Research Article

Statistics in Medicine

Received 16 December 2013, Accepted 21 May 2014 Published online 17 June 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6228

Bayesian hierarchical model for multiple repeated measures and survival data: an application to Parkinson's disease

Sheng Luo[a*†] **and Jue Wang**[b]

Multilevel item response theory models have been increasingly used to analyze the multivariate longitudinal data of mixed types (e.g., continuous and categorical) in clinical studies. To address the possible correlation between multivariate longitudinal measures and time to terminal events (e.g., death and dropout), joint models that consist of a multilevel item response theory submodel and a survival submodel have been previously developed. However, in multisite studies, multiple patients are recruited and treated by the same clinical site. There can be a significant site correlation because of common environmental and socioeconomic status, and similar quality of care within site. In this article, we develop and study several hierarchical joint models with the hazard of terminal events dependent on shared random effects from various levels. We conduct extensive simulation study to evaluate the performance of various models under different scenarios. The proposed hierarchical joint models are applied to the motivating deprenyl and tocopherol antioxidative therapy of Parkinsonism study to investigate the effect of tocopherol in slowing Parkinson's disease progression. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: clinical trial; item-response theory; failure time; latent variable; MCMC; nested random effects

1. Introduction

Parkinson's disease (PD) is a chronic progressive neurodegenerative disorder manifested clinically by tremor, rigidity, slow movement, and impaired balance [1]. In the USA alone, PD affects about 1% of people older than 60 years [2]. Morbidity and mortality associated with PD burden result in an estimated six billion dollars in healthcare cost in the USA annually. Because the risk of PD increases with age, the financial and public health burden of PD is expected to increase as the population ages. The impairment caused by PD is multidimensional and progressive. Its multidimensional nature precludes a single outcome to measure disease progression. Many clinical trials of PD often collect multiple longitudinal outcomes of mixed types (categorical and continuous) to assess the treatment effects on overall improvement.

To analyze such multivariate longitudinal data, multilevel item response theory (MLIRT) models have been increasingly used [3–9]. According to MLIRT, 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 first level measurement model. Then, 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 [10–14]. Advantages of the MLIRT model include the following: (i) it avoids adjustment of type I error rates and provides improved treatment effect estimate [15]; (ii) it separates the measurement-specific parameters and subject-specific predictors so that both may be understood and studied separately; (iii) it accounts for all sources of correlation within the same subject via the random effects and it utilizes the full longitudinal information of all subjects; and (iv) each subject's

[a]*Division of Biostatistics, The University of Texas Health Science Center, Houston, 1200 Pressler St, Houston, TX 77030, U.S.A.*
[b]*Division of Biostatistics, The University of Texas Health Science Center, Houston, TX, U.S.A.*
[*]*Correspondence to: Sheng Luo, Division of Biostatistics, The University of Texas Health Science Center, Houston, 1200 Pressler St, Houston, TX 77030, U.S.A.*
[†]*E-mail: sheng.t.luo@uth.tmc.edu*

disease progression can be readily monitored using the latent disease severity. The MLIRT model has been increasingly used in studying many diseases such as PD disability [6, 16], Alzheimer's disease [17], Huntington's disease [18], and dementia [19].

During the follow-up process of PD studies, some patients may have terminal events such as death, dropout, and withdrawal of consent. The terminal events are often correlated with the multivariate longitudinal measurements. The presence of such a dependent failure time is often termed 'informative censoring', ignoring which could lead to biased estimates [20]. To address this issue, joint analysis of survival with repeated measures has been increasingly common [20–24]. In the IRT modeling framework, Wang *et al.* [4] proposed a joint model to analyze multiple-item ordinal quality of life data in the presence of death. He and Luo [8] developed a joint model for multiple longitudinal outcomes of mixed types, subject to outcome-dependent terminal events. Luo [9] relaxed the PH assumption in the work of He and Luo [8] and developed a joint modeling framework replacing the PH model by various parametric accelerated failure time models.

However, most joint models developed in the literature have focused on the subject level. In many multisite PD studies such as deprenyl and tocopherol antioxidative therapy of Parkinsonism (DATATOP) study [25], ELLDOPA study [1], and LS-1 study [26], multiple patients are recruited and treated by the same clinical site. There can be a significant site correlation because the patients in the same clinical site may share common environmental (e.g., medical environment related variables) and socioeconomic status and have similar quality of care. The existing MLIRT models ignore site correlation and potentially increase the variability of parameters, resulting in considerable bias in parameter estimates [27–29].

In this article, we will develop a hierarchical joint modeling framework in which a hierarchical MLIRT model is used for the multilevel multivariate longitudinal outcomes, and a Cox PH model is used for the dependent terminal event. The two models are linked via multilevel random effects. The rest of the article proceeds as follows. In Section 2, we describe a motivating clinical trial, the data structure, and the dependent terminal event. Section 3 discusses the joint random effects model, Bayesian inference, and Bayesian model selection criteria. Section 4 provides a simulation study to assess the performance of the proposed joint model. In Section 5, we apply the proposed model to the motivating clinical trial dataset. Section 6 gives some concluding remarks and discussions. To facilitate easy reading and implementation of the proposed models, the codes have been posted at the web supplement.

2. A motivating clinical trial

The methodological development in this article is motivated by the DATATOP study, a double-blind, placebo-controlled multicenter clinical trial to determine if deprenyl and/or tocopherol administered to patients with early PD will slow the progression of PD. A total of 800 patients were randomly assigned in a 2×2 factorial design to receive double-placebo, active tocopherol alone, active deprenyl alone, and both active tocopherol and deprenyl. The patients who did not receive tocopherol (double-placebo and active deprenyl alone groups, 401 patients) are combined and referred to as placebo group. The patients who received tocopherol (active tocopherol alone and both active tocopherol and deprenyl groups, 399 patients) are combined and referred to as treatment group. Please refer to the work of Shoulson *et al.* [25] for details of the DATATOP trial.

The multiple outcomes collected include Unified PD Rating Scale (UPDRS) total score, Schwab and England activities of daily living (SEADL), modified Hoehn and Yahr (HY) scale, measured at 10 visits (baseline, months 1, and every 3 months starting from month 3 to month 24). UPDRS is the sum of 44 questions each measured on a 5-point scale (0–4), and it is approximated by a continuous variable with integer value from 0 (not affected) to 176 (most severely affected). SEADL is a measurement of activities of daily living [30], and it is an ordinal variable with integer values from 0 to 100 incrementing by 5, with larger values reflecting better clinical outcomes. HY is a scale describing how the symptoms of PD progresses. It is an ordinal variable with possible values at 1, 1.5, 2, 2.5, 3, 4, and 5, with higher values being clinically worse outcome. We have recoded SEADL variable so that higher values in all outcomes correspond to worse clinical conditions. In addition, we have combined some categories with zero or small counts so that SEADL and HY have eight and five categories, respectively.

Before the end of the study, some patients (193 and 183 patients in the placebo and treatment groups, respectively) reached the primary study endpoint, for example, in the judgment of the blinded investigator, the patients reached a level of functional disability sufficient to initiate levodopa therapy, which can ameliorate the disease symptoms. We hence define time to levodopa therapy as the terminal event. To

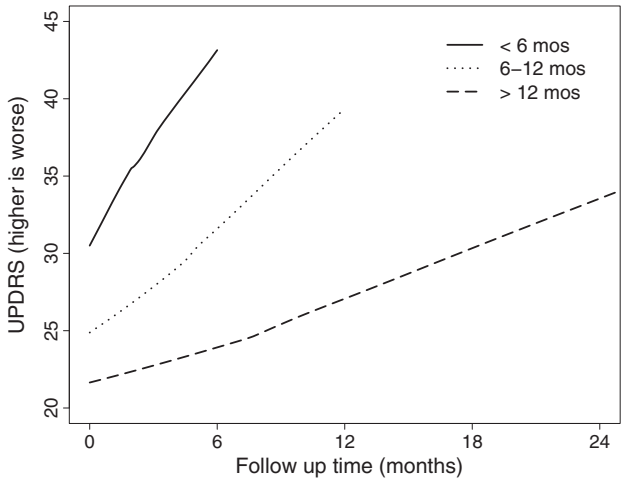

Figure 1. Mean UPDRS values over time.

visualize the correlation between the terminal event and the outcome measurements, Figure 1 displays mean UPDRS values over time for DATATOP patients with follow-up time less than 6 months (solid line), 6–12 months (dotted line), and more than 12 months (dashed line). It indicates that patients with shorter follow-up had higher UPDRS values (more severe PD).

Patients were recruited and treated in 28 research centers in the USA and Canada. There can be a significant clustering effect, which may represent homogeneity of quality of care in the same research center. In addition, patients treated in the same research center may share a common environmental and socioeconomic status. To this end, we will develop a hierarchical joint modeling framework of multivariate longitudinal outcomes and survival data that accounts for four levels of nesting, that is, multiple outcomes (level 1) are nested within visits (level 2), which are nested within patients (level 3) clustered within centers (level 4).

3. Model and estimation

3.1. The multilevel item response model

Let y_{hijk} (binary, ordinal, and continuous) be the observed outcome k ($k = 1, \ldots, K$) at visit j ($j = 1, \ldots, J_i$, where $j = 1$ is baseline and J_i is the number of visits of patient i) from patient i ($i = 1, \ldots, I_h$, where I_h is the number of patients in center h) within center h ($h = 1, \ldots, H$). Let $\boldsymbol{y}_{hij} = (y_{hij1}, \ldots, y_{hijK})'$ and let $\boldsymbol{y}_{hi} = (\boldsymbol{y}'_{hi1}, \ldots, \boldsymbol{y}'_{hiJ_i})'$. All outcomes are coded so that larger observation values correspond to worse clinical conditions. Let θ_{hij} be a univariate latent variable measuring disease severity of patient i from center h at visit j, with a higher value denoting more severe status. In the first level measurement model, we model the binary outcomes, the cumulative probabilities of ordinal outcomes, and the continuous outcomes as functions of θ_{hij} and outcome-specific parameters.

$$\text{logit}\left\{p(y_{hijk} = 1|\theta_{hij})\right\} = a_k + b_k\theta_{hij}, \tag{1}$$

$$\text{logit}\left\{p(y_{hijk} \leqslant l|\theta_{hij})\right\} = a_{kl} - b_k\theta_{hij}, \text{ with } l = 1, 2, \ldots, n_k - 1, \tag{2}$$

$$y_{hijk} = a_k + b_k\theta_{hij} + \epsilon_{hijk}, \tag{3}$$

where $\epsilon_{hijk} \sim N\left(0, \sigma_k^2\right)$, a_k is the outcome-specific 'difficulty' parameter and b_k (always positive) is the outcome-specific 'discriminating' parameter representing the discrimination of outcome k, that is, the degree to which outcome k discriminates between patients with different latent disease severity θ_{hij}. In model (2), the kth ordinal outcome has n_k categories and $n_k - 1$ thresholds $a_{k1}, \ldots, a_{kl}, \ldots, a_{kn_k-1}$ with the order constraint $a_{k1} < \ldots < a_{kl} < \ldots < a_{kn_k-1}$. The probability that patient i is in category l on outcome k at visit j is $p(Y_{hijk} = l|\theta_{hij}) = p(Y_{hijk} \leqslant l|\theta_{hij}) - p(Y_{hijk} \leqslant l - 1|\theta_{hij})$.

In the second-level structural multilevel model, the latent disease severity θ_{hij} is regressed on covariates of interest, visit time, and random effects.

$$\theta_{hij} = \boldsymbol{X}_{hi0}\boldsymbol{\beta}_0 + u_{hi0} + v_h + \left(\boldsymbol{X}_{hi1}\boldsymbol{\beta}_1 + u_{hi1}\right) t_{hij}, \tag{4}$$

where $\boldsymbol{X}_{hi0}$ and $\boldsymbol{X}_{hi1}$ are vectors of covariates associated with disease severity and progression rate, respectively, $\boldsymbol{X}_{hi0}$ may or may not be identical to $\boldsymbol{X}_{hi1}$, t_{hij} is visit time variable with $t_{hi1} = 0$ for baseline, random intercept u_{hi0} and random slope u_{hi1} determine the subject-specific disease severity and progression rate, respectively, and the center-specific random effects v_h follows $N(0, \sigma_v^2)$. It is of note that IRT models are over-parameterized. To make the model identifiable, we set $Var(u_{hi0}) = 1$ and set $a_{k1} = 0$ in one ordinal outcome. The random effects vector $\boldsymbol{u}_{hi} = (u_{hi0}, u_{hi1})'$ follows $N_2(0, \boldsymbol{\Sigma})$, with covariance matrix $\boldsymbol{\Sigma}$ being denoted by $\{(1, \rho\sigma_u), (\rho\sigma_u, \sigma_u^2)\}$ with σ_u^2 being the variance of u_{hi1} and ρ being the correlation coefficient. The random effects vector $\boldsymbol{u}_{hi}$ accounts for three sources of correlation within the same patient, that is, longitudinal (same outcome at different visits), inter-source (different outcomes at the same visit), and cross correlation (different outcomes at different visits). The random effects v_h accounts for the within-center correlation. Under the conditional independence assumption (conditional on the random effects $\boldsymbol{u}_{hi}$ and v_h, all measurements from each patient are independent) [31], the conditional likelihood of the multiple longitudinal outcomes for patient i is

$$L_y\left(\boldsymbol{\Phi}_y | \boldsymbol{y}_{hi}, \boldsymbol{u}_{hi}, v_h\right) = \prod_{j=1}^{J_i}\prod_{k=1}^{K} p\left(y_{hijk} | \boldsymbol{u}_{hi}, v_h\right), \tag{5}$$

where $p(y_{hijk}|\boldsymbol{u}_i)$ is the conditional density function of y_{hijk} obtained from models (1) to (4) and $\boldsymbol{\Phi}_y$ is the parameter vector. For notation convenience, we let the difficulty parameter vector be $\boldsymbol{a} = (\boldsymbol{a}_1', \ldots, \boldsymbol{a}_k', \ldots, \boldsymbol{a}_K')'$, with $\boldsymbol{a}_k$ being numeric for binary and continuous outcomes and $\boldsymbol{a}_k = (a_{k1}, \ldots, a_{kn_k-1})'$ for ordinal outcomes. Let the discrimination vector be $\boldsymbol{b} = (b_1, \ldots, b_K)'$, and let the regression vector be $\boldsymbol{\beta} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1')'$. So $\boldsymbol{\Phi}_y = (\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\beta}', \rho, \sigma_u, \sigma_v, \sigma_k)'$.

Let t_{hi} denote the time to terminal event for patient i within center h, δ_{hi} (1 if the terminal event is observed and 0 if not needed) denote the censoring indicator for t_{hi}, and $\boldsymbol{X}_{hi}$ denote vector of possible risk factors. Vector $\boldsymbol{X}_{hi}$ can share part of or all covariates in $\boldsymbol{X}_{hi0}$ and $\boldsymbol{X}_{hi1}$. The hazard of having a terminal event at time t_{hi} is

$$\lambda(t_{hi}) = \lambda_0(t_{hi}) \exp\left(\boldsymbol{X}_{hi}\boldsymbol{\gamma} + \nu_0 u_{hi0} + \nu_1 u_{hi1} + \nu_2 v_h\right), \tag{6}$$

where $\boldsymbol{\gamma}$, $\boldsymbol{\nu} = (\nu_0, \nu_1, \nu_2)'$ are unknown parameters and $\lambda_0(t_{hi})$ is the baseline hazard function. The associated survival function is $S(t_{hi}) = \exp[-\int_0^{t_{hi}} \lambda(s)ds]$. The likelihood of event outcome t_{hi} and δ_{hi} for patient i is

$$L_s\left(\boldsymbol{\Phi}_s | t_{hi}, \delta_{hi}, \boldsymbol{u}_{hi}, v_h\right) = \lambda(t_{hi})^{\delta_{hi}} S(t_{hi}), \tag{7}$$

where the parameter vector $\boldsymbol{\Phi}_s = (\boldsymbol{\gamma}', \boldsymbol{\nu}')'$. Conditional on the random effects $\boldsymbol{u}_{hi}$ and v_h, $\boldsymbol{y}_{hi}$ is assumed to be independent of t_{hi}. The full likelihood of the joint model for patient i within center h is

$$L\left(\boldsymbol{\Phi} | \boldsymbol{y}_{hi}, t_{hi}, \delta_{hi}, \boldsymbol{u}_{hi}, v_h\right) = L_y\left(\boldsymbol{\Phi}_y | \boldsymbol{y}_{hi}, \boldsymbol{u}_{hi}, v_h\right) L_s\left(\boldsymbol{\Phi}_s | t_{hi}, \delta_{hi}, \boldsymbol{u}_{hi}, v_h\right) p(\boldsymbol{u}_{hi}) p(v_h), \tag{8}$$

where $p(\boldsymbol{u}_{hi})$ and $p(v_h)$ are the density functions of $\boldsymbol{u}_{hi}$ and v_h, the parameter vector $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_y', \boldsymbol{\Phi}_s')'$. The 'cross-equation correlation' between models (4) and (6) is introduced by the center level random effects v_h and subject level random effects vector $\boldsymbol{u}_{hi}$. Specifically, positive parameters ν_0 and ν_1 indicates that the patients with worse disease severity (higher u_{hi0}) and faster disease progression rate (higher u_{hi1}) tend to need levodopa therapy earlier. Similarly, positive parameter ν_2 suggests that the patients within the center with more severe PD patients tend to have higher hazard of needing levodopa therapy. We thereafter refer to the propose joint modeling framework (8) as 'full model'.

Moreover, we consider two special cases of the full model. The first case (refer to as 'reduced model A') is that the hazard of survival event only depends on the subject level random effects vector $\boldsymbol{u}_{hi}$.

$$\begin{aligned} \theta_{hij} &= \boldsymbol{X}_{hi0}\boldsymbol{\beta}_0 + u_{hi0} + v_h + \left(\boldsymbol{X}_{hi1}\boldsymbol{\beta}_1 + u_{hi1}\right) t_{hij} \\ \lambda(t_{hi}) &= \lambda_0(t_{hi}) \exp\left(\boldsymbol{X}_{hi}\boldsymbol{\gamma} + \nu_0 u_{hi0} + \nu_1 u_{hi1}\right). \end{aligned} \tag{9}$$

Another special case (refer to as 'reduced model B') is that the hazard of survival event only depends on the center-level random effects v_h.

$$\begin{aligned}\theta_{hij} &= \boldsymbol{X}_{hi0}\boldsymbol{\beta}_0 + u_{hi0} + v_h + \left(\boldsymbol{X}_{hi1}\boldsymbol{\beta}_1 + u_{hi1}\right) t_{hij} \\ \lambda(t_{hi}) &= \lambda_0(t_{hi}) \exp\left(\boldsymbol{X}_{hi}\boldsymbol{\gamma} + \nu_2 v_h\right).\end{aligned} \tag{10}$$

3.2. Bayesian inference

The full likelihood formulation in model (8) involves unspecified baseline hazard function $\lambda_0(\cdot)$. We adopt a piecewise constant function to approximate $\lambda_0(\cdot)$. It has been shown that survival models using a piecewise constant baseline hazard function yield good estimators for both fixed effects and frailty [32, 33], although fixed cut points need to be specified *a priori*. It is more flexible than the *a priori* choices of baseline hazard distribution (e.g., Weibull distribution), and it retains enough model structure [34]. Specifically, we divide the follow-up period into m intervals by every $1/m$th quantiles denoted by $\tau_1, \tau_2, \ldots, \tau_m$ and $\tau_0 = 0$ or the smallest failure time. The baseline hazard vector is $\boldsymbol{g} = (g_1, g_2, \ldots, g_m)$, and the piecewise constant baseline hazard function is defined as $\lambda_0(t) = \sum_{l=1}^{m} g_l I_l(t)$, with indicator function $I_l(t) = 1$ if $\tau_{l-1} \leqslant t < \tau_l$ and 0 otherwise.

To infer the unknown parameter vector $\boldsymbol{\Phi}$, we use Bayesian inference based on MCMC posterior simulations. We use vague priors on all elements in the parameter vector $\boldsymbol{\Phi}$. Specifically, the prior distributions of all elements in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, ν_0, ν_1, and ν_2 are $N(0, 100)$. We use the prior distribution $b_k \sim \text{Uniform}(0, 30)$, $k = 1, \ldots, K$, to ensure positivity. The prior distribution for the difficulty parameter a_k of the continuous outcomes is $a_k \sim N(0, 100)$. To obtain the prior distributions for the threshold parameters of ordinal outcome k, we let $a_{k1} \sim N(0, 100)$, and $a_{kl} = a_{k,l-1} + \delta_l$ for $l = 2, n_k - 1$, with $\delta_l \sim N(0, 100)I(0,)$, that is, normal distribution left censored at 0. We use the prior distribution $\rho \sim \text{Uniform}[-1, 1]$, and $\sigma_k, \sigma_u, \sigma_v \sim \text{Gamma}(0.01, 0.01)$.

The model fitting is performed in `OpenBUGS` [35] (`OpenBUGS` version 3.2.2) by specifying the full likelihood function and the prior distributions of all unknown parameters. We use the history plots available in `OpenBUGS` and view the absence of apparent trend in the plot as evidence of convergence. In addition, we use the Gelman–Rubin diagnostic to ensure the scale reduction $\widehat{R}$ of all parameters are smaller than 1.1 [36].

3.3. Bayesian model selection criteria

There are a wide variety of model selection criteria in Bayesian inference. We adopt a model selection approach using the deviance information criterion (DIC) proposed by Spiegelhalter *et al.* [37]. The DIC provides an assessment of model fitting and a penalty for model complexity. The deviance statistic 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 function for the observed data $\boldsymbol{y}$ given the parameter vector $\boldsymbol{\Phi}$ and $h(\boldsymbol{y})$ denotes a standardizing function of the data alone that has no impact on model selection [38]. The DIC is defined as $\text{DIC} = 2\overline{D} - D(\overline{\boldsymbol{\Phi}}) = \overline{D} + p_D$, where $\overline{D} = E_{\boldsymbol{\Phi}|\boldsymbol{y}}[D]$ is the posterior mean of the deviance, $D(\overline{\boldsymbol{\Phi}}) = D(E_{\boldsymbol{\Phi}|\boldsymbol{y}}[\boldsymbol{\Phi}])$ is the deviance evaluated at the posterior mean $\overline{\boldsymbol{\Phi}}$ of the parameter vector, and $p_D = \overline{D} - D(\overline{\boldsymbol{\Phi}})$ is the effective number of parameters. A smaller value of DIC indicates a better-fitting model. In addition, we use the expected AIC (EAIC) and the expected Bayesian (or Schwarz) information criterion (EBIC) as model selection tools [38]. The EAIC and EBIC can be estimated as $\text{EAIC} = \overline{D} + 2p$ and $\text{EBIC} = \overline{D} + p\log(N)$, where p is the number of elements in the parameter vector $\boldsymbol{\Phi}$. Smaller values of EAIC and EBIC indicate better predictive ability of the model.

Moreover, Bayes factors (BF) is a Bayesian alternative to p-values for testing hypotheses and for quantifying the degree to which observed data support or conflict with a hypothesis. Let two competing models be M_1 and M_2. The BF in favor of model M_1 over M_2 is defined as

$$\text{BF}_{12} = \text{BF}(M_1; M_2) = \frac{p(M_1|\boldsymbol{y})/p(M_2|\boldsymbol{y})}{p(M_1)/p(M_2)} = \frac{p(\boldsymbol{y}|M_1)}{p(\boldsymbol{y}|M_2)}, \tag{11}$$

where $p(M_i)$ is the prior probability of model M_i, $i = 1, 2$, $p(M_i|\boldsymbol{y})$ is the posterior probability of model M_i, and $p(\boldsymbol{y}|M_i)$ is the predictive probability of observing data $\boldsymbol{y}$ under model M_i. $p(\boldsymbol{y}|M_i) = \int f(\boldsymbol{y}|\boldsymbol{\Phi}_i, M_i)p(\boldsymbol{\Phi}_i|M_i)d\boldsymbol{\Phi}_i$, where $p(\boldsymbol{\Phi}_i|M_i)$ is the prior distribution for parameter vector $\boldsymbol{\Phi}_i$ under model M_i. When $\text{BF}_{12} > 150$, decisive evidence is shown in favor of model M_1; when BF_{12} is between 20 and 150, strong evidence is shown in favor of model M_1 [39]. To avoid the integral involved in computation

of BF, the Laplace–Metropolis estimator based on the normal distribution [40] is adopted to approximate the predictive probability. Specifically, $p(\boldsymbol{y}|M_i) \approx (2\pi)^{d_i/2}|\boldsymbol{\Sigma}_i|^{1/2} f(\boldsymbol{y}|\overline{\boldsymbol{\Phi}}_i, M_i) p(\overline{\boldsymbol{\Phi}}_i|M_i)$, where d_i is the number of the parameters in $\boldsymbol{\Phi}_i$, $\boldsymbol{\Sigma}_i$ is the posterior covariance matrix of $\boldsymbol{\Phi}_i$, $\overline{\boldsymbol{\Phi}}_i$ is the posterior mean of parameters, $p(\overline{\boldsymbol{\Phi}}_i|M_i)$ is the prior probability of parameters evaluated at $\overline{\boldsymbol{\Phi}}_i$, and $f(\boldsymbol{y}|\overline{\boldsymbol{\Phi}}_i, M_i)$ is the likelihood when parameters are at the posterior mean values.

4. Simulation studies

In this section, we conduct the simulation study under two settings to compare the performance of the proposed hierarchical joint models (the full model and two reduced models). In both settings, we generate 200 datasets with sample size $H = 25$ centers and $I_h = 32$ patients per center, with a total of 800 patients. The data structure is similar to the motivating dataset, and it has one continuous outcome and three ordinal outcomes with 7, 7, 10 categories, and five visits.

We generate data by the following model:

$$\begin{aligned} \theta_{hij} &= \beta_{01} x_{hi} + u_{hi0} + v_h + \left(\beta_{10} + \beta_{11} x_{hi} + u_{hi1}\right) t_{hij} \\ \lambda(t_{hi}) &= \lambda_0 \exp\left(\gamma x_{hi} + \nu_0 u_{hi0} + \nu_1 u_{hi1} + \nu_2 v_h\right). \end{aligned} \tag{12}$$

The covariate x_{hi} takes value 0 or 1 each with probability $1/2$ to mimic treatment assignment. The time vector $t_{hi} = (t_{hi1}, \ldots, t_{hi5})' = (0, 0.17, 0.5, 1.5, 2.5)'$. We set the coefficients to be $\boldsymbol{\beta} = (\beta_{01}, \beta_{10}, \beta_{11})' = (-1, 0.4, -0.5)'$, $\gamma = -0.7$. The parameters for the continuous outcome are set to be $a_1 = 3$, $b_1 = 1.5$, and $\sigma_1 = 1$. The parameters for the ordinal outcomes are set to be $\boldsymbol{a}_2 = (0, 1, 1.8, 2.6, 3.3, 4)$, $\boldsymbol{a}_3 = (-2.7, -0.6, 2, 2.8, 5, 6)$, $\boldsymbol{a}_4 = (-1, -0.1, 0.5, 1, 1.5, 2, 2.4, 2.8, 3.3)$, $b_2 = 2$, $b_3 = 0.4$, and $b_4 = 0.7$. Assume that center-specific random effects $v_h \sim N(0, \sigma_v^2)$ with $\sigma_v = 1$, subject-specific random effects vector $\boldsymbol{u}_{hi} = (u_{hi0}, u_{hi1})' \sim N_2(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \{(1, \rho\sigma_u), (\rho\sigma_u, \sigma_u^2)\}$ with $\rho = 0.4$ and $\sigma_u = 1.3$. The baseline hazard λ_0 is set to 0.07. The independent censoring time is sampled from uniform(10, 20).

In setting I, we assume $\boldsymbol{\nu} = (0.4, 1.0, 1.0)'$ so that the hazard of terminal event depends on both the subject-specific and center-specific random effects. Table I displays bias (the average of the posterior means minus the true values), standard error (SE; the square root of the average of the posterior variance), standard deviation (SD; the SD of the posterior means), and coverage probabilities (CP) of 95% equal-tail credible intervals of two reduced models and the full model (true model). The results suggest that the full model provides parameter estimates with very small biases, SE being close to SD, and the CP being close to the nominal level 0.95. In contrast, when all components in $\boldsymbol{\nu}$ are nonzero as in setting I, incorrectly constraining some of the model association to 0 lead to inaccurate results. For example, both reduced models A and B give biased estimates and poor CP for the regression vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, in additional to the variance-related parameters ρ, σ_u, and σ_v. The estimates of the difficulty parameter

Table I. Simulations results from reduced model A, reduced model B, and the full model for setting I (the terminal event dependent on both the subject-specific and center-specific random effects).

	Reduced model A				Reduced model B				Full model			
	Bias	SE	SD	CP	Bias	SE	SD	CP	Bias	SE	SD	CP
For longitudinal outcomes												
$\beta_{01} = -1.0$	0.037	0.078	0.082	0.920	0.007	0.079	0.082	0.935	0.009	0.079	0.084	0.925
$\beta_{10} = 0.4$	0.140	0.091	0.105	0.640	−0.501	0.079	0.093	0.000	−0.007	0.086	0.088	0.945
$\beta_{11} = -0.5$	−0.029	0.116	0.129	0.910	0.124	0.106	0.106	0.785	−0.001	0.112	0.121	0.925
$\rho = 0.4$	0.009	0.043	0.044	0.945	−0.092	0.052	0.050	0.550	−0.006	0.044	0.042	0.940
$\sigma_u = 1.3$	0.068	0.071	0.078	0.825	−0.172	0.055	0.054	0.135	0.001	0.063	0.061	0.960
$\sigma_v = 1.0$	−0.247	0.126	0.117	0.600	0.047	0.165	0.157	0.940	0.030	0.162	0.142	0.965
For survival												
$\gamma = -0.7$	0.055	0.153	0.150	0.945	0.225	0.098	0.122	0.395	0.011	0.150	0.159	0.930
$\nu_0 = 0.4$	0.110	0.094	0.114	0.730					−0.000	0.079	0.076	0.940
$\nu_1 = 1.0$	−0.032	0.082	0.093	0.905					0.001	0.083	0.076	0.965
$\nu_2 = 1.0$					−0.303	0.067	0.065	0.010	0.009	0.079	0.081	0.935

Table II. Simulations results from reduced model A and the full model for setting II (the terminal event only dependent on the subject-specific random effects).

	Reduced model A				Full model			
	Bias	SE	SD	CP	Bias	SE	SD	CP
For longitudinal outcomes								
$\beta_{01} = -1.0$	0.000	0.079	0.085	0.940	0.008	0.079	0.084	0.925
$\beta_{10} = 0.4$	0.003	0.086	0.090	0.930	−0.009	0.086	0.092	0.930
$\beta_{11} - 0.5$	−0.001	0.112	0.110	0.930	0.011	0.112	0.121	0.925
$\rho = 0.4$	−0.002	0.044	0.037	0.980	−0.000	0.044	0.038	0.975
$\sigma_u = 1.3$	0.003	0.064	0.065	0.940	−0.002	0.065	0.065	0.955
$\sigma_v = 1.0$	0.027	0.162	0.162	0.930	0.025	0.162	0.152	0.945
For survival								
$\gamma = -0.7$	−0.007	0.149	0.165	0.935	0.007	0.149	0.159	0.945
$\nu_0 = 0.4$	−0.005	0.077	0.074	0.950	−0.001	0.078	0.075	0.955
$\nu_1 = 1.0$	0.002	0.083	0.084	0.950	0.000	0.084	0.085	0.955
$\nu_2 = 0.0$					−0.009	0.065	0.068	0.940

vector $\boldsymbol{a}$, the discriminating vector $\boldsymbol{b}$, and σ_1 from both reduced models and the full model have small bias, SE close to SD, and CP close to 95% (please refer Table 1 in the Web supplement).

In setting II, we assume $\boldsymbol{\nu} = (0.4, 1.0, 0)'$ so that the hazard of terminal event only depends on the subject-specific but not the center-specific random effects. Table II displays the simulation results of setting II from reduced model A (true model) and the full model. The results suggest that both models generate comparable results, that is, the bias is negligible, SE is close to SD, and the credible interval CP are reasonably close to 95%. Under model overparameterization, the estimate of η_2 from the full model is correctly close to zero. The estimates of the difficulty parameter vector $\boldsymbol{a}$, the discrimination vector $\boldsymbol{b}$, and σ_1 from both reduced model A and the full model have small bias, SE close to SD, and CP close to 95% (please refer Table 2 in the Web supplement).

From the simulation study, we conclude that when the hazard of the terminal event depends on both the subject-specific and center-specific random effects, the full model provides satisfactory results while reduced models A and B with incorrect assumption on the dependence structure result in considerable biases and poor CP. On the other hand, when the hazard of the terminal event only depends on the subject-specific, but not the center-specific random effects, the overparameterized full model provides results comparable with the reduced model, which has correct assumption on the dependence structure.

5. Real data analysis

In this section, we apply the proposed hierarchical joint models and the Bayesian inference framework to the motivating DATATOP study. Specifically, we compare the full model, reduced models A and B, and the model without accounting for center information (referred to as 'no-center model'). For all the results in this section, we use two parallel MCMC chains with overdispersed initial values, and run each chain for 40,000 iterations. The first 30,000 iterations are discarded as burn-in, and the inference is based on the remaining 10,000 iterations from each chain. Good mixing properties of the chains for the model parameter are observed in the trace plots. The scale reduction $\widehat{R}$ of all parameters are smaller than 1.1. To specify the baseline hazard functions $\lambda_0(t_{hi})$, we divide the time to levodopa therapy into $M = 3$ intervals by every $1/M$th quantiles. We have also explored other selections of M and the results are very similar.

The PD patients in the DATATOP study were recruited and treated by a total of 28 research centers. The number of patients in each center varies from 10 (center 4) to 44 (center 19) with mean and SD being 24.9 and 9.8, respectively. There are 104 patients with missing center information, which we collapse into one 'artificial' center to obtain a total of 29 research centers. Table III displays the research center numbers and the number of patients in each center. We consider the treatment assignment variable x_{hi} (1 treatment, and 0 if placebo) as a single covariate so that $\boldsymbol{X}_{hi0} = x_{hi}$, $\boldsymbol{X}_{hi1} = (1, x_{hi})'$, and $\boldsymbol{\gamma} = x_{hi}$. Hence, model (4) is $\theta_{hij} = \beta_{01}x_{hi} + u_{hi0} + v_h + (\beta_{10} + \beta_{11}x_{hi} + u_{hi1})t_{hij}$, with visit times being transformed in 2-year unit as $t_{ij} = (0, 1, 3, 6, 9, 12, 15, 18, 21, 24)/24$. The Cox model for the time to levodopa therapy is

Table III. Research center numbers and the number of patients in each center.

Center #	1	2	4	5	6	7	8	9	10	11	12	14	15	16	17
Count	23	35	10	43	23	18	22	18	19	41	16	37	25	20	26
Center #	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Count	34	44	31	20	19	18	30	19	14	18	11	42	20	104	

Center 31 is the 'artificial' center with all patients without center information.

Table IV. Model comparison statistics for the DATATOP dataset.

	Dbar	EAIC	EBIC	DIC	BF
Reduced model A	**53876.9**	**53930.9**	**54057.4**	**55186.9**	Ref
Reduced model B	54576.9	54628.9	54750.7	55846.9	>> 150
Full model	53876.9	53932.9	54064.1	55196.9	38.9
no-center model	53976.9	54028.9	54150.7	55326.9	>> 150

Dbar, the posterior mean of the deviance; EAIC, expected AIC; EBIC, expected BIC; DIC, deviance information criterion; BF, Bayes factor in favor of reduced model A over other models. Boldface indicates the preferred model.

Table V. Results of fitting reduced model A, reduced model B, and the full model in the DATATOP dataset.

	Reduced model A				Reduced model B				Full model			
	Mean	SD	95% CI		Mean	SD	95% CI		Mean	SD	95% CI	
For longitudinal outcomes												
β_{01}	−0.021	0.067	−0.167	0.102	−0.027	0.068	−0.183	0.095	−0.027	0.056	−0.127	0.091
β_{10}	2.527	0.145	2.261	2.822	2.258	0.127	2.007	2.503	2.520	0.138	2.251	2.782
β_{11}	−0.129	0.167	−0.467	0.190	−0.085	0.164	−0.423	0.224	−0.109	0.150	−0.410	0.176
ρ	0.450	0.039	0.372	0.523	0.422	0.044	0.334	0.507	0.450	0.041	0.370	0.529
σ_u	2.190	0.108	1.987	2.413	1.995	0.102	1.800	2.205	2.185	0.112	1.969	2.400
σ_v	1.757	0.250	1.350	2.318	1.819	0.260	1.392	2.412	1.763	0.255	1.346	2.336
For survival												
γ	−0.029	0.148	−0.331	0.253	−0.044	0.102	−0.247	0.152	−0.024	0.141	−0.307	0.250
ν_0	0.349	0.070	0.211	0.489					0.358	0.071	0.211	0.492
ν_1	0.660	0.050	0.563	0.760					0.665	0.050	0.571	0.771
ν_2					0.097	0.108	−0.122	0.308	0.021	0.160	−0.293	0.338

$\lambda(t_{hi}) = \lambda_0(t_{hi})\exp(\gamma x_{hi}+\nu_0 u_{hi0}+\nu_1 u_{hi1}+\nu_2 v_h)$. Table IV compares all the hierarchical joint models using the model selection criteria discussed in Section 3.2. All of the proposed hierarchical joint models perform significantly better than the no-center model with smaller Dbar, EAIC, EBIC, and DIC values, suggesting that the hierarchical joint models accounting for the center information are more preferable than the no-center model. The BF in favor of reduced model A over reduced model B or the no-center model are much larger than 150, suggesting decisive evidence in favor of reduced model A. In comparison, the BF in favor of reduced model A over the full model is 38.9, suggesting strong evidence in favor of reduced model A [39]. Thus, reduced model A is selected as the final model because it has the best predictive ability with the smallest Dbar, EAIC, EBIC, and DIC values, and the BF.

Table V compares the posterior mean, SD, and 95% equal-tail credible intervals from the full model and reduced models A and B. It is observed that all models give different parameter estimates, although the same set of parameters are identified for significance by all models. For instance, reduced model A suggests that tocopherol is associated with 0.021 unit decrease ($\widehat{\beta}_{01}$, 95% CI: $[-0.167, 0.102]$) in disease severity, comparing with placebo, while it is associated with 0.027 (95% CI: $[-0.183, 0.095]$) and 0.027 (95% CI: $[-0.127, 0.091]$) unit decrease in disease severity in reduced model B and the full model, respectively. The placebo patients have significant disease progression at the rate of 2.527 units per 2 years

($\widehat{\beta}_{10}$, 95% CI: [2.261, 2.822]) in reduced model A, whereas the rates of progression are 2.258 (95% CI: [2.007, 2.503]) and 2.520 (95% CI: [2.251, 2.782]) units per 2 years in reduced model B and the full model, respectively. In comparison, reduced model A suggests that the tocopherol patients have disease progression rate of 2.398 units per 2 years ($\widehat{\beta}_{10} + \widehat{\beta}_{11}$, 95% CI: [1.794, 3.012]) with insignificant tocopherol treatment effect of slowing down the disease progression rate by −0.129 per two year ($\widehat{\beta}_{11}$, 95% CI: [−0.467, 0.190]). Similar conclusions can be made from reduced model B and the full model. Moreover, tocopherol has insignificant effects on the probability of having levodopa therapy, that is, the time to levodopa therapy for tocopherol patients is estimated to be 0.971 times (exp($\widehat{\gamma}$), 95% CI: [0.718, 1.288]) that of nontocopherol patients in reduced model A versus 0.957 times (95% CI: [0.781, 1.164]) in reduced model B and 0.976 times (95% CI: [0.746, 1.402]) in the full model.

Table V also suggests that patients with worse disease severity (higher u_{hi0}) and faster disease progression rate (higher u_{hi1}) tend to need levodopa therapy earlier, as indicated by the significance of the estimates of parameters ν_0 (0.349, 95% CI: [0.211, 0.489] in reduced model A and 0.358, 95% CI: [0.211, 0.492] in the full model) and ν_1 (0.660, 95% CI: [0.563, 0.760] in reduced model A and 0.665, 95% CI: [0.571, 0.771]). However, there is no significant association between disease severity and survival at the research center level, as indicated by the insignificance of $\widehat{\nu}_2$ (0.097, 95% CI: [−0.122, 0.308] in reduced model B and 0.021, 95% CI: [−0.293, 0.338] in the full model).

To provide parameter interpretation on clinically interpretable scales, we tabulate in Table VI the change from baseline to each follow-up visit for the outcome UPDRS, and the odds ratio of the cumulative probability at certain thresholds of the outcomes SEADL and HY, all estimated from reduced model A. At month 3, the placebo patients are expected to increase 3.035 (95% CI: [2.758, 3.339]) units in UPDRS, and the odds ratio of the cumulative probabilities at certain thresholds of SEADL and HY are expected to be 0.596 (95% CI: [0.563, 0.628]) and 0.678 (95% CI: [0.649, 0.705]), respectively, compared with baseline, whereas the treatment patients are expected to increase 2.880 (95% CI: [2.573, 3.161]) units in UPDRS, and the odds ratio of the cumulative probabilities at certain thresholds of SEADL and HY are expected to be 0.612 (95% CI: [0.582, 0.647]) and 0.692 (95% CI: [0.664, 0.722]), respectively, compared with baseline. At month 15, the placebo patients are expected to increase 15.177 (95% CI: [13.790, 16.697]) units in UPDRS, and the odds ratio of the cumulative probabilities at certain thresholds of SEADL and HY are expected to be 0.076 (95% CI: [0.056, 0.097]) and 0.144 (95% CI: [0.115, 0.174]), respectively, compared with baseline, whereas the treatment patients are expected to increase 14.400 (95% CI: [12.865, 15.804]) units in UPDRS, and the odds ratio of the cumulative probabilities at certain thresholds of SEADL and HY are expected to be 0.087 (95% CI: [0.067, 0.113]) and 0.159 (95% CI: [0.129, 0.196]), respectively, compared with baseline. At the end of the study at month 24, the placebo

Table VI. Change from baseline to each follow-up visit for the outcome UPDRS and the odds ratio (OR) of the cumulative probability at certain thresholds of the outcomes SEADL and HY, estimated from reduced model A.

	UPDRS		OR$\{p(\text{SEADL} \leqslant l)\}$		OR$\{p(\text{HY} \leqslant l)\}$	
	Placebo	Treatment	Placebo	Treatment	Placebo	Treatment
Month 3	$3.035_{0.150}$ [2.758, 3.339]	$2.880_{0.150}$ [2.573, 3.161]	$0.596_{0.017}$ [0.563, 0.628]	$0.612_{0.017}$ [0.582, 0.647]	$0.678_{0.015}$ [0.649, 0.705]	$0.692_{0.015}$ [0.664, 0.722]
Month 9	$9.106_{0.451}$ [8.274, 10.018]	$8.640_{0.449}$ [7.719, 9.482]	$0.213_{0.018}$ [0.178, 0.247]	$0.230_{0.019}$ [0.197, 0.271]	$0.312_{0.020}$ [0.273, 0.351]	$0.332_{0.021}$ [0.293, 0.376]
Month 15	$15.177_{0.752}$ [13.790, 16.697]	$14.400_{0.748}$ [12.865, 15.804]	$0.076_{0.011}$ [0.056, 0.097]	$0.087_{0.012}$ [0.067, 0.113]	$0.144_{0.015}$ [0.115, 0.174]	$0.159_{0.017}$ [0.129, 0.196]
Month 21	$21.248_{1.053}$ [19.307, 23.376]	$20.160_{1.047}$ [18.010, 22.126]	$0.027_{0.005}$ [0.018, 0.038]	$0.033_{0.006}$ [0.022, 0.047]	$0.067_{0.010}$ [0.049, 0.087]	$0.077_{0.011}$ [0.057, 0.102]
Month 24	$24.283_{1.203}$ [22.065, 26.716]	$23.040_{1.196}$ [20.583, 25.287]	$0.016_{0.004}$ [0.010, 0.024]	$0.020_{0.004}$ [0.013, 0.031]	$0.045_{0.008}$ [0.031, 0.061]	$0.053_{0.009}$ [0.038, 0.074]

The numbers in the subscript are the standard deviations (SD). The numbers within the square brackets are 95% equal-tailed credible intervals.

patients are expected to increase 24.283 (95% CI: [22.065, 26.716]) units in UPDRS, and the odds ratio of the cumulative probabilities at certain thresholds of SEADL and HY are expected to be 0.016 (95% CI: [0.010, 0.024]) and 0.045 (95% CI: [0.031, 0.061]), respectively, compared with baseline, whereas the treatment patients are expected to increase 23.040 (95% CI: [20.583, 25.287]) units in UPDRS, and the odds ratio of the cumulative probabilities at certain thresholds of SEADL and HY are expected to be 0.020 (95% CI: [0.013, 0.031]) and 0.053 (95% CI: [0.038, 0.074]), respectively, compared with baseline. Similar conclusions can be made at months 9 and 21.

Note that the SE of the random intercept (u_{hi1}) from reduced model A is 2.190 (95% CI: [1.987, 2.413]), and the estimate of the correlation coefficient ρ between two subject-specific random effects (u_{hi0} and u_{hi1}) is 0.450 (95% CI: [0.372, 0.523]). The significant positive correlation indicates that the patients with worse PD disease severity (higher u_{hi0}) tend to have faster disease progression rate (higher u_{hi1}) and vice versa. To gain further insight into u_{hi0}, u_{hi1}, and ρ, we plot in Web Figure 1 the rankings of patients' subject-specific disease severity (u_{hi0}, upper panel) and disease progression rate (u_{hi1}, lower panel). Each patient is ordered by his or her rank: patients at the bottom left corner show milder disease severity (lower rank) and slower disease progression rate (lower rank), whereas patients at the upper right corner have poorer disease severity (higher rank) and faster disease progression rate (higher rank). To visualize the effect of the correlation coefficient ρ, we have selected two patients as examples. Patient 213 has the worst disease severity and he or she ranks number 9 in disease progression rate. Patient 105 has the fastest disease progression rate and he or she ranks number 30 in disease severity.

Web Figure 2 displays the ranking of the posterior estimates (with 95% equal-tail credible intervals) of the center-specific random effects (v_h) on disease severity. We observe large variability of the center-specific random effects, in addition to different lengths of the credible intervals as a result of different numbers of patients in each center. Center 1 has the most patients with severe PD, whereas center 18 has the most patients with mild PD. It is interesting to note that the estimate of v_h of center 31 (the 'artificial' center) ranks number 17, and its length of credible interval is the smallest because the number of patients in this center is much larger than all other centers.

6. Discussion

In this article, we propose several hierarchical joint models that consist of an MLIRT submodel for multivariate longitudinal outcomes and a Cox submodel for a dependent terminal event while accounting for the possible correlation within each research center. We assume that the survival time may be correlated with the multivariate longitudinal outcomes at various subject and center levels. We adopt Bayesian inference framework based on MCMC simulation for parameter estimation. This Bayesian framework work provides accurate parameter estimates, in additional to various subject-specific PD disease severity estimation and center profiling. It also provides additional insight into the correlation between the multivariate longitudinal outcomes and the dependent terminal event at both subject and center levels. It has been shown in the extensive simulation studies that incorrect assumption on the correlation structure between longitudinal and survival outcomes results in considerable biases and poor CP in many parameters of interest. We apply the hierarchical joint models to analyze the motivating DATATOP study dataset and use various model selection criteria to select among all the candidate models. All proposed hierarchical joint models have better fit than the model without accounting for the within-center correlation. The treatment tocopherol has insignificant effects in either reducing the PD disease severity or slowing the PD disease progression. We identify significant correlation between the longitudinal and survival outcomes at the subject level but not at the center level. We provide subject-specific disease severity and disease progression rate estimates for all patients at each visit and the center profiling information, in addition to the figures for clear visualization. The proposed models have the capability to accommodate hierarchically structured data, to estimate latent disease severity at different levels, and to investigate the relationships between disease severity and predictors at different levels. The proposed model can be easily implemented using the publicly available `BUGS` language and can be readily accessible to, modified, and extended by applied researchers.

Our modeling framework provides great flexibility for extensions. For example, this article only considers a single-type terminal event. In the presence of multiple failure types, for example, levodopa therapy and dropout due to disease-related causes, the proposed joint model can be extended to accommodate competing risks survival data. Further, the treatment main effect ($\boldsymbol{\beta}_0$) is assumed constant over the follow-up period. This convenient assumption may not be realistic and may lead to bias in the coefficient

estimates [41, 42]. Varying-coefficient models have been studied in the literature [43–51]. Developing a formal nonparametric statistical model in the framework of MLIRT models to define and estimate the time-dependent treatment effect is part of our ongoing research. Moreover, it is of our interest to develop statistical tools that account for covariate measurement error in the MLIRT models.

Inference from the proposed Bayesian MLIRT modeling framework is valid when the model fits the data and the model assumptions are met. In this paper, we have assumed unidimensionality (i.e., there exists a single latent variable θ_{hij} to measure the disease severity). But there may be multiple latent variables representing multidimensional (e.g., sensoria, functions, and cognition) impairment caused by PD. Expanding the unidimensional MLIRT model to the multidimensional one is an interesting direction of future research. Also, we have assumed that local dependence (i.e., condition on random effects $\boldsymbol{u}_{hi}$ and v_h), all measurements y_{hijk} and time t_{hi} are independent. However, there may be multiple variables that are constructed and implemented as a unit of measurement (i.e., measuring the same ability such as activities of daily living). Therefore, these variables may have 'testlet' effect. How to detect the testlet effect and testlet dimension in the proposed model warrants further investigation. Moreover, differential item functioning (people with the same latent disease severity have different outcome measurements) detection in this complex hierarchical model is also 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 author thanks the editor, associate editor, and three anonymous referees for their reading and valuable comments. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performing computing resources that have contributed to the research results reported within this article. URL: http://www.tacc.utexas.edu.

References

1. Fahn S, Oakes D, Shoulson Ira, Kieburtz K, Rudolph A, Lang A, Olanow CW, Tanner C, Marek K, Parkinson Study Group. Levodopa and the progression of Parkinson's disease. *The New England Journal of Medicine* 2004; **351**(24):2498–2508.
2. Samii A, Null JG, Ransom BR. Parkinson's disease. *The Lancet* 2004; **363**:1783–1793.
3. Douglas JA. Item response models for longitudinal quality of life cata in clinical trials. *Statistics in Medicine* 1999; **18**: 2917–2931.
4. Wang C, Douglas J, Anderson S. Item response models for joint analysis of quality of life and survival. *Statistics in Medicine* 2002; **21**(1):129–142.
5. Glas CAW, Geerlings H, van de Laar MAFJ, Taal E. Analysis of longitudinal randomized clinical trials using item response models. *Contemporary Clinical Trials* 2009; **30**(2):158–170.
6. Luo S, Lawson AB, He B, Elm JJ, Tilley BC. Bayesian multiple imputation for missing multivariate longitudinal data from a Parkinson's disease clinical trial. *Statistical Methods in Medical Research* 2012; published online on Dec 12, 2012, PMID: 23242384 PMCID: PMC3883900.
7. Luo S, Ma J, Kieburtz KD. Robust Bayesian inference for multivariate longitudinal data using normal/independent distributions. *Statistics in Medicine* 2013; **33**(22):3812–3828.
8. He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research* 2013; published online on April 16, 2013, PMID: 23592717 PMCID: PMC3883896.
9. Luo S. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine* 2014; **33**:580–594.
10. Mislevy RJ. Estimation of latent group effects. *Journal of the American Statistical Association* 1985; **80**:993–997.
11. Adams RJ, Wilson M, Wu M. Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics* 1997; **22**(1):47–76.
12. Anderson TW. *An Introduction to Multivariate Statistical Analysis* 3rd ed. John Wiley & Sons: Hoboken, New Jersey, 2003.
13. Andersen EB. Latent regression analysis based on the rating scale model. *Psychology Science* 2004; **46**:209–226.
14. Christensen KB, Bjorner JB, Kreiner S, Petersen JH. Latent regression in loglinear Rasch models. *Communications in Statistics-Theory and Methods* 2004; **33**(6):1295–1313.
15. Pastor DA. The use of multilevel item response theory modeling in applied research: An illustration. *Applied Measurement in Education* 2003; **16**(3):223–243.
16. Weisscher N, Glas CA, Vermeulen M, De Haan RJ. The use of an item response theory–based disability item bank across diseases: accounting for differential item functioning. *Journal of Clinical Epidemiology* 2010; **63**(5):543–549.
17. Snitz BE, Yu L, Crane PK, Chang CCH, Hughes TF, Ganguli M. Subjective cognitive complaints of older adults at the population level: An item response theory analysis. *Alzheimer Disease & Associated Disorders* 2012; **26**(4):344–351.

18. Vaccarino AL, Anderson K, Borowsky B, Duff K, Giuliano J, Guttman M, Ho AK, Orth M, Paulsen JS, Sills T, van Kammen DP, Evans KR, PREDICT-HD and REGISTRY Investigators Coordinators. An item response analysis of the motor and behavioral subscales of the unified Huntington's disease rating scale in Huntington disease gene expansion carriers. *Movement Disorders* 2011; **26**(5):877–884.
19. Miller TM, Balsis S, Lowe DA, Benge JF, Doody RS. Item response theory reveals variability of functional impairment within clinical dementia rating scale stages. *Dementia and Geriatric Cognitive Disorders* 2012; **32**(5):362–366.
20. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**(4):465–480.
21. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**(15):1663–1685.
22. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–339.
23. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 1997; **16**:239–257.
24. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* 1997; **16**:259–272.
25. Shoulson I. DATATOP: a decade of neuroprotective inquiry. Parkinson Study Group. Deprenyl And Tocopherol Antioxidative Therapy Of Parkinsonism. *Annals of Neurology* 1998; **44**(3 Suppl 1):S160-6.
26. Elm JJ. 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* 2012; **27**(12): 1513–1521.
27. Liu L, Ma JZ, O'Quigley J. Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (ESRD) data. *Statistics in Medicine* 2008; **27**(27):5679–5691.
28. Liu L, Ma JZ, Johnson BA. A multi-level two-part random effects model, with application to an alcohol-dependence study. *Statistics in Medicine* 2008; **27**(18):3528–3539.
29. Kahan BC, Morris TP. Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects?. *Statistics in Medicine* 2013; **32**:1136–1149.
30. Schwab RS, England AC. Projection technique for evaluating surgery in Parkinson's disease. In *Third Symposium on Parkinson's Disease*. Livingstone: Edinburgh, 1969; 152–157.
31. Fox JP. *Bayesian Item Response Modeling: Theory and Applications*. Springer Verlag: New York, New York, 2010.
32. Lawless JF, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics* 1998; **26**(4):549–565.
33. Feng S, Wolfe RA, Port FK. Frailty survival model analysis of the national deceased donor kidney transplant dataset using poisson variance structures. *Journal of the American Statistical Association* 2005; **100**(471):728–735.
34. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2009; **58**(1):65–81.
35. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. *R news* 2006; **6**(1):12–17.
36. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. CRC Press: Boca Raton, FL, 2004.
37. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B:Statistical Methodology* 2002; **64**(4):583–639.
38. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC: Boca Raton, FL, 2009.
39. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association* 1995; **90**:773–795.
40. Lewis SM, Raftery AE. Estimating bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association* 1997; **92**:648–655.
41. Fan J, Zhang W. Statistical methods with varying coefficient models. *Statistics and its Interface* 2008; **1**(1):179–195.
42. Yu Z, Liu L, Bravata DM, Williams LS, Tepper RS. A semiparametric recurrent events model with time-varying coefficients. *Statistics in Medicine* 2013; **32**:1016–1026.
43. Amorim LD, Cai J, Zeng D, Barreto ML. Regression splines in the time-dependent coefficient rates model for recurrent event data. *Statistics in Medicine* 2008; **27**(28):5890–5906.
44. Cai J, Fan J, Zhou H, Zhou Y. Hazard models with varying coefficients for multivariate failure time data. *The Annals of Statistics* 2007; **35**(1):324–354.
45. Chiang CT, Wang MC. Varying-coefficient model for the occurrence rate function of recurrent events. *Annals of the Institute of Statistical Mathematics* 2009; **61**(1):197–213.
46. Kauermann G. Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics & Data Analysis* 2005; **49**(1):169–186.
47. Sun L, Zhou X, Guo S. Marginal regression models with time-varying coefficients for recurrent event data. *Statistics in Medicine* 2011; **30**(18):2265–2277.
48. Sun L, Zhu L, Sun J. Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Analysis* 2009; **100**(10):2214–2223.
49. Sun Y, Wu H. Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics* 2005; **32**(1):21–47.
50. Wong M, Lam KF, Lo E. Analysis of multilevel grouped survival data with time-varying regression coefficients. *Statistics in Medicine* 2011; **30**(3):250–259.
51. Yu Z, Lin X. Semiparametric regression with time-varying regression coefficients for failure time data analysis. *Statistica Sinica* 2010; **20**:853–869.

Statistics in Medicine

Supporting information

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