0.118	0.109	0.850
	β_{11}	−0.500	−0.075	0.151	0.143	0.910
Indep-CN-MLIRT	β_{10}	0.400	**0.008**	**0.095**	**0.092**	**0.940**
	β_{11}	−0.500	**-0.006**	**0.119**	**0.126**	**0.960**

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

	LPML	DIC	EAIC	EBIC	BF
N-MLIRT	−27312.50	53975.06	52743.02	52822.64	>>100
Dep-CN-MLIRT	−26930.44	53158.58	51736.53	51825.51	>>100
Indep-CN-MLIRT	**-26873.84**	**52818.56**	**51149.79**	**51248.14**	Ref

Table 5: 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 ν and γ are the tuning parameters in the Dep-CN-MLIRT model. Parameters ν_1, γ_1 and ν_2, γ_2 are the turning parameters for the random effects and UPDRS in the Dep-CN-MLIRT and Indep-CN-MLIRT models, respectively.

	N-MLIRT		Dep-CN-MLIRT		Indep-CN-MLIRT	
	Mean (SE)	95% CI	Mean (SE)	95% CI	Mean (SE)	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
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
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
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

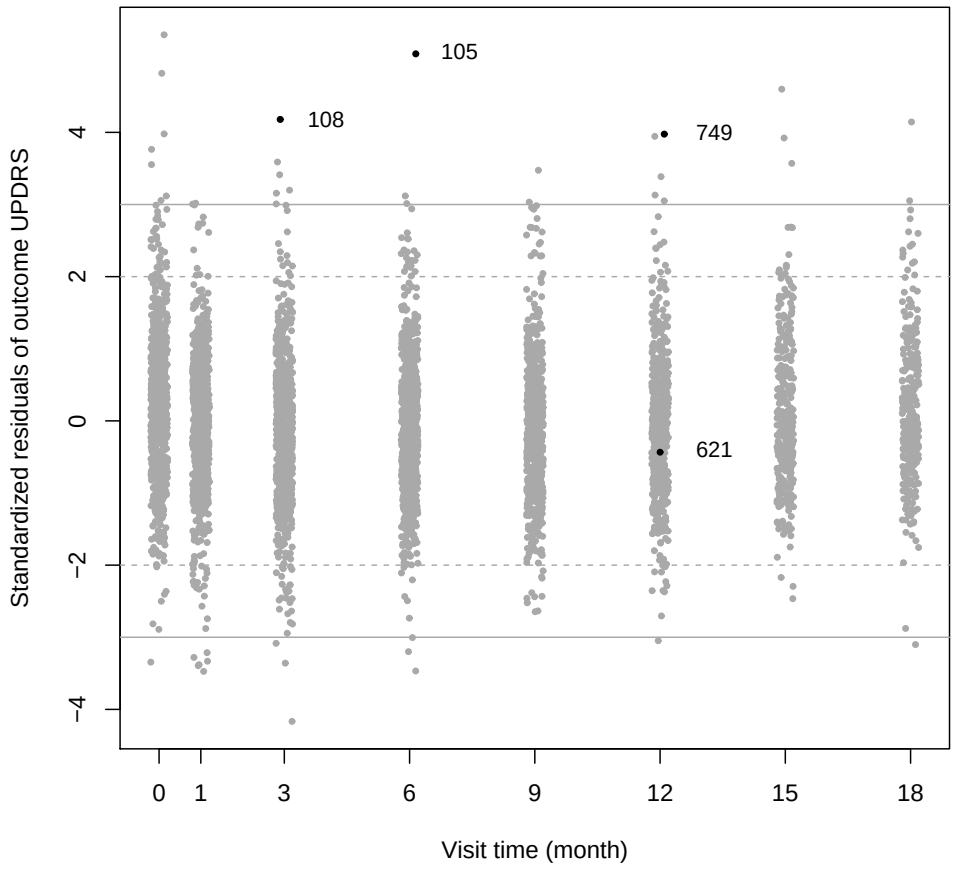

Figure 3: Standardized residuals of UPDRS for all patients at each visit when fitting N-MLIRT model, the dashed lines are horizontal lines at -2 and 2 and the solid lines are horizontal lines at -3 and 3. Numbers 105, 108, 621 and 749 denotes four patients to be used for further discussion.

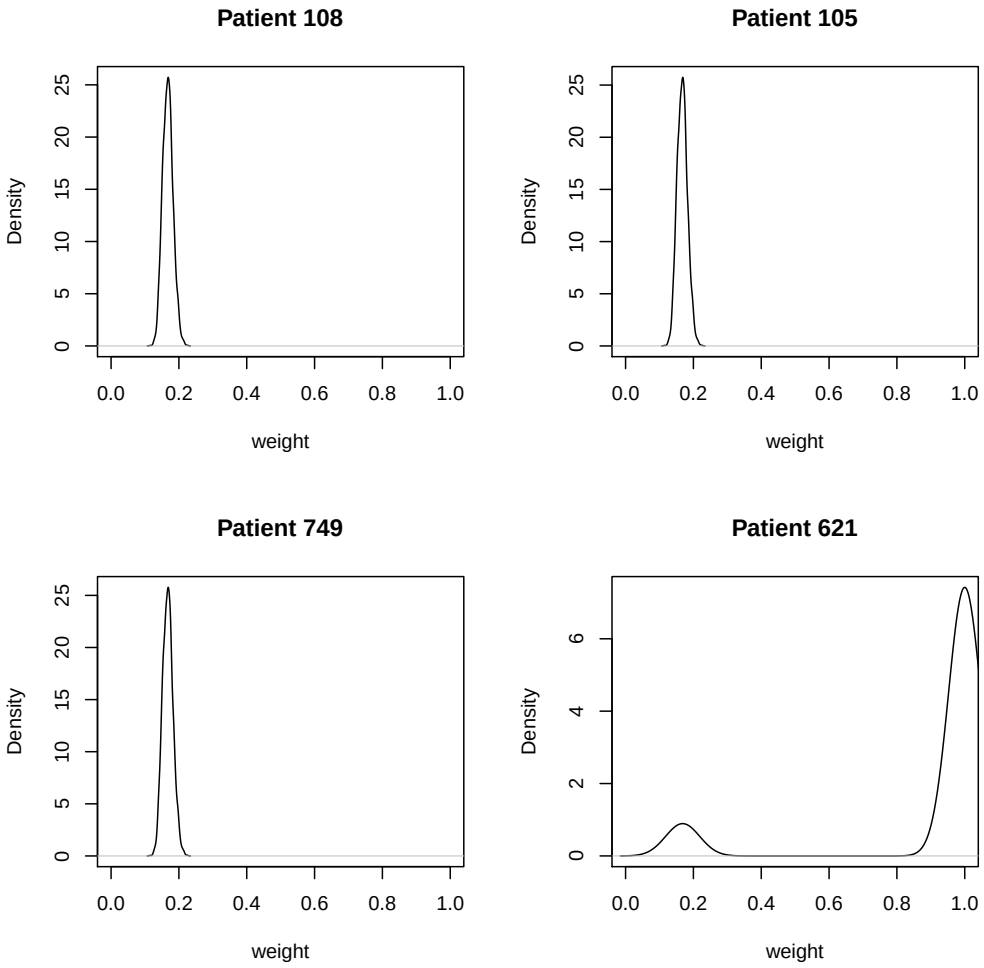

Figure 4: Estimates of the weight variable ω_{ijk} for patient 105, 108, 749 and 621.

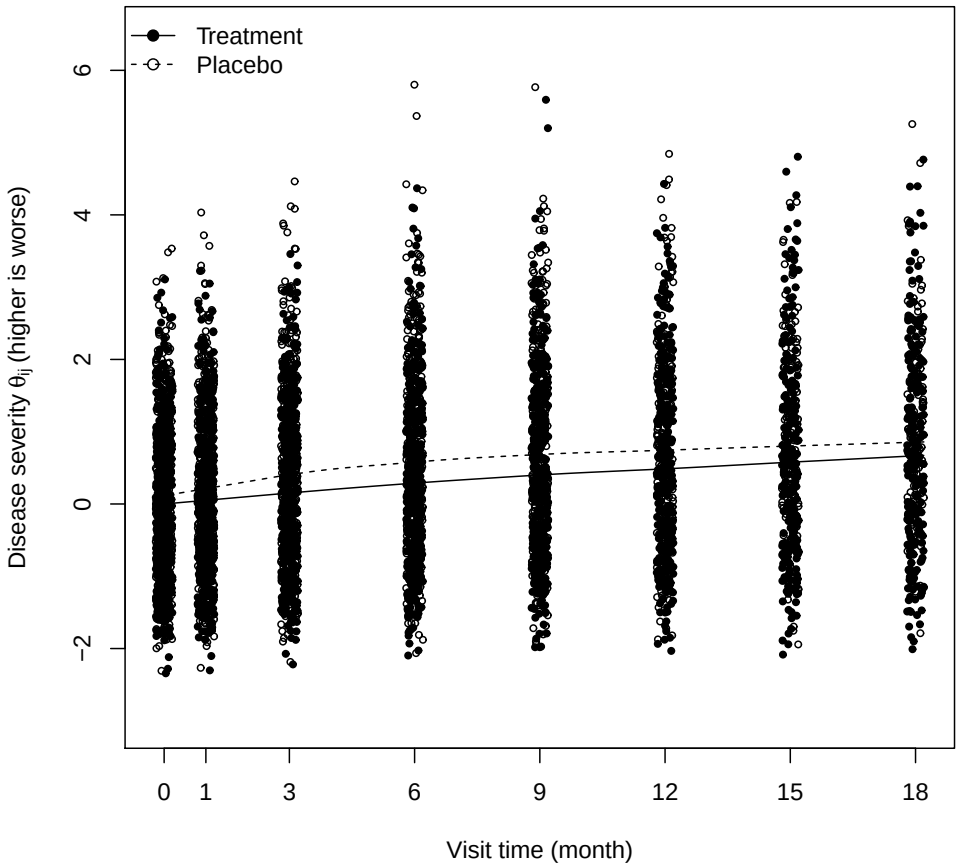

Figure 5: Bayesian posterior estimates of the subject-specific disease severity θ_{ij} at each visit and the lowess smooth curves for treatment and placebo group from Indep-CN-MLIRT model.

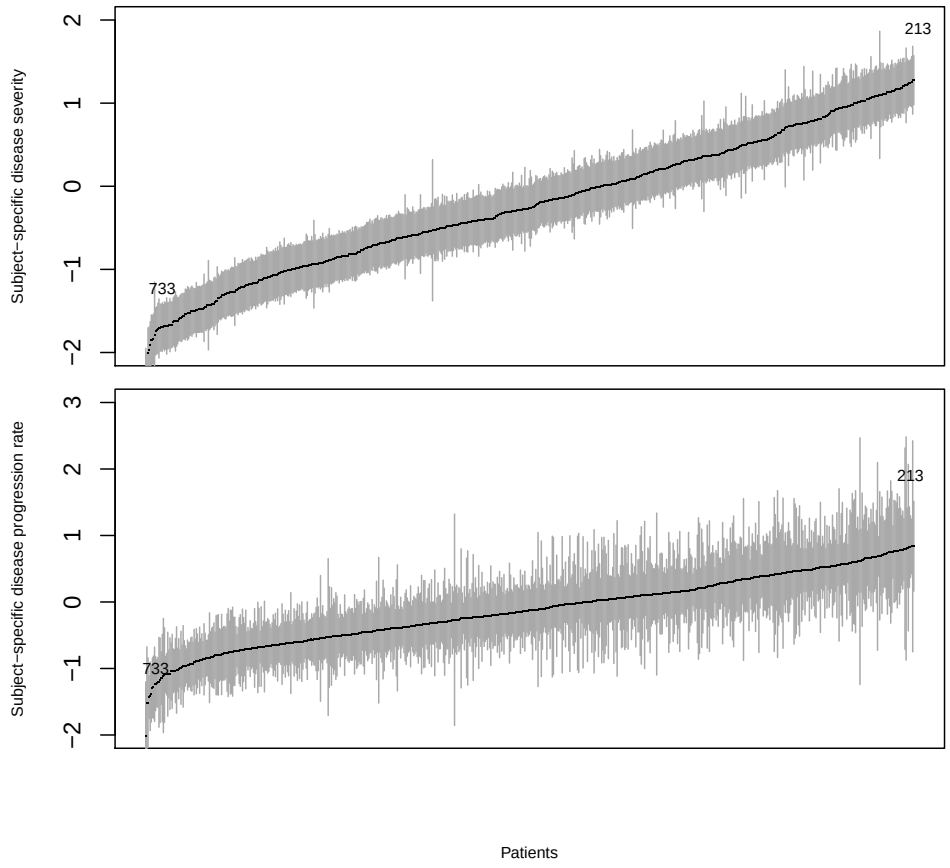

Figure 6: Bayesian posterior estimates of the rank of subject-specific disease severity u_{i0} (upper panel) and disease progression rate u_{i1} (lower panel) with 95% CI from Indep-CN-MLIRT model. The number in the figures are patient number.

Appendix

R code for data generation in simulation study

```
for (set in 1:nset)
{
    y1<-matrix(NA,nSub,nTime) #measurement 1, continuous outcome
    y2<-matrix(NA,nSub,nTime) #measurement 2, ordinal outcome
    y3<-matrix(NA,nSub,nTime) #measurement 3, ordinal outcome
    y4<-matrix(NA,nSub,nTime) #measurement 4, continuous outcome

# generate random effects
    U<- matrix(NA,nSub,2)
    for (i in 1:nSub)
        {
            U[i,] <- mvrnorm(1,mu=rep(0,2),Sigma=sigma)
        }

# generate latent variable theta
    theta <- NULL
    for (j in 1:nTime) theta <- cbind(theta, U[,1]+
                                (beta0+beta1*treat+U[,2])*t[j])

    # generate longitudinal outcomes y1, y2, y3 and y4
    for (i in 1:nSub)
        {
            # generate ordinal outcomes y2 and y3
            #for measurement 2, cumulative probability
            Q2<-matrix(NA,nTime,(n[1]-1))
            #for measurement 3, cumulative probability
            Q3<-matrix(NA,nTime,(n[2]-1))
            prob.y2<-matrix(NA,nTime,n[1])  #for measurement 2
            prob.y3<-matrix(NA,nTime,n[2])  #for measurement 2
            for (j in 1:nTime)
            {
            for (l in 1:(n[1]-1))
            {
            Q2[j,l]<-exp(a2[l]-b[2]*theta[i,j])/(1+exp(a2[l]-b[2]*theta[i,j]))
            }
            for (l in 1:(n[2]-1))
            {
            Q3[j,l]<-exp(a3[l]-b[3]*theta[i,j])/(1+exp(a3[l]-b[3]*theta[i,j]))
```

```
            }
prob.y2[j,1]<-Q2[j,1]
prob.y3[j,1]<-Q3[j,1]

for (k in 2:(n[1]-1))
{
prob.y2[j,k]<-Q2[j,k]-Q2[j,k-1]
}
for (k in 2:(n[2]-1))
{
prob.y3[j,k]<-Q3[j,k]-Q3[j,k-1]
}
prob.y2[j,n[1]]<-1-Q2[j,n[1]-1]
prob.y3[j,n[2]]<-1-Q3[j,n[2]-1]
y2[i,j]<-sample(1:n[1],1,prob=prob.y2[j,],replace=TRUE)
y3[i,j]<-sample(1:n[2],1,prob=prob.y3[j,],replace=TRUE)

                # generate continuous outcomes y1 and y4
#measure 1-continous
mu1<-a1+b[1]*theta[i,j]
                y1[i,j]<-rnorm(1,mu1,sd1)
mu4<-a4+b[4]*theta[i,j]
y4[i,j]<-rnorm(1,mu4,sd4)
                }
        }
outname <- set
    data<-list("nSub","n","y1","y2","y3","y4","zero","treat","t")
outputname <- paste("BUGSdata/", "data", outname, ".txt", sep="")
bugs.data(data, data.file=outputname)
}
```

BUGS code for fitting Indep-CN-MLIRT model

```
model
{
  for (i in 1:obs) # obs: number of total observations
  {
    # Contaminated Normal distribution
    u.contam2[i] ~ dbern(nu2)
    w2[i]<-gamma2*u.contam2[i] + (1-u.contam2[i])
    Y.conti[i] ~ dnorm(mu.conti[i], taudy[i]) # continuous outcome
    taudy[i]<-w2[i]*tau.conti
    # K.ordi: number of ordinal variables, K=2 in DATATOP study
```

```
    for (k in 1:K.ordi) { Y.ordi[i, k] ~ dcat(prob.y[i, k, 1:n[k]]) }
  }

  # construct mean of the continuous variable
  for (i in 1:obs)
    {
      mu.conti[i] <- a.conti + b.conti * theta[i]
    }

    # Construct the probability vector for the ordinal variables
    for (i in 1:obs)
    {
        for (k in 1:K.ordi)
            {
            for (l in 1:(n[k]-1)) { logit(psi[i, k, l]) <- a.ordi[k,l]
                                                    - b.ordi[k]*theta[i] }
            prob.y[i, k, 1] <- psi[i, k, 1]
            for (l in 2:(n[k]-1)) { prob.y[i, k, l] <- psi[i, k, l]
                                                    - psi[i, k, l-1] }
            prob.y[i, k, n[k]] <- 1 - psi[i, k, (n[k]-1)]
            }
    }

  # construct random effects
  for (i in 1:N)
    {
  u[i, 1:2] ~ dmnorm(zero[], precision[i, ,])
  u.contam1[i] ~ dbern(nu1)
  w1[i] <- gamma1*u.contam1[i] + (1-u.contam1[i]) # Contaminated Normal

 #construct the variance -covariance matrix for random effects
 precision[i,1:2,1:2]<-inverse(Sigma[i,,])
 Sigma[i,1,1]<-1/w1[i]
 Sigma[i,1,2]<-rho*sig/w1[i]
 Sigma[i,2,1]<-Sigma[i,1,2]
 Sigma[i,2,2]<-sig*sig/w1[i]
}

  # construct theta, the latent variable of subject i at time j
  for (i in 1:obs)
  {
     theta[i] <- u[subject[i], 1] + (beta[1] + beta[2]*treat[i]
                + u[subject[i], 2])*time[i]
```

```
  }

  # prior for regression coefficients
  for (i in 1:2)
  {
      beta[i] ~ dnorm(0, 0.01)
  }

  # specify prior distributions
  rho ~ dunif(-1, 1)
  sig ~  dgamma(0.01, 0.01)

  # prior for continuous variable's parameters
  b.conti ~ dgamma(0.001,0.001)
  a.conti ~ dnorm(0, 0.0005)
  tau.conti ~ dgamma(0.001,0.001)
  sd.conti <- 1/sqrt(tau.conti)

  #prior for ordinal variables' parameters
  for (k in 1:K.ordi)
  {
    b.ordi[k] ~ dgamma(0.001,0.001)
    a.ordi[k, 1] ~ dnorm(0,0.001)
    for (l in 2:(n[k]-1)) { a.ordi[k, l] <- a.ordi[k, l-1] + delta[k, l-1] }
    for (i in 1:(n[k]-2)) {delta[k, i] ~ dnorm(0,0.01)I(0,) }
  }

#prior for contaminated normal tuning parameters
    nu1~dbeta(1,1)
    gamma1~dbeta(1,1)
    nu2~dbeta(1,1)
    gamma2~dbeta(1,1)
}
```

3 Article 2: Bayesian hierarchical joint modeling using skew-normal/independent distributions (proposed for article submission to Statistical Methods in Medical Research)

Bayesian hierarchical joint modeling using skew-normal/independent distributions

Abstract

Many clinical trials often collect information on multiple longitudinal outcomes. Multilevel item response theory (MLIRT) models have been increasingly used to analyze these multivariate longitudinal data. Moreover, patients' follow-up may be stopped by some terminal events such as death or dropout and the time to the terminal events may be dependent on the multiple longitudinal outcomes. Joint modeling frameworks with a MLIRT sub-model for the multiple longitudinal outcomes and a survival sub-model have been previous developed to account for the dependent censoring. The continuous outcomes in the MLIRT sub-models are often assumed to be normally distributed. However, the normality assumption is often violated due to skewness and/or outliers and thus may produce biased results. The skew-normal/independent (SNI) distribution has been increasingly used to handle

the skewness and outlier problems to produce robust inference. In this article, we developed and studied several joint models with the continuous outcomes having the SNI distribution to account for three data features: skewness, outliers, and dependent censoring. Extensive simulation studies were conducted to evaluate the performance of various models. Specifically, in the simulation study we considered one continuous outcome and two ordinal outcomes with the continuous outcome has outliers and skewness. Our proposed methods were applied to the motivating DATATOP study for Parkinson's disease to investigate the effect of deprenyl in slowing down the disease progression.

Keywords. Clinical trial; Item-response theory; Latent variable; joint model; Skew-normal/independent distributions

3.1 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, e.g., binary, ordinal, and continuous. Clinical studies on Parkinson's disease (PD), for example, is 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 (Cummings, 1992; Fahn et al., 2004). The multidimensional nature of the disease exclude the method of using a single outcome to summarize or represent the overall disease severity (Huang et al., 2005; Luo et al., 2012). Therefore, clinical trials searching for a 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 (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).

The multivariate longitudinal data structure has three sources of correlation within and between outcomes of the same patient, i.e., inter-source (different measures at the same visit), longitudinal (same measure at different visits), and cross correlation (different 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) (Huang et al., 2005, 2009). 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 (Huang et al., 2009; Bandyopadhyay et al., 2011). The method of choosing one outcome as primary and other outcomes as secondary may encounter problems when the conclusion 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 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; Wang et al., 2002; Reise and Waller, 2009). 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; Luo et al., 2013; He and Luo, 2013). The MLIRT model consists of two levels of models. The first level of the MLIRT models 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 (He and Luo, 2013; Maier, 2001; Kamata, 2001).

The follow-up in longitudinal clinical studies may be stopped by terminal events. The terminal events could be noninformative (e.g. moving to other states or countries or traffic accidents) 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 referred

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; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997). In the MLIRT modeling framework, Wang *et al* (Wang et al., 2002) proposed a joint model to analyze multiple-item ordinal quality of life data in the presence of death. He and Luo (He and Luo, 2013) developed a joint model for multiple longitudinal outcomes of mixed types, subject to outcome-dependent terminal events. Luo (Luo, 2014) relaxed the proportional hazard (PH) assumption in He and Luo (He and Luo, 2013) and developed a joint modeling framework replacing the PH model by various parametric accelerated failure time (AFT) models.

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 due to skewness, heavy tails, or both (Lachos et al., 2011). Common approaches 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; Lachin, 2000). Data transformation may reduce information and the transformation may vary with different datasets (Lachos et al., 2011; Bandyopadhyay et al., 2010). Therefore, it is essential to replace the normality assumption with a more flexible distribution that accounts for the skewness and heavy tails in the continuous outcomes and to produce

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 (Bandyopadhyay et al., 2010; Lachos et al., 2009, 2010; Azzalini and Capitanio, 2003; Sahu et al., 2003; Bandyopadhyay et al., 2012). To the best of our knowledge, there is no literature discussing the SNI distribution in the MLIRT model framework. In this article, we propose a robust Bayesian parametric joint MLIRT model that accounts for both dependent terminal event time and the violation of the normality assumption due to skewness and outliers.

The rest of the article proceeds as follow. In Section 3.2, we describe a clinical trial that motivated the method development. In Sections 3.3, we discuss the MLIRT models, the SNI distributions, the Bayesian inference, as well as various model selection criteria. In Section 3.7, we present an extensive simulation study comparing the performance of various models. In section 3.8, we apply our proposed method to the DATATOP study dataset. Section 3.9 summarizes the main findings and discusses our future research.

3.2 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 2×2 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. Eight hundred 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, 1989, 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 in this article 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, started 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 and outcomes after the initiation of levodopa therapy were excluded from our study. Figure 7 shows the plot of mean UPDRS values at 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 that there is a strong association between the longitudinal outcomes and the time to the initiation of levodopa therapy.

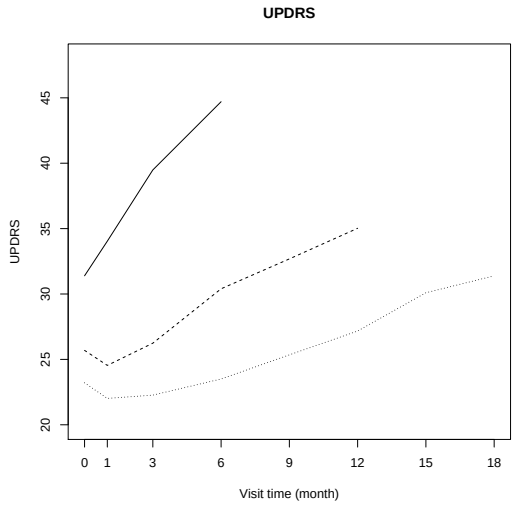

Figure 7: Mean UPDRS values with follow-up time less than 9 months (solid line), 9 to 15 months (dashed line) and more than 15 months (dotted line).

Figure 8 displays the histograms and the Q-Q plots of the residuals of the continuous outcome UPDRS by fitting the MLIRT model (18) (with normal assumption on the continuous outcome UPDRS) to the motivating DATATOP study. Figure 8 suggests the presence of skewness and outliers in the continuous outcome UPDRS. So in the following

sections, we will investigate how all three data features, i.e., skewness, outliers, and dependent censoring, would affect the inference of the MLIRT model. In addition, the validity of the asymptotic distribution of the GST test is based on the equal variance and covariance assumption on the longitudinal outcomes, however, this assumption remain to be checked in the DATATOP study.

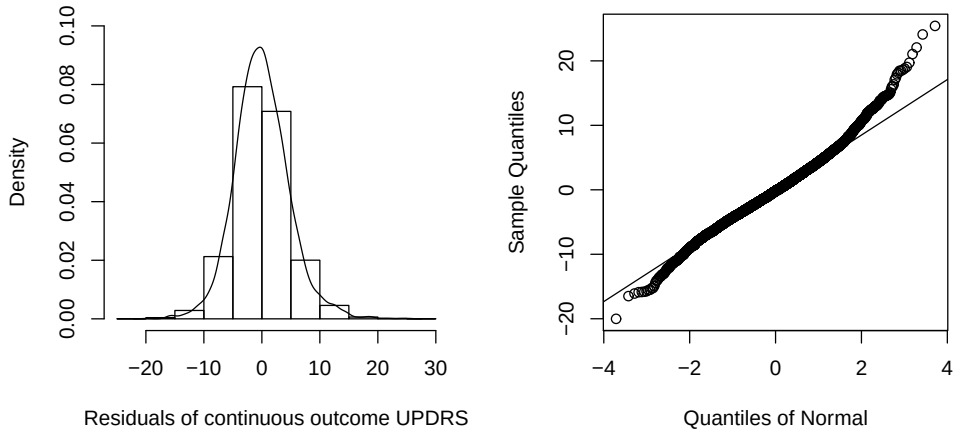

Figure 8: Histogram and Q-Q plot of residuals of the continuous outcome UPDRS obtained by fitting a joint MLIRT model using normal assumptions.

3.3 The model and estimation

3.4 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 ordinal or continuous and it is recoded 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 $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iK})'$ be

the outcome vector across visits. Let θ_{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 using θ_{ij} and subject-specific parameters. Specifically, we model the cumulative probabilities of ordinal outcomes using graded response sub-model (Samejima, 1997), and the continues outcomes using common factor sub-model (Lord et al., 1968) as follows.

$$\text{logit}\big\{p(y_{ijk} \leq l|\theta_{ij})\big\} = a_{kl} - b_k\theta_{ij}, \text{ with } l = 1, 2, \ldots, n_k - 1, \tag{14}$$

$$y_{ijk} = a_k + b_k\theta_{ij} + \epsilon_{ijk}, \tag{15}$$

where random error $\epsilon_{ijk} \sim N(0, \sigma_k^2)$ with σ_k^2 being variance of continuous outcome k, a_k is the outcome-specific "difficulty" parameter and b_k is the "discriminating" parameter that is always positive and describes the intensity to which outcome k discriminates between patients with latent disease severity θ_{ij}. Suppose the ordinal outcome k in model (14) 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 that patient i being in category l on outcome k at visit j is defined as $p(Y_{ijk} = l|\theta_{ij}) = p(Y_{ijk} \leq l|\theta_{ij}) - p(Y_{ijk} \leq l - 1|\theta_{ij})$.

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

$$\theta_{ij} = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0 + u_{i0} + (\boldsymbol{X}_{i1}\boldsymbol{\beta}_1 + u_{i1})t_{ij}, \tag{16}$$

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 which represents the subject-specific disease severity and u_{i1} is the random slope which 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 (Lord et al., 1968; Samejima, 1997). We impose the constraint $\text{Var}(u_{i0}) = 1$ to ensure model identifiability.

Let t_i denote the time to a terminal event for patient i, ζ_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 can be written as

$$h(t_i) = h_0(t_i)\exp(\boldsymbol{X}_i\gamma + \eta_0 u_{i0} + \eta_1 u_{i1}), \tag{17}$$

where γ is the unknown parameter vector for the potential risk factors $\boldsymbol{X}_i$ and η_0 and η_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 η_0 and η_1 are not equal to zero, the correlation between the survival time and longitudinal outcomes is incorporated. Specifically, positive parameters η_0 and η_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. To allow different baseline hazard rates for different time periods, 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 $0 = \tau_0 < \tau_1 < \ldots < \tau_m$ and the baseline hazard vector $\boldsymbol{g} = (g_0, g_1, \ldots, g_{m-1})'$, the piecewise constant hazard function can be defined as $h_0(t) = \sum_{l=0}^{m-1} g_l I_l(t)$, where $I_l(t)$ is an indicator function with $I_l(t) = 1$ if $\tau_l \leq t < \tau_{l+1}$ 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

$$L(\boldsymbol{y}_i, \boldsymbol{u}_i) = \left[\prod_{j=1}^{J_i}\prod_{k=1}^{K} P(y_{ijk}|\boldsymbol{u}_i)\right] \cdot h(t_i)^{\zeta_i} S(t_i) \cdot P(\boldsymbol{u}_i), \tag{18}$$

where the survival function $S(t_i) = \exp[-\int_0^{t_i} h(s)ds]$ 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 $\boldsymbol{\beta} = (\boldsymbol{\beta_0}^{'},\ \boldsymbol{\beta_1}^{'})^{'}$. Since we assume normal random errors for the continuous outcome in model (15), we refer to this joint modeling framework as model $\mathrm{JM_N}$ with the parameter vector $\boldsymbol{\Phi} = (\boldsymbol{a}^{'}, \boldsymbol{b}^{'}, \boldsymbol{\beta}^{'},\ \rho,\ \sigma_u,\ \sigma_k, \gamma^{'}, \eta_0, \eta_1, \boldsymbol{g}^{'})^{'}$. We refer to as reduced model $\mathrm{RM_N}$, the model assuming the survival time is independent to the longitudinal outcomes (i.e., $\eta_0 = \eta_1 = 0$). The parameter vector of model $\mathrm{RM_N}$ is $\boldsymbol{\Phi} = (\boldsymbol{a}^{'}, \boldsymbol{b}^{'}, \boldsymbol{\beta}^{'},\ \rho,\ \sigma_u,\ \sigma_k, \gamma^{'}, \boldsymbol{g}^{'})^{'}$.

3.5 The skew-normal/independent (SNI) distributions

The skew-normal/independent (SNI) distribution is an attractive class of asymmetric heavy-tailed distributions that includes the skew-t (ST), skew-slash, and skew-contaminated normal distributions. In this article, we adopt the ST distribution, while the extension

to other SNI distributions is straightforward. It is noted that ST distribution reduces to skew-normal distribution (SN, asymmetric but not heavy-tailed) when degrees of freedom are large. Different methods of specifying the SNI distributions have been discussed in literatures (Azzalini and Capitanio, 2003; Sahu et al., 2003; Arellano-Valle et al., 2007; Ghosh et al., 2007; Huang and Dagne, 2012). We consider the ST and SN distributions introduced by Sahu et al (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 to a continuous outcome k in model (15). Let δ_k be the skewness parameter for continuous outcome y_{ijk}, and ν be the degree of freedom for the ST distribution, then following Sahu et al. (Sahu et al., 2003), the Bayesian full model specification for ST distribution for y_{ijk} is given by

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

where the weight variable ω_{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 δ_k indicates the skewness of outcome k, with positive δ_k representing a right skewed distribution and negative δ_k representing a left skewed distribution. The parameter

Γ determines the prior variance information for δ_k. It is of note that when the degree of freedom $\nu \to \infty$, the distribution Gamma$(\nu/2, \nu/2)$ degenerates to 1, i.e., $\omega_{ijk} \equiv 1$. In this case, the ST distribution reduces to the SN distribution. Moreover, when the skewness parameter $\delta_k = 0$, the ST distribution reduces to the symmetric and heavy-tailed student-t distribution.

Incorporating the ST distribution, the full likelihood in (18) becomes

$$L(\boldsymbol{y}_i, \boldsymbol{u}_i, \omega_{ijk}, z_{ik}, \delta_k) = \prod_{j=1}^{J_i} \left[\prod_{k=1}^{K} P(y_{ijk}|\boldsymbol{u}_i, \omega_{ijk}, z_{ik}, \delta_k) P(\omega_{ijk}) P(z_{ik}) \right] h(t_i)^{\zeta_i} S(t_i) P(\boldsymbol{u}_i). \tag{19}$$

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

In summery, we consider the following models which account for various combinations of the three data features, i.e., skewness, outliers, and dependent censoring.

JM_{ST}: accounts for all three features;

JM_{SN}: accounts for skewness and dependent censoring, but not outliers;

JM_{T}: accounts for outliers and dependent censoring, but not skewness;

JM_{N}: accounts for dependent censoring, but not skewness and outliers;

RM_{ST}: accounts for skewness and outliers, but not dependent censoring;

RM_N: does not account for any of the three features.

3.6 Bayesian inference and model selection criteria

To infer the unknown parameter vector $\mathbf{\Phi}$, 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 $\mathbf{\Phi}$. Specifically, the prior distribution of all parameters in $\boldsymbol{\beta}$ and η_0, η_1 and γ are $N(0, 100)$. We use the prior distribution Gamma(0.001, 0.001) for all components in $\boldsymbol{b}$ and σ_u to ensure positivity, and use Uniform$[-1, 1]$ for ρ. 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 ν and δ_k. 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., 2004).

Among the various model selection methods available in Bayesian inference, we select the following five: the log pseudo-marginal likelihood (LPML), deviance information criterion (DIC), expected Akaike information criterion (EAIC), expected Bayesian information criterion (EBIC), and Bayes factor (BF). Conditional predictive ordinate (CPO) (Geisser, 1993; Carlin and Louis, 2011) is a cross-validation predictive method that evaluates the predictive distribution of the model conditioning on the data but with single data point

deleted (Lachos et al., 2009; Chen et al., 2000). 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 $CPO_i = p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}) = \int p(\boldsymbol{y_i}|\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\boldsymbol{y}_{(i)})d\boldsymbol{\Phi}$. Large CPO suggests that the data for subject i can be well predicted by the model based on the data from all the other subjects. Thus lager CPO indicates a better model fit. Because there is no close form for CPO_i for our proposed models, we adopt a Monte Carlo estimation obtained from the posterior distribution $p(\boldsymbol{\Phi}|\boldsymbol{y})$. Since the function of CPO_i can be further derived as $CPO_i = p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}) = p(\boldsymbol{y})/p(\boldsymbol{y}_{(i)}) = 1/\int p(\boldsymbol{\Phi}|\boldsymbol{y})/p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)}, \boldsymbol{\Phi})d\boldsymbol{\Phi}$, then a harmonic-mean approximation of CPO_i is $\widehat{CPO_i} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(\boldsymbol{y_i}|\boldsymbol{y}_{(i)},\boldsymbol{\Phi}^{(t)})})^{-1} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(\boldsymbol{y_i}|\boldsymbol{\Phi}^{(t)})})^{-1}$, where M be the total number of post burn-in samples (Luo et al., 2013; Chen et al., 2000). A summary statistics of CPO_i is log pseudo-marginal likelihood (LPML), defined as $LPML = \sum_{i=1}^{N}\log(\widehat{CPO_i})$. 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). The deviance statistics is defined as $D(\boldsymbol{\Phi}) = -2\log f(\boldsymbol{y}|\boldsymbol{\Phi}) + 2\log h(\boldsymbol{y})$, where $f(\boldsymbol{y}|\boldsymbol{\Phi})$ is the likelihood of the observed data $\boldsymbol{y}$ given parameter vector $\boldsymbol{\Phi}$, $h(\boldsymbol{y})$ is a standardized function of the data alone and have no impact on the assessment of the model fit. Let $\bar{D}(\boldsymbol{\Phi}) = E_{\boldsymbol{\Phi}|\boldsymbol{y}}[D(\boldsymbol{\Phi})]$ be the posterior mean of the deviance and let $D(\bar{\boldsymbol{\Phi}}) = D(E_{\boldsymbol{\Phi}|\boldsymbol{y}}[\boldsymbol{\Phi}])$ be the deviance evaluated at the posterior mean of the parameter vector $\boldsymbol{\Phi}$. The DIC is defined as $DIC = \bar{D}(\boldsymbol{\Phi}) + p_D$, where $p_D = \bar{D}(\boldsymbol{\Phi}) - D(\bar{\boldsymbol{\Phi}})$ is the number of effective parameters and it captures the complexity of the model. A smaller value of DIC indicates better model fit. Moreover, the expected Akaike information criterion (EAIC) and the

expected Bayesian information criterion (EBIC) (Carlin and Louis, 2011) are defined as $EAIC = \bar{D}(\mathbf{\Phi}) + 2p$ and $EBIC = \bar{D}(\mathbf{\Phi}) + p \log N$, respectively, where p is the total number of parameters in the model and N is the sample size. Smaller values of EAIC and EBIC 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 (Lavine and Schervish, 1999; Lewis and Raftery, 1997). Let two competing models be M_1 and M_2. Then for observed data $\boldsymbol{y}$, BF in favor of model M_1 over M_2 is defined as

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

where $\mathbf{\Phi_i}$ is the parameter vectors for model M_i for $i = 1, 2$; $f(\boldsymbol{y}|\mathbf{\Phi_i}, M_i)$ is the likelihood of model M_i; and $f(\mathbf{\Phi_i}|M_i)$ is the posterior density of $\mathbf{\Phi_i}$ for model M_i (Lewis and Raftery, 1997; Gelman et al., 2013). 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 $f(\boldsymbol{y}|M_i)$ (Lewis and Raftery, 1997). Specifically, the $f(\boldsymbol{y}|M_i) \approx (2\pi)^{d_i/2} |\mathbf{\Sigma_i}|^{1/2} f(\boldsymbol{y}|\bar{\mathbf{\Phi}}_\mathbf{i}, M_i) f(\bar{\mathbf{\Phi}}_\mathbf{i}|M_i)$, where d_i is the number of parameters in $\mathbf{\Phi_i}$, $\mathbf{\Sigma_i}$ is the posterior covariance matrix of $\mathbf{\Phi_i}$, $\bar{\mathbf{\Phi}}_\mathbf{i}$ is the posterior mean of $\mathbf{\Phi_i}$, $f(\bar{\mathbf{\Phi}}_\mathbf{i}|M_i)$ is the prior probability of parameters evaluated at $\bar{\Phi}_i$, and $f(\boldsymbol{y}|\bar{\mathbf{\Phi}}_\mathbf{i}, M_i)$ is the likelihood evaluated at the posterior mean $\bar{\Phi}_i$ (He and Luo, 2013; Lewis and Raftery, 1997). The interpretation of the BF is summarized by Kass and Raftery (1995)(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.

3.7 Simulation studies

In this section, we conducted an extensive simulation study with two settings to compare the performance of the models JM_{ST}, JM_{SN}, JM_{T}, JM_{N}, RM_{ST}, and RM_{N}. In both settings, we generated 400 datasets with a sample size of 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 $h(t_i) = h_0(t_i)\exp(\boldsymbol{X}_i\gamma + \eta_0 u_{i0} + \eta_1 u_{i1})$. 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 3.5 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 $\mu = a_1 + b_1\theta_{ij}$, $\sigma_1 = 5$ and skewness parameter $\delta = 7$. Then, to generate additional 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 6 we only presented the results from the four joint models, i.e., JM_N, JM_T, JM_SN, and JM_ST, while the results from the reduced models RM_N and RM_ST can be found in the appendix. 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. Overall, the models which ignored skewness in the continuous outcome (models JM_T, JM_N, and RM_N) overestimated the difficulty parameter a_1, while the models which ignored the outlier issue in the continuous variable (models JM_SN, JM_N, and RM_N) overestimated the standard deviation of the continuous outcome σ_1. All models except model JM_ST had large bias on the parameters b_1. Comparing with model JM_ST, the reduced model RM_ST had large bias on the parameters β_{10} (disease progression rate for the patients in the placebo group), β_{11} (the change in disease progression rate by the treatment), and γ (treatment effect on the survival time), a phenomenon also reported in He and Luo (He and Luo, 2013). This is because that by ignoring the dependent censoring (the subjects with more severe disease severity and faster disease

Table 6: 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
σ_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
β_{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
β_{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
ρ	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
σ_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
γ	−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
η_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
η_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 CR are highlighted in bold.

progression tend to have a terminal event earlier), the reduced model tends to reduce the difference between two groups and therefore underestimate the treatment effect in both the longitudinal and survival models. Moreover, Table 6 suggests that by ignoring the issues of skewness and/or outliers in the continuous outcome, the joint models $\mathrm{JM_N}$, $\mathrm{JM_T}$, and $\mathrm{JM_{SN}}$ gave biased estimates for the parameters β_{10} and β_{11}, and the CPs are away from the nominal value of 0.95. This previously unknown and commonly ignored phenomenon indicates that it is essential to account for the skewness and outliers in the continuous outcome in the MLIRT joint modeling framework to obtain correct estimates in disease progression and treatment effects.

In setting II, we generated data from model $\mathrm{JM_N}$ with no skewness and outliers in the continuous outcome and presented the results from models $\mathrm{RM_{ST}}$, $\mathrm{JM_N}$, and $\mathrm{JM_{ST}}$ in Table 7. Due to space constraint, we did not present the outcome-specific parameters $\boldsymbol{a}$ and $\boldsymbol{b}$, but the estimation of all these parameters was seasonally good. Table 7 suggests that comparing with model $\mathrm{JM_N}$, the proposed model $\mathrm{JM_{ST}}$ produced comparable and satisfactory results, i.e., the bias was negligible, SE was close to SD, and the overage probabilities were reasonably close to nominal level of 0.95. Under the model overparameterization, model $\mathrm{JM_{ST}}$ is still a reasonable choice. However, model $\mathrm{RM_{ST}}$ 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 $\mathrm{JM_{ST}}$ produced accurate parameter estimates while all other models provided severely biased estimates for some parameters. In the absence of skewness and

Table 7: Setting II: simulation results from models RM_{ST}, JM_N, and JM_{ST} when there were no skewness and outliers in the continuous outcome.

		RM_{ST}				JM_N				JM_{ST}			
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
β_{10}	0.400	**−1.229**	0.163	0.170	**0.000**	−0.009	0.183	0.179	0.921	−0.000	0.190	0.182	0.924
β_{11}	−0.500	**0.367**	0.188	0.196	**0.532**	−0.003	0.190	0.197	0.958	−0.010	0.206	0.199	0.965
γ	−1.000	**0.300**	0.169	0.109	**0.327**	−0.006	0.179	0.175	0.929	−0.014	0.175	0.175	0.935
η_1	0.400	—	—	—	—	0.009	0.104	0.106	0.955	0.021	0.105	0.107	0.953
η_2	1.000	—	—	—	—	0.015	0.110	0.110	0.950	0.003	0.112	0.110	0.959

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

outliers in the continuous outcome, the overparameterized model JM_{ST} provided results comparable with model JM_N. In both simulation settings, the reduced models had large bias in the parameters related to the disease progression and treatment effects.

3.8 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 two reduced models RM_N and RM_{ST}. To fit all these models, we combined some categories with zero or small counts so that HY and SEADL have 5 and 6 categories, respectively. We also reverse-coded the SEADL variable so that all three outcomes have higher value indicating worse clinical condition. We removed one patient which has no UPDRS measurements across all visits. Therefore, there were 398 and 401 patients in the treatment and placebo groups, and respectively. There were 153 and 222 patients in

treatment and placebo groups, respectively, experienced the terminal event, i.e., 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 (16) 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 $h(t_i) = h_0(t_i)\exp(\gamma x_i + \eta_0 u_{i0} + \eta_1 u_{i1})$. To specify the baseline hazard $h_0(t_i)$, we divided the survival time into $K = 3$ intervals by every $1/Kth$ 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 8: 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	BF
JM_{ST}	**−24832.28**	**51079.95**	**48319.02**	**48431.42**	Ref
RM_{ST}	−25103.57	51514.16	48897.54	49000.57	>>100
JM_{SN}	−25416.64	52187.91	50163.75	50271.47	>>100
JM_{T}	−26154.57	53863.16	51772.63	51880.35	>>100
JM_{N}	−26571.84	54689.14	53295.50	53398.54	>>100
RM_{N}	−26820.56	55150.33	53919.96	54013.62	>>100

Table 8 displays various model comparison criteria obtained from different models.

The proposed model JM_{ST} performed significantly better than all the other models with largest LPML value and smallest DIC, EAIC and EBIC values, suggesting the need of accounting for the outliers and skewness in the continuous outcome UPDRS, as well as the correlation between the longitudinal outcomes and the time to the initiation of levodopa therapy. Hence, we selected model JM_{ST} as the final model.

As introduced in Section 3.6, BF is a Bayesian alternative to p value for hypothesis testing for competing models. In our model settings, when $\delta = 0$, model JM_{ST} reduces to model JM_{T}; when $\omega_{ijk} = 1$ or equivalently $\nu \to \infty$, model JM_{ST} reduces to model JM_{SN}; when $\eta_0 = \eta_1 = 0$, model JM_{ST} reduces to model RM_{ST}; when both $\delta = 0$ and $\omega_{ijk} = 1$, model JM_{ST} reduces to model JM_{N}; and when $\delta = \eta_0 = \eta_1 = 0$ and $\omega_{ijk} = 1$, model JM_{ST} reduces to model RM_{N}. We computed the BFs to compare all the models stated above with our proposed model JM_{ST}. The BF results were summarized in Table 8. Decisive evidence (BF $>>$ 100) was shown in favor of JM_{ST} over all other models, suggesting that it was necessary to account for all three data features (outliers, skewness, and dependent censoring) in the DATATOP study.

Table 9 displays the posterior means and standard deviation (SD) from all six models. The skewness parameter δ was significantly positive in both models JM_{SN} and JM_{ST}, suggesting the existence of right skewness in the outcome UPDRS. To visualize this, Figure 9 displays the posterior density distribution of δ from models JM_{SN} (left panel) and JM_{ST} (right panel). It suggests that the UPDRS variable followed a right skewed distribution with 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 10

Table 9: 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 ν and δ are the tuning parameter and skewness parameter, respectively, for the outcome UPDRS in model JM_{ST}.

	RM_{N} 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)	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)	9.691(0.355)	10.731(0.290)	10.661(0.257)	9.804(0.288)	9.568(0.294)
σ_1	5.231(0.075)	3.419(0.085)	5.224(0.075)	3.849(0.108)	4.861(0.062)	3.449(0.087)
ν	—	3.552(0.240)	—	4.250(0.368)	—	3.668(0.262)
δ	—	12.258(0.430)	—	—	11.958(0.410)	12.159(0.411)
a_{21}	−0.871(0.068)	−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)	0.118(0.056)	0.108(0.065)	0.095(0.055)	0.138(0.058)	0.128(0.054)
a_{23}	3.218(0.086)	3.047(0.077)	3.230(0.084)	3.197(0.078)	3.082(0.079)	3.071(0.078)
a_{24}	5.471(0.135)	5.296(0.128)	5.480(0.132)	5.439(0.129)	5.327(0.132)	5.307(0.129)
b_2	1.400(0.053)	1.323(0.059)	1.385(0.050)	1.368(0.046)	1.317(0.051)	1.324(0.052)
a_{31}	−2.555(0.094)	−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)	−0.563(0.112)	−0.501(0.080)	−0.503(0.067)	−0.531(0.115)	−0.548(0.106)
a_{33}	1.932(0.090)	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)	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)	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)	3.019(0.123)	1.814(0.062)	1.759(0.058)	2.999(0.099)	2.993(0.104)
β_{10}	1.286(0.068)	1.272(0.070)	1.433(0.068)	1.367(0.063)	1.504(0.063)	1.415(0.063)
β_{11}	−0.609(0.071)	−0.579(0.072)	−0.681(0.073)	−0.630(0.069)	−0.730(0.080)	−0.655(0.064)
ρ	0.365(0.043)	0.329(0.044)	0.389(0.042)	0.388(0.040)	0.338(0.040)	0.350(0.040)
σ_u	0.817(0.042)	0.827(0.046)	0.897(0.045)	0.861(0.043)	0.968(0.045)	0.905(0.044)
γ	−0.680(0.105)	−0.685(0.106)	−1.046(0.143)	−1.059(0.148)	−1.059(0.150)	−1.066(0.141)
η_1	—	—	0.333(0.068)	0.324(0.070)	0.428(0.070)	0.448(0.072)
η_2	—	—	1.466(0.120)	1.534(0.130)	1.309(0.104)	1.418(0.115)

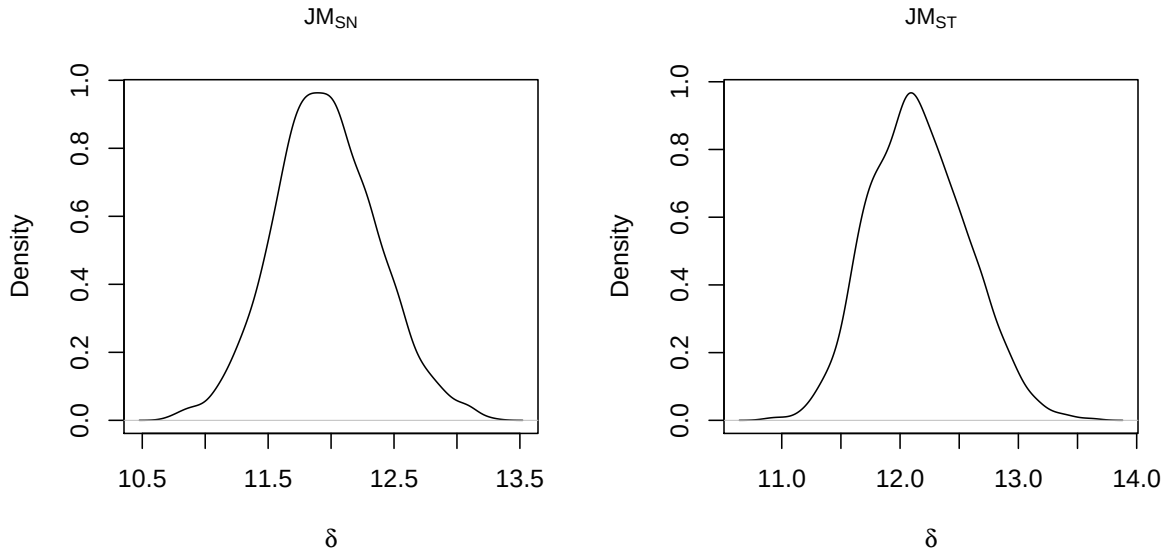

Figure 9: Posterior density functions of the skewness parameter δ from models JM_{SN} (left panel) and JM_{ST} (right panel).

displays the posterior density distributions of the degree of freedom (df) ν from models JM_{T} (left panel) and JM_{ST} (right panel). For both models, the densities of ν 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 in the UPDRS variable. Comparing with model JM_{ST}, the models without accounting for skewness in the UPDRS variable (models JM_{T}, JM_{N}, and RM_{N}) provided larger estimate to its difficulty parameter a_1, a phenomenon also observed in simulation setting I. Moreover, comparing with model JM_{ST}, the models without considering heavy tails in the UPDRS variable (models JM_{SN}, JM_{N}, and RM_{N}) provided larger estimate to its standard deviation (σ_1), which was also reported in the simulation setting I.

Table 9 suggested that all six models demonstrated significant disease progression rate

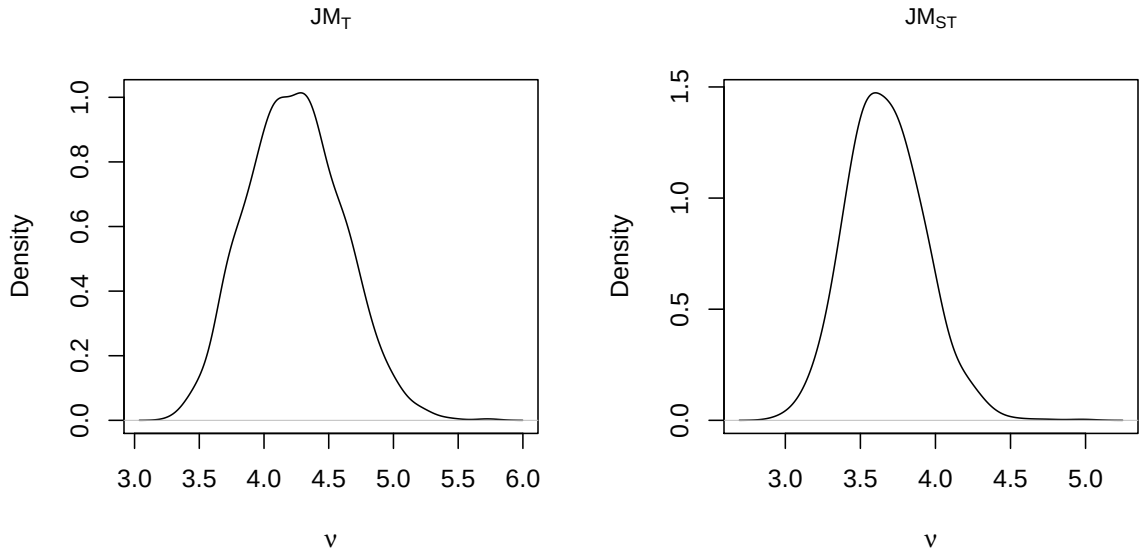

Figure 10: Posterior density functions of the degree of freedom (df) ν from models $\mathrm{JM_T}$ (left panel) and $\mathrm{JM_{ST}}$ (right panel).

for placebo patients (β_{10}) and significant deprenyl effects in delaying the progression of PD symptoms (β_{11}). The results were consistent with the findings from the DATATOP study analysis (Parkinson Study Group, 1993). The reduced models $\mathrm{RM_N}$ and $\mathrm{RM_{ST}}$ provided smaller estimates of the of the treatment effects (β_{10}, β_{11}, and γ), as comparing with their joint model counterparts. Similar phenomenon was also observed in the simulation setting I. In model $\mathrm{JM_{ST}}$, the disease progression rate for placebo patients (β_{10}) was 1.415 (95% Cl: [1.302, 1.536]) and the changes in disease progression rate by deprenyl (β_{11}) was -0.655 (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 η_0 (mean: 0.448, 95%CI: [0.302, 0.586]) and η_1 (mean: 1.418, 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.

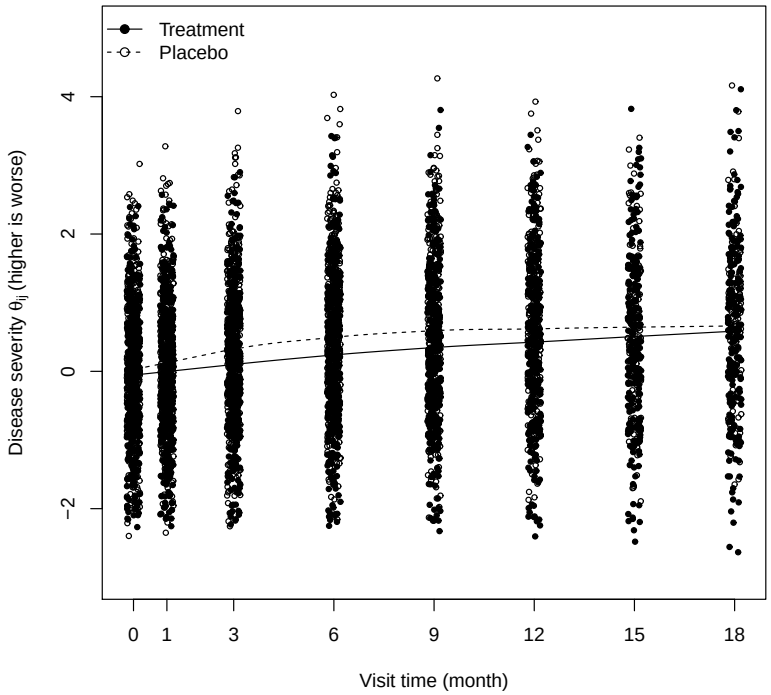

Figure 11: Bayesian posterior estimates of the subject-specific disease severity θ_{ij} at each visit and the lowess smooth curves for treatment and placebo groups from model JM_{ST}.

To provide insight into the latent disease progression at each visit, Figure 11 displays the posterior estimates of the subject-specific latent disease severity θ_{ij} for all patients at all visits. The estimates of θ_{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 the treatment and placebo groups, respectively. Figure 11 suggested 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

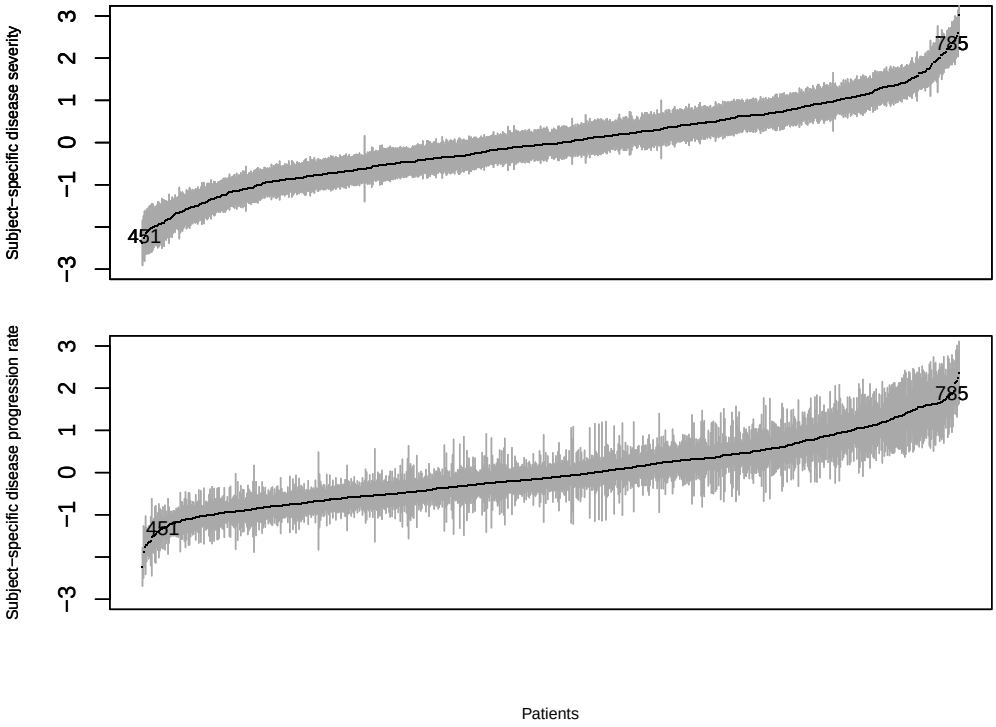

Figure 12: The ranking of the Bayesian posterior estimates of the subject-specific disease severity u_{i0} (upper panel) and disease progression rate u_{i1} (lower panel) with 95% CI, obtained from model JM_{ST}. The numbers in the figures are patient numbers.

slow progression rate overtime. Overall, the treatment patients had lower disease severity across all visits than the placebo patients. In addition, from Table 9, 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 ρ, 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 12 displays the ranked u_{i0} (upper panel) and ranked u_{i1} (lower panel) for each patient. To visualize the effect of positive correlation coefficient ρ, 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.

3.9 Conclusions

In this article, we developed a joint modeling framework that consists of an MLIRT submodel for multivariate longitudinal outcomes and a Cox submodel for the dependent terminal event while accounting for the skewness and heavy tails in the continuous outcomes by using the more flexible skew-normal/independent (SNI) distributions. Simulation studies demonstrated that when there were one continuous outcome and two ordinal outcomes, and both skewness and outliers exist 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 in the continuous outcome UPDRS. We demonstrated that model JM_{ST} was the best fitting model to the DATATOP dataset based on the four model selection criteria. The BF results indicated decisive evidences in favor of model JM_{ST} over all other models which did not account for all three data

features. 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 offered the 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` provide 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 to account for multiple competing dependent censoring events. 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. It is of interest to investigate our joint model's performance when the underlying random effects distribution is symmetric non-normal

or even asymmetric.

Acknowledgements

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 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.

Appendix

Table 10: Setting I: simulation results from models RM_N and RM_ST when there were skewness and 5% outliers in the continuous outcome.

		RM_N				RM_ST			
	True	Bias	SD	SE	CP	Bias	SD	SE	CP
a_1	25.000	**4.390**	0.774	0.713	0.000	0.421	1.055	1.074	0.941
b_1	10.000	0.345	0.801	0.657	0.843	−0.147	0.525	0.509	0.946
σ_1	5.000	**7.447**	1.332	0.376	0.000	−1.369	0.247	0.257	0.008
a_{21}	−2.700	0.099	0.205	0.183	0.858	−0.045	0.197	0.187	0.925
a_{22}	−0.600	0.018	0.149	0.140	0.925	−0.029	0.148	0.142	0.936
a_{23}	2.000	−0.060	0.158	0.160	0.923	−0.001	0.167	0.163	0.941
a_{24}	2.800	−0.069	0.187	0.183	0.915	0.018	0.192	0.187	0.938
a_{25}	5.000	−0.074	0.345	0.316	0.918	0.086	0.353	0.321	0.928
a_{26}	6.000	0.076	0.512	0.467	0.948	0.256	0.518	0.473	0.922
b_2	2.000	−0.103	0.152	0.145	0.871	−0.005	0.148	0.144	0.936
a_{31}	−0.100	0.009	0.074	0.076	0.954	0.004	0.074	0.075	0.949
a_{32}	1.000	0.012	0.082	0.083	0.954	0.010	0.083	0.082	0.952
a_{33}	1.800	0.012	0.096	0.103	0.961	0.009	0.096	0.102	0.971
a_{34}	2.600	0.032	0.142	0.140	0.951	0.031	0.142	0.140	0.968
a_{35}	3.300	0.069	0.196	0.193	0.938	0.068	0.197	0.193	0.941
a_{36}	4.000	0.149	0.285	0.278	0.928	0.149	0.289	0.279	0.928
b_3	0.400	0.004	0.060	0.060	0.954	−0.007	0.057	0.056	0.946
β_{10}	0.400	**−1.459**	0.191	0.188	**0.000**	**−1.278**	0.173	0.173	**0.000**
β_{11}	−0.500	**0.464**	0.214	0.207	**0.374**	**0.388**	0.194	0.199	**0.504**
ρ	0.400	−0.223	0.147	0.164	0.714	−0.212	0.107	0.111	0.477
σ_u	1.300	−0.526	0.124	0.118	0.034	−0.334	0.101	0.101	0.147
γ	−1.000	**0.304**	0.158	0.109	**0.273**	**0.305**	0.157	0.109	**0.273**
η_1	0.400	—	—	—	—	—	—	—	—
η_2	1.000	—	—	—	—	—	—	—	—

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

R code for data generation in simulation study

```
for (myloop in 1:nsim)
{
  set.seed(mySeeds[myloop])
  # generate random effects
  U <- mvrnorm(N, mu=rep(0,2), Sigma=Sigma.U)

  # generate latent variable theta
  theta <- NULL
  for (j in 1:T) theta <- cbind(theta, U[,1]+
        (beta1[1] + beta1[2]*treat.pts + U[,2])*time[j])

  # generate longitudinal outcomes
  Y.conti <- array(NA,dim = c(N, T))
  Y.conti.2 <- array(NA,dim = c(N, T))
  Y.ordi <- array(NA,dim = c(N, T, K.ordi))
  for (i in 1:N) {
    for (j in 1:T) {
      Y.conti[i,j] <- rnorm(1, mean=a.conti+b.conti*theta[i,j],
                            sd=sd.conti)
      for (k in 1:K.ordi) {
        psi <- c(rep(0,n.ordi[k]-1),1)
        prob.y <- rep(0,n.ordi[k])
        for (l in 1:(n.ordi[k]-1)) psi[l] <- expit(a.ordi[k,l]
                                   - b.ordi[k]*theta[i,j])
        prob.y[1] <- psi[1]
        for (l in 2:n.ordi[k]) prob.y[l] <- psi[l]-psi[l-1]
        Y.ordi[i,j,k] <- sample(1:n.ordi[k],1,prob=prob.y)
      }
    }
  }

  # turn response Y and theta into long format
  Y.conti.temp <-  Y.conti.2.temp <- Y.ordi.temp <- theta.temp <- NULL
  for (i in 1:N) {
    Y.conti.temp <- c(Y.conti.temp, Y.conti[i,])
    Y.ordi.temp <- rbind(Y.ordi.temp, Y.ordi[i,,])
    theta.temp <- c(theta.temp, theta[i,])
  }
  Y.conti <- Y.conti.temp; rm(Y.conti.temp)
  Y.ordi <- Y.ordi.temp; rm(Y.ordi.temp)
  theta <- theta.temp; rm(theta.temp)
```

```
# simulation for survival part
S <- runif(N) #simualte survival function from unif(0,1)
Ti.obs <- -log(S)/(exp(gam*treat.pts+ omega2*U[,1] + omega3*U[,2])*h0)
C <- runif(N, 10, 20)
tee <- pmin(Ti.obs, C) #observed time is min of censored and true
event <- ifelse(tee==Ti.obs, 1, 0) #0: censored ; 1: event;

# Combined longitudinal and survival data,
# measures after event/censor are deleted
subject <- rep(1:N, each = T)
treat <- rep(treat.pts, each = T)
t <- rep(time, N)
obs <- length(t)  # obs: total number of observations across subjects
zero <- rep(0, 2)
month <- rep(c(0,1,3,9,15), N)

all <- cbind(Y.conti, Y.ordi, subject, treat, t, month)
unbalance.all <- NULL
for (i in 1:obs)
{
  #keep the measures before event/censor
  if (all[i, 7] < tee[subject[i]])
  {unbalance.all <- rbind(unbalance.all, all[i,])}
  #delete measures after/at the same time of event/censor
  else {unbalance.all <- unbalance.all}
}

obs <- nrow(unbalance.all)
subject <- unbalance.all[,4]
treat <- unbalance.all[,5]
t <- unbalance.all[,6]
Y.conti <- unbalance.all[,1]
Y.ordi <- unbalance.all[,2:3]

# output BUGS data
zeros <-  numeric(N)
n <- n.ordi
data <- list("N", "n", "subject", "obs", "K.ordi",
             "Y.conti", "Y.ordi", "treat", "zero",
             "t", "treat.pts", "tee","event","zeros")

## generate data for BUGS
outname <- myloop
```

```
  outputname <- paste("BUGSdata/", "data", outname, ".txt", sep="")
  bugs.data(data, data.file=outputname)

  ## transfer BUGS data to JAGS format
  bugs.data.path <- paste("./", "BUGSdata/", "data", outname, ".txt", sep="")
  jags.data.path <- paste("./", "JAGSdata/", "data", outname, ".txt", sep="")
  bugs2jags(bugs.data.path, jags.data.path)
}
```

JAGS code for fitting $\mathrm{JM}_{\mathrm{ST}}$ model

```
model
{
  for (i in 1:obs)  # obs: number of total observations
  {
    w1[i] ~ dgamma(nu,nu)
    Y.conti[i] ~ dnorm(mu.conti[i], taudy[i]) # continuous outcome
    taudy[i]<-w1[i]*tau.conti
    # two ordinal outcomes
    Y.ordi.1[i] ~ dcat(prob.y.1[i, 1:n.1])
    Y.ordi.2[i] ~ dcat(prob.y.2[i, 1:n.2])
  }

  # construct ST distribution on the mean of the continuous outcome
  for (i in 1:obs)
{
  mu.conti[i] <- a.conti + b.conti * theta[i] + delta.sn*(w.sn[subject[i]])
}

    # Construct the probability vector for the ordinal variables
    for (i in 1:obs)
{
for (l in 1:(n.1-1)) { logit(psi.1[i, l]) <- a.ordi.1[l] - b.ordi.1*theta[i] }
prob.y.1[i, 1] <- psi.1[i, 1]
for (l in 2:(n.1-1)) { prob.y.1[i, l] <- psi.1[i, l] - psi.1[i, l-1] }
prob.y.1[i, n.1] <- 1 - psi.1[i,(n.1-1)]

for (l in 1:(n.2-1)) { logit(psi.2[i, l]) <- a.ordi.2[l] - b.ordi.2*theta[i] }
prob.y.2[i, 1] <- psi.2[i, 1]
for (l in 2:(n.2-1)) { prob.y.2[i, l] <- psi.2[i, l] - psi.2[i, l-1] }
prob.y.2[i, n.2] <- 1 - psi.2[i,(n.2-1)]
}
```

```
  for (i in 1:N)
    {
    # construct random effects
    u[i, 1:2] ~ dmnorm(zero[], precision[,])
    # construct variable for the skewness parameter
    w.sn[i]~dnorm(0,1)I(0,)
}

    # construct the variance-covariance matrix for random effects
    precision[1:2,1:2]<-inverse(sigma[,])
    sigma[1,1]<-1
    sigma[1,2]<-rho*sig1
    sigma[2,1]<-sigma[1,2]
    sigma[2,2]<-sig1*sig1

  # construct theta, the latent variable of subject i at time j
  for (i in 1:obs)
  {
     theta[i] <- u[subject[i], 1] + (beta[1] + beta[2]*treat[i]
                 + u[subject[i], 2])*t[i]
  }

  # construct survival part
  for (i in 1:N)
  {
    # use zero-trick to specify the likelihood
    phi[i] <- -lL[i]
    zeros[i] ~ dpois(phi[i])

    # k is the number of time interval for baseline step function
    for (k in 1:3) {
      h0[i,k] <- inprod(g[k],I0[i,k])
      gt[i,k] <- inprod(g[k],dt1[i,k])
    }

## take log of the survival function
    lh[i] <- gam*treat.pts[i] + omega2*u[i, 1] + omega3*u[i,2] + log(sum(h0[i,]))
    lS[i] <- -(exp(gam*treat.pts[i] + omega2*u[i, 1] + omega3*u[i,2])*sum(gt[i,]))
    # event=1 for event; 0 for censored
    lL[i] <- event[i]*lh[i] + lS[i] -log(1.0E+08)
  }
```

```
  # prior for g
  for (k in 1:3)
  {
    g[k] ~dunif(0,20)
  }
  # prior for parameters gam, omega2, and omega3
  gam ~ dnorm(0, 0.01)
  omega2 ~ dnorm(0, 0.01)
  omega3 ~ dnorm(0, 0.01)

  # prior for regression coefficients
  for (i in 1:2)
  {
      beta[i] ~ dnorm(0, 0.01)
  }

  # specify prior distributions
  rho ~ dunif(-1, 1)
  sig1 ~  dgamma(0.01, 0.01)

  # prior for continuous variable's parameters
  b.conti ~ dgamma(0.001,0.001)
  a.conti ~ dnorm(0, 0.0005)
  tau.conti ~ dgamma(0.001,0.001)
  sd.conti <- 1/sqrt(tau.conti)
  b.ordi.1 ~ dgamma(0.001,0.001)
  b.ordi.2 ~ dgamma(0.001,0.001)

  a.ordi.1[1] ~ dnorm(0,0.001)
  for (l in 2:(n.1-1)) { a.ordi.1[l] <- a.ordi.1[l-1] + delta.1[l-1] }
  for (i in 1:(n.1-2)) {delta.1[i] ~ dnorm(0,0.01)I(0,) }
  a.ordi.2[1] ~ dnorm(0,0.001)
  for (l in 2:(n.2-1)) { a.ordi.2[l] <- a.ordi.2[l-1] + delta.2[l-1] }
  for (i in 1:(n.2-2)) {delta.2[i] ~ dnorm(0,0.01)I(0,) }

  # prior distribution for df and skewness parameters
  nu~dgamma(0.001,0.001)
  delta.sn~dnorm(0,0.001)I(0,)
}
```

4 CONCLUSION AND FUTURE RESEARCH

In this research we first provided a new statistical technique for the analysis of multivariate longitudinal data when the continuous outcomes and random effects have heavy tails and outliers. We performed an extensive simulation study to illustrate how our proposed methods handle the heavy tails and outlier problems and provide robust parameter estimations. We compared the simulation results and picked the Indep-CN-MLIRT model, which maintains the flexibility of specifying different tuning parameters for continuous outcome and random effects, as our best model in handling outliers and providing accurate parameter estimates. We applied our method to the DATATOP study on Parkinson's disease and demonstrated how the implementation of the NI distribution to the MLIRT model obtains robust inference.

Furthermore, we provided a model framework for multivariate longitudinal data that accounts for the three data features, outliers, skewness and dependent censoring by using joint MLIRT model with SNI distribution assumption. We performed simulation studies to demonstrate how our proposed method provide robust parameter estimations and explored the effect of the three features (outliers, skewness and dependent censoring) to our model estimations. In addition, the simulation results showed that when there were no outliers or skewness, the proposed JM_{ST} model still produced satisfactory results. We applied our method to the motivating DATATOP study and confirmed the positive skewness and heavy tail nature of the dataset by providing visualization on the skewness parameter δ and degree of freedom ν. We demonstrated that the JM_{ST} model was

the best fitting model to the DATATOP dataset based on the model selection criterion (LPML,DIC,EAIC and EBIC). The BF results indicated decisive evidences in favor of our proposed JM_{ST} over all the other models (JM_{SN}, JM_{T}, JM_{N}, RM_{ST}, and RM_{N}).

Overall, The Bayesian hierarchical implementation of NI and SNI distributions to the MLIRT model is fairly straightforward. The easy access of publicly available software such as `WinBUGS`, `OpenBUGS` and `JAGS` provide a practical and feasible platform for practitioner and researchers to perform analysis using our proposed method.

For our future research, the underlying linear disease progression assumption on θ_{ij} may be relaxed by add the quadratic or high order term of time t, e.g., $\theta_{ij} = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0 + u_{i0} + (\boldsymbol{X}_{i1}\boldsymbol{\beta}_1 + u_{i1})t_{ij} + (\boldsymbol{X}_{i2}\boldsymbol{\beta}_2 + u_{i1})t_{ij}^2$, where $\boldsymbol{X}_{i0}$, $\boldsymbol{X}_{i1}$, and $\boldsymbol{X}_{i2}$ are the covariate vectors that can share part or all the covariates. In addition, non-parametric smoothing methods such as B-spline or P-spline may also be used to further relax the assumptions on θ_{ij}. Moreover, predictive inference using the MLIRT framework maybe another field we may explore. The MLIRT model will allow us to make predictions not only to the overall disease condition but also to the measures on specific outcomes. Furthermore, in our simulation study, 5% outliers on were considered and discussed. It would be interesting to study the robustness of NI or SNI distributions under various degree of outliers and skewness.

References

Arellano-Valle, R., Bolfarine, H., and Lachos, V. (2007). Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics*, 34(6):663–682.

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(3):579–602.

Azzalini, A. and 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 (Statistical Methodology)*, 65(2):367–389.

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(6):844–865.

Bandyopadhyay, D., Lachos, V. H., Abanto-Valle, C. A., and 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., and 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., and Chatterjee, A. (2011). A review of multivariate longitudinal data analysis. *Statistical Methods in Medical Research*, 20(4):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(4):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.

Chen, R. (2012). *Bayesian inference on mixed-effects models with skewed distributions for HIV longitudinal data*. PhD thesis, University of South Florida.

Cummings, J. L. (1992). Depression and Parkinson's disease: A review. *The American Journal of Psychiatry*, 149(4):443–454.

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. 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(12):1513–1521.

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(24):2498–2508.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685.

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*. CRC press.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman&Hall, Boca Raton, FL.

Ghosh, P., Branco, M. D., and Chakraborty, H. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA. *Statistics in Medicine*, 26(6):1255–1267.

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(2):158–170.

Group, P. S. (2002). A controlled trial of rasagiline in early Parkinson disease: the TEMPO study. *Archives of Neurology*, 59(12):1937.

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*.

Henderson, R., Diggle, P., and 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., and 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., and 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.

Huang, Y. and Dagne, G. (2012). Bayesian semiparametric nonlinear mixed-effects joint models for data with skewness, missing responses, and measurement errors in covariates. *Biometrics*, 68(3):943–953.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38(1):79–93.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Lachin, J. M. (2000). Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials*, 21(3):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(4):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 & Data Analysis*, 64:237–252.

Lachos, V. H., Dey, D. K., and 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., and Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, 20(1):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(2):175–198.

Lavine, M. and Schervish, M. J. (1999). Bayes factors: what they are and what they are not. *The American Statistician*, 53(2):119–122.

Lawless, J. and 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. and Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the LaplaceMetropolis estimator. *Journal of the American Statistical Association*, 92(438):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(2):484–495.

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

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, 91(435):1219–1227.

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

Luo, S. (2014). A bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine*, 33(4):580–594.

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*.

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(22):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(2):335–336.

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(6):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(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.

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(5):573–590.

Sahu, S. K., Dey, D. K., and 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*. Springer, New York, New York.

Song, X., Davidian, M., and 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., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Wang, C., Douglas, J., and Anderson, S. (2002). Item response models for joint analysis of quality of life and survival. *Statistics in Medicine*, 21(1):129–142.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339.

Zeng, D. and 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.