Statistics
in Medicine

Research Article

Received 20 November 2012, Accepted 30 July 2013 Published online 6 September 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5956

A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time‡

Sheng Luo*†

Impairment caused by Parkinson's disease (PD) is multidimensional (e.g., sensoria, functions, and cognition) and progressive. Its multidimensional nature precludes a single outcome to measure disease progression. Clinical trials of PD use multiple categorical and continuous longitudinal outcomes to assess the treatment effects on overall improvement. A terminal event such as death or dropout can stop the follow-up process. Moreover, the time to the terminal event may be dependent on the multivariate longitudinal measurements. In this article, we consider a joint random-effects model for the correlated outcomes. A multilevel item response theory model is used for the multivariate longitudinal outcomes and a parametric accelerated failure time model is used for the failure time because of the violation of proportional hazard assumption. These two models are linked via random effects. The Bayesian inference via MCMC is implemented in 'BUGS' language. Our proposed method is evaluated by a simulation study and is applied to DATATOP study, a motivating clinical trial to determine if deprenyl slows the progression of PD. © 2013 The authors. Statistics in Medicine published by John Wiley & Sons, Ltd.

Keywords: clinical trial; item response theory; failure time; latent variable; Markov chain Monte Carlo

1. Introduction

The impairment caused by many neurodegenerative diseases such as Parkinson's disease (PD), Alzheimer's disease, and Huntington's disease is multidimensional (e.g., sensoria, functions, and cognition) and progressive. Its multidimensional nature precludes a single outcome to measure disease progression. Many clinical trials studying these diseases have been conducted to search for neuroprotective treatments capable of halting or slowing down disease progression (e.g., deprenyl and tocopherol antioxidative therapy of parkinsonism (DATATOP) study [1], ELLDOPA study [2], PRECEPT study [3], TEMPO study [4], and ADAGIO study [5]). Clinical trials of neurodegenerative diseases often collect multiple longitudinal outcomes of mixed types (categorical and continuous) to assess the treatment effects on overall improvement.

The multivariate longitudinal data structure of these studies has three levels of nesting, that is, multiple outcomes (level 1) are nested within visits (level 2) that are nested within individuals (level 3). To determine the overall treatment effects, the analysis model needs to account for three sources of correlation within the same individual: (1) inter-source (different outcomes at the same visit); (2) longitudinal (same outcome at different visits); and (3) cross correlation (different outcomes at different visits) [6]. A univariate analysis (generalized estimating equations and mixed effects models) that deals with each outcome separately ignores the inter-source and cross correlations, fails to provide an overall treatment effects estimate, and is subject to type I error inflation [7]. Another commonly used approach of reducing the multivariate outcomes into a single summary outcome (e.g., weighted average) results in substantial

Division of Biostatistics, University of Texas School of Public Health, 1200 Pressler St., Houston, TX 77030, U.S.A.
**Correspondence to: Sheng Luo, Division of Biostatistics, University of Texas School of Public Health, 1200 Pressler St., Houston, Texas 77030, U.S.A.*
†*E-mail: sheng.t.luo@uth.tmc.edu*

loss of information and more importantly, results cannot be interpreted on the original outcome scale. Rank-based tests [8–10] for multiple outcomes have been used in several clinical studies [11–14]. But they neither utilize the full longitudinal data information nor describe the disease progression process.

Multilevel item response theory (MLIRT) models have been recently used to analyze such multivariate longitudinal data [15]. It is assumed that the multivariate outcomes are clinical manifestations of a univariate latent variable measuring disease severity. The MLIRT model consists of two levels. The first level measurement model quantifies the relationship between an individual's latent disease severity and the response to the multivariate outcomes. In the second level structural multilevel model, the latent disease severity is regressed on covariates (e.g., treatment and disease duration), time, and subject-specific random effects to study the treatment effects [16–20]. The three sources of correlation are accounted for via the random effects. Advantages of the MLIRT models include better reflection of multilevel data structure, simultaneous estimation of measurement-specific parameters and covariate effects, and accurate inference about high-level measures [21–23]. To obtain valid inference from the MLIRT models, marginal maximum likelihood methods [20], and Bayesian methods [23–29] have been widely used. Skrondal and Rabe-Hesketh [30] and [31] provide good description of the IRT models.

During the course of clinical trials, the follow-up of some individuals could be stopped by a terminal event such as death, dropout due to adverse event or severe adverse event, or some other events. Because the terminal event may be related to the individual's underlying disease severity, the terminal mechanism is nonignorable. The dependent terminal event time is often termed 'dependent censoring' or 'informative censoring'. Ignoring the dependent censoring leads to biased estimates [32, 33]. To address this issue, joint analysis of survival with repeated measures has been increasingly common [32–36]. Tsiatis and Davidian [37], and Yu *et al*. [38] give excellent review of joint modeling research. In the IRT modeling framework, Wang *et al*. [39] proposed a joint model to analyze multiple-item ordinal quality of life data in the presence of death. He and Luo [40] developed a joint model for multiple longitudinal outcomes of mixed types, subject to outcome-dependent terminal events. However, all these references focus on the proportional hazard (PH) model or its extensions. When the PH assumption is violated, the accelerated failure time (AFT) model is an attractive alternative approach. Tseng et al [41] proposed a joint modeling framework replacing the PH models by semiparametric AFT models. The advantage of the AFT model is that the interpretation of risk factors on the failure time is easy, because the AFT model simply regresses the logarithm of the survival time onto covariates and random effects.

In this article, we propose a joint modeling framework in which a MLIRT model is used for the multivariate longitudinal outcomes and a parametric AFT model is used for the dependent terminal event. The two models are linked via 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 a motivating clinical trial dataset. Section 6 gives some concluding remarks and discussions. To facilitate easy reading and implementation of the proposed methodology, the codes have been posted at the supporting information[§].

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 individuals with early PD will slow the progression of PD. A total of 800 individuals 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 individuals who did not receive deprenyl (double-placebo and active tocopherol alone groups, 401 individuals) are combined and referred to as placebo group. The individuals who received deprenyl (active deprenyl alone and both active tocopherol and deprenyl groups, 399 individuals) are combined and referred to as treatment group. Please refer to Shoulson *et al*.[1] for details of the DATATOP trial.

The multiple outcomes collected include Unified Parkinson's Disease Rating Scale (UPDRS) total score, Schwab and England activities of daily living (SEADL), Mini-mental State Exam (MMSE), and Hamilton rating scale for depression (HRSD), measured at five visits (baseline, months 1, 3, 9, and 15).

[§] *Supporting information may be found in the online version of this article.*

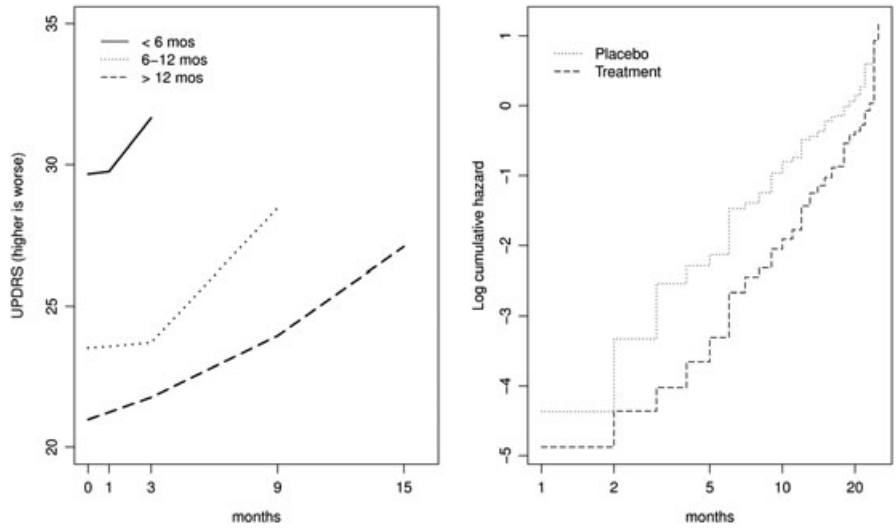

Figure 1. Mean Unified Parkinson's Disease Rating Scale (UPDRS) values over time (left panel), and log cumulative hazard curves for two treatment groups (right panel).

UPDRS is the sum of 44 questions each measured on a five-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 [42] and it is an ordinal variable with integer value from 0 to 100 incrementing by 5, with larger value reflecting better clinical outcomes. MMSE is a measurement of cognitive impairment and is an ordinal variable with integer value from 0 (severe) to 30 (normal). HRSD is a depression test measuring the severity of clinical depression symptoms and is an ordinal variable with integer value from 0 (normal) to 52 (severe). The outcomes SEADL and MMSE are recoded 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, MMSE, and HRSD have 7, 7, and 10 categories, respectively.

Before the end of the study, some individuals (223 and 154 individuals in the placebo and treatment groups, respectively) reached a level of functional disability sufficient to warrant the initiation of dopaminergic therapy, which is a symptomatic therapy to provide temporary relief of PD symptoms for a short period. In this case, only the observed outcomes before the initiation of dopaminergic therapy can be used in the assessment of treatment efficacy because dopaminergic therapy can artificially change the values of the outcomes collected. Therefore, these individuals would have missing data after the initiation of dopaminergic therapy. In the DATATOP study, the time to terminal event is defined as time to dopaminergic therapy.

The left panel of Figure 1 shows plot of mean UPDRS measurements over time for DATATOP individuals with follow-up time less than 6 months (219 individuals, solid line), 6–12 months (253 individuals, dotted line), and more than 12 months (328 individuals, dashed line). Individuals with shorter follow-up had higher UPDRS measurement, indicating that individuals with more severe PD symptom were more likely to have dopaminergic therapy. This phenomenon manifests strong correlation between the multivariate longitudinal values and the time to dopaminergic therapy. The right panel of Figure 1 displays the product-limit estimates of log cumulative hazard rates for the treatment variable. The cross-over of two curves and highly significant test based on scaled Schoenfeld residuals ($p < 0.001$) [43] suggest the violation of PH assumption if the treatment variable is the only covariate. To this end, an AFT model is an alternative approach to the Cox PH model.

3. Model and estimation

3.1. The multilevel item response model

We first introduce the MLIRT model. Let y_{ijk} (binary, ordinal, and continuous) be the observed outcome $k(k = 1, \ldots, K)$ from individual $i(i = 1, \ldots, N)$ at visit $j(j = 1, \ldots, J_i$, where $j = 1$ is baseline and J_i is the number of visits of individual i). Let $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijK})'$ be the vector of observations for individual i at visit j and let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iK})'$ be the outcome vector across visits. Throughout,

all outcomes are coded so that larger observation values correspond to worse clinical conditions. Let θ_{ij} be a univariate latent variable measuring disease severity of individual i 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_{ij} and outcome-specific parameters.

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

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

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

where $\epsilon_{ijk}\sim N\left(0,\sigma_k^2\right)$, 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 individuals with different latent disease severity θ_{ij}. 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 individual i is in category l on outcome k at visit j is $p(Y_{ijk}=l|\theta_{ij})=p(Y_{ijk}\leqslant l|\theta_{ij})-p(Y_{ijk}\leqslant l-1|\theta_{ij})$.

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

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

where $\boldsymbol{X}_{i0}$ and $\boldsymbol{X}_{i1}$ are vectors of covariates associated with baseline disease severity and disease progression rate, respectively, $\boldsymbol{X}_{i0}$ may or may not be the same as $\boldsymbol{X}_{i1}$, t_{ij} is visit time variable with $t_{i1}=0$ for baseline, random intercept u_{i0} and random slope u_{i1} determine the subject-specific baseline disease severity and disease progression rate, respectively. It is assumed that $\boldsymbol{u}_i=(u_{i0},u_{i1})'$ are from $N_2(0,\boldsymbol{\Sigma})$, with covariance matrix $\boldsymbol{\Sigma}$ being denoted by $((1,\rho\sigma_u),(\rho\sigma_u,\sigma_u^2))$, where the variance of u_{i0} is set to 1 for identifiability, σ_u^2 is the variance of u_{i1}, and ρ is the correlation coefficient. The random effects vector $\boldsymbol{u}_i$ takes into account all three sources of correlations illustrated in Section 1. For example, if $\theta_{ij}=u_{i0}+(\beta_{10}+\beta_{11}x_i+u_{i1})t_{ij}$, where x_i is treatment indicator (1 if treatment and 0 otherwise), then significant negative coefficient β_{11} indicates that the treatment slows down the disease progression. The combined level 1 and level 2 models are MLIRT with subject-specific covariance (referred to as subject-specific MLIRT models) [23–27,44,45].

Let t_i denote the time to terminal event for individual i, δ_i (1 if the terminal event is observed and 0 if not needed) denote the censoring indicator for t_i, and $\boldsymbol{X}_i$ denote vector of possible risk factors with the first element being 1. Vector $\boldsymbol{X}_i$ can share part of or all covariates in $\boldsymbol{X}_{i0}$ and $\boldsymbol{X}_{i1}$. The regular AFT model can be expressed as $\log(t_i)=\boldsymbol{X}_i\boldsymbol{\gamma}+\sigma_\epsilon\epsilon_i$, where $\boldsymbol{\gamma}$ is the unknown coefficient, ϵ_i is independent random error, and σ_ϵ is the scale parameter. The correlation between time to a terminal event and the longitudinal outcomes can be accounted for by sharing the random effects u_{i0} and u_{i1} in the AFT model as

$$\log(t_i|\boldsymbol{u}_i)=\boldsymbol{X}_i\boldsymbol{\gamma}+\eta_0u_{i0}+\eta_1u_{i1}+\sigma_\epsilon\epsilon_i. \tag{5}$$

When η_0 and η_1 are not equal to zero, the correlation between the survival time and longitudinal outcomes is incorporated, and the random effects have different effects on θ_{ij} and t_i.

Let $f_0(\cdot)$, $S_0(\cdot)$, and $h_0(\cdot)$ denote the density, survival, and hazard functions of random error ϵ in model (5), respectively. Let $f(\cdot)$, $S(\cdot)$, and $h(\cdot)$ denote the density, survival, and hazard functions of T, respectively. Then we have the following relationships: $f(t_i|\boldsymbol{u}_i)=\frac{1}{\sigma_\epsilon t_i}f_0\left(\frac{\log(t_i)-\lambda(\boldsymbol{X}_i)}{\sigma_\epsilon}\right)$, $S(t_i|\boldsymbol{u}_i)=S_0\left(\frac{\log(t_i)-\lambda(\boldsymbol{X}_i)}{\sigma_\epsilon}\right)$, and $h(t_i|\boldsymbol{u}_i)=\frac{1}{\sigma_\epsilon t_i}h_0\left(\frac{\log(t_i)-\lambda(\boldsymbol{X}_i|\boldsymbol{u}_i)}{\sigma_\epsilon}\right)$, where $\lambda\left(\boldsymbol{X}_i|\boldsymbol{u}_i\right)=\boldsymbol{X}_i\boldsymbol{\gamma}+\eta_0u_{i0}+\eta_1u_{i1}$. The common density distributions and the corresponding survival functions for ϵ_i are summarized in Table I. When ϵ_i follows normal, logistic, and extreme value distributions, event time t_i follows log-normal, log-logistic, and Weibull distributions, respectively. For example, if ϵ_i follows normal distribution with $S_0(\epsilon_i)=1-\Phi(\epsilon_i)$, where $\Phi(\cdot)$ denotes cumulative standard normal distribution function, then $S(t_i|\boldsymbol{u}_i)=1-\Phi[(\log(t_i)-\boldsymbol{X}_i\boldsymbol{\gamma}-\eta_0u_{i0}-\eta_1u_{i1})/\sigma_\epsilon]$. To solve this equation for t_i, $t_i=\exp[\sigma_\epsilon\Phi^{-1}(1-S)+\boldsymbol{X}_i\boldsymbol{\gamma}+\eta_0u_{i0}+\eta_1u_{i1}]$, where Φ^{-1} is the inverse function of $\Phi(\cdot)$. Under

Table I. Common distributions of ϵ_i in the parametric AFT models. $\phi(\cdot)$ and $\Phi(\cdot)$ denotes the probability density function and cumulative distribution function of the standard normal distribution.

Distribution	$f_0(\cdot)$	$S_0(\cdot)$
Normal	$\phi(\epsilon_i)$	$1-\Phi(\epsilon_i)$
Logistic	$\exp(\epsilon_i)/[1+\exp(\epsilon_i)]^2$	$1/(1+\exp(\epsilon_i))$
Extreme value	$\exp(\epsilon_i-\exp(\epsilon_i))$	$\exp(-\exp(\epsilon_i))$

the local independence assumption (i.e., conditioning on the random effects vector $\boldsymbol{u}_i$, all components in $\boldsymbol{y}_{ij}$ and t_i are independent), the full likelihood for individual i is

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

For notation convenience, we let the observed data be $\boldsymbol{y}=\{y_{ijk}\}\cup\{t_i\}\cup\{\delta_i\}$, the difficulty parameter vector be $\boldsymbol{a}=(\boldsymbol{a}_1',\ldots,\boldsymbol{a}_k',\ldots,\boldsymbol{a}_K')'$, with $\boldsymbol{a}_k=(a_{k1},\ldots,a_{kn_k-1})'$, the discrimination vector be $\boldsymbol{b}=(b_1,\ldots,b_K)'$, and $\boldsymbol{\beta}=(\boldsymbol{\beta}_0',\boldsymbol{\beta}_1')'$, and the parameter vector $\boldsymbol{\Psi}=(\boldsymbol{a}',\boldsymbol{b}',\boldsymbol{\beta}',\rho,\sigma_u,\sigma_k,\boldsymbol{\gamma},\sigma_\epsilon,\eta_0,\eta_1)'$. We refer to the proposed joint modeling framework assuming the log-normal, log-logistic, and Weibull distributions for survival time as models $\mathrm{JM_{LN}}$, $\mathrm{JM_{LL}}$, and $\mathrm{JM_W}$, respectively. In addition, we consider reduced models assuming independence between the survival time and longitudinal outcomes (i.e., $\eta_0=\eta_1=0$). We refer to the reduced models assuming the log-normal, log-logistic, and Weibull distributions for survival time as models $\mathrm{RM_{LN}}$, $\mathrm{RM_{LL}}$, and $\mathrm{RM_W}$, respectively.

3.2. *Bayesian inference*

To infer the unknown parameter vector $\boldsymbol{\Psi}$, we use Bayesian inference on the basis of MCMC posterior simulations. We use vague priors on all elements in the parameter vector $\boldsymbol{\Psi}$. Specifically, the prior distributions of all elements in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, η_0, and η_1 are $N(0,100)$. We use the prior distribution $b_k\sim\mathrm{Gamma}(0.01,0.01)$, $k=1,\ldots,K$, to ensure positivity. The prior distribution for the difficulty parameter a_k of the continuous outcomes is $a_k\sim N(0,2000)$ because some continuous measurements are quite large. To obtain the prior distributions for the threshold parameters of ordinal outcome k, we let a_{k1} $N(0,100)$, 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 $\rho\sim\mathrm{Uniform}[-1,1]$, and $\sigma_k,\sigma_u,\sigma_\epsilon\sim\mathrm{Gamma}(0.01,0.01)$.

The model fitting is performed in `OpenBUGS` (`OpenBUGS` version 3.2.2) by specifying the likelihood function and the prior distribution 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 $\hat{R}$ of all parameters are smaller than 1.1 [46].

There are a wide variety of model selection criteria in Bayesian inference. We adopt a model selection approach by using the deviance information criterion (DIC) proposed by Spiegelhalter *et al*. [47]. The DIC provides an assessment of model fitting and a penalty for model complexity. The deviance statistic is defined as $D(\boldsymbol{\Psi})=-2\log f(\boldsymbol{y}|\boldsymbol{\Psi})+2\log h(\boldsymbol{y})$, where $f(\boldsymbol{y}|\boldsymbol{\Psi})$ is the likelihood function for the observed data $\boldsymbol{y}$ given the parameter vector $\boldsymbol{\Psi}$ and $h(\boldsymbol{y})$ denotes a standardizing function of the data alone that has no impact on model selection [48]. The DIC is defined as $\mathrm{DIC}=2\overline{D}-D(\overline{\boldsymbol{\Psi}})=\overline{D}+p_D$, where $\overline{D}=E_{\boldsymbol{\Psi}|\boldsymbol{y}}[D]$ is the posterior mean of the deviance, $D(\overline{\boldsymbol{\Psi}})=D(E_{\boldsymbol{\Psi}|\boldsymbol{y}}[\boldsymbol{\Psi}])$ is the deviance evaluated at the posterior mean $\overline{\boldsymbol{\Psi}}$ of the parameter vector, and $p_D=\overline{D}-D(\overline{\boldsymbol{\Psi}})$ 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 BIC (or Schwarz) (EBIC) as model selection tools [48]. The EAIC and EBIC can be estimated as $EAIC=\overline{D}+2p$ and $EBIC=\overline{D}+p\log(N)$, where p is the number of elements in the parameter vector $\boldsymbol{\Psi}$. Smaller values of EAIC and EBIC indicate better predictive ability of the model.

4. Simulation study

In this section, we report results from a simulation study of two settings to compare the performance of the proposed joint models and reduced models. In each setting, we generate 100 datasets with sample size $N = 800$. The simulated data structure is similar to the DATATOP study with one continuous outcome and three ordinal outcomes with 7, 7, and 10 categories, and five visits (baseline, months 1, 3, 9, and 15). Each dataset is generated using the following algorithm.

(1) Consider a single binary covariate x_i as the treatment variable (1 if treatment and 0 if placebo), simulate $x_i \sim$ Bernoulli(0.5).
(2) Set $\boldsymbol{\beta} = (1, -0.5)'$, $\rho = 0.5$, and $\sigma_u = 2$, simulate the random effects vector $\boldsymbol{u}_i \sim N_2(0, \boldsymbol{\Sigma})$, and generate θ_{ij} for $j = 1, \ldots, J_i = 5$ from model (4) with $\boldsymbol{X}_{i0} = 0$ and $\boldsymbol{X}_{i1} = x_i$.
(3) Set $\sigma_1 = 5, a_1 = 25, b_1 = 10$, and generate the continuous outcome y_{ij1} from model (3).
(4) Set $\boldsymbol{a}_2 = (-2.6, -0.6, 2, 2.8, 4.9, 5.9)$, $\boldsymbol{a}_3 = (-0.1, 1, 1.8, 2.6, 3.3, 4.1)$, $\boldsymbol{a}_4 = (-0.9, 0, 0.5, 1, 1.5, 1.9, 2.4, 2.8, 3.3)$, $b_2 = 2.0, b_3 = 0.4, b_4 = 0.65$, and simulate ordinal outcomes y_{ij2}, y_{ij3} and y_{ij4} from model (2) for $j = 1, \ldots, J_i = 5$.
(5) Set $\boldsymbol{X}_i = (1, x_i)$. Set $\boldsymbol{\gamma} = (0.6, 0.5)'$, $\eta_0 = \eta_1 = 0$ for setting I and set $\boldsymbol{\gamma} = (1.5, 0.5)'$, $\eta_0 = -0.2$, $\eta_1 = -0.8$ for setting II. Simulate S_i from Uniform[0, 1] and generate t_i in months. Let $\delta_i = 1$ if $t_i \leqslant 15$ months, and $\delta_i = 0$, otherwise.
(6) Repeat steps 1 to 5 until all individuals' data are generated.

We apply the Bayesian framework in Section 3.2 to obtain inference. To determine the burn-in iterations and assess the MCMC convergence and mixing properties, we examine the trace plots and the autocorrelations. We find that the chains converge reasonably fast and all achieve stationarity within 5000 iterations with $\hat{R}$ of all parameters smaller than 1.1. We run two parallel MCMC chains with diverse initial values and choose 5000 iterations for burn-in and the inference is based on the subsequent 5000 iterations from both chains. The running time for each MCMC chain is around 11 h.

In setting I, there is no correlation between the survival time and the longitudinal outcomes (i.e., $\eta_0 = \eta_1 = 0$) and there are around 20% of individuals experiencing a terminal event. In setting II, we set $\eta_0 = -0.2$ and $\eta_1 = -0.8$, indicating that the individuals who are more severe in disease are more likely to have a terminal event. For this setting, around 20% of individuals experience a terminal event. The estimation results with 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 standard deviation of the posterior means and coverage probabilities (*CP*) of 95% equal-tail credible intervals of all reduced and joint models are displayed in Tables II and III.

Table II suggests that in setting I with no correlation (i.e., $\eta_0 = \eta_1 = 0$), all reduced and joint models generate comparable results, that is, the bias is negligible, *SE* is close to *SD*, and the credible interval coverage probabilities are reasonably close to 95%. Under model overparameterization, the estimates of η_0 and η_1 from all joint models are correctly close to zero. The estimates of the difficulty parameter vector $\boldsymbol{a}$, the discrimination vector $\boldsymbol{b}$, and σ_1 from all reduced and joint models have small bias, *SE* close to *SD*, and *CP* close to 95% (results not shown due to the space constraint).

Table III suggests that in setting II with some correlation (i.e., $\eta_0 = -0.2, \eta_1 = -0.8$), all joint models provide estimates of all parameters with negligible bias, *SE* being close to *SD*, *CP* being reasonably around the nominal value. In contrast, the reduced models give biased estimates and low coverage probabilities, especially for $\beta_{10}, \beta_{11}, \rho, \sigma_u, \gamma_0$, and σ_ϵ. Specifically, the estimate of treatment effects parameter ($\hat{\beta}_{11}$) is biased toward zero. By setting negative β_{11}, the treatment is effective and the individuals in the placebo group are more likely to have the terminal event. Ignoring this phenomenon and treating the missing data after the terminal event as missing at random, the reduced models tend to reduce the difference between two groups and therefore underestimate the treatment effects. This new finding in the MLIRT framework is consistent with what has been reported in the analysis of univariate longitudinal data with dependent dropouts using marginal or random effects models [49]. The estimates of the difficulty parameter vector $\boldsymbol{a}$, the discrimination vector $\boldsymbol{b}$, and σ_1 from all reduced and joint models have small bias, *SE* close to *SD*, and *CP* close to 95% (results not shown because of the space constraint).

Because the survival time is modeled by parametric AFT model, it is essential to assess the robustness of the proposed joint models under model misspecification. Table IV displays the results of fitting model JM_{LN} when data are simulated from model JM_{LL} under setting II with some correlation. The results indicate that all parameters, except the scale parameter σ_ϵ, have reasonable estimates with small bias,

Table II. Results of fitting various reduced and joint models for setting I in which the terminal event is *independent* on the longitudinal outcomes.

	RM_{LN}				JM_{LN}			
	BIAS	*SE*	*SD*	*CP*	*BIAS*	*SE*	*SD*	*CP*
For longitudinal outcomes								
$\beta_{10} = 1.000$	0.005	0.102	0.107	0.960	−0.001	0.103	0.108	0.950
$\beta_{11} = -0.500$	0.002	0.133	0.137	0.950	0.009	0.133	0.129	0.970
$\rho = 0.500$	−0.005	0.030	0.031	0.920	−0.000	0.030	0.031	0.930
$\sigma_u = 2.000$	−0.004	0.069	0.065	0.960	−0.002	0.069	0.066	0.960
For survival								
$\gamma_0 = 0.400$	0.003	0.032	0.030	0.980	0.006	0.033	0.032	0.950
$\gamma_1 = 0.500$	0.009	0.058	0.050	0.980	0.016	0.060	0.051	0.990
$\sigma_\epsilon = 0.400$	0.003	0.028	0.026	0.960	0.006	0.028	0.024	0.950
$\eta_0 = 0$					0.000	0.029	0.028	0.980
$\eta_1 = 0$					-0.000	0.016	0.016	0.950
	RM_{LL}				JM_{LL}			
	BIAS	*SE*	*SD*	*CP*	*BIAS*	*SE*	*SD*	*CP*
For longitudinal outcomes								
$\beta_{10} = 1.000$	0.014	0.104	0.111	0.920	0.007	0.105	0.107	0.960
$\beta_{11} = -0.500$	−0.005	0.136	0.142	0.950	0.006	0.134	0.143	0.960
$\rho = 0.500$	−0.000	0.030	0.032	0.950	−0.002	0.031	0.031	0.940
$\sigma_u = 2.000$	−0.001	0.070	0.064	0.960	−0.007	0.070	0.065	0.950
For survival								
$\gamma_0 = 0.600$	0.004	0.057	0.051	0.970	0.014	0.059	0.054	0.970
$\gamma_1 = 0.500$	0.018	0.089	0.083	0.980	0.022	0.090	0.082	0.960
$\sigma_\epsilon = 0.400$	0.006	0.031	0.030	0.930	0.008	0.032	0.031	0.930
$\eta_0 = 0$					0.002	0.052	0.051	0.960
$\eta_1 = 0$					0.000	0.029	0.028	0.960
	RM_{W}				JM_{W}			
	BIAS	*SE*	*SD*	*CP*	*BIAS*	*SE*	*SD*	*CP*
For longitudinal outcomes								
$\beta_{10} = 1.000$	0.011	0.105	0.114	0.940	0.012	0.105	0.107	0.970
$\beta_{11} = -0.500$	0.006	0.135	0.140	0.960	−0.018	0.136	0.125	1.000
$\rho = 0.500$	−0.001	0.031	0.031	0.930	−0.001	0.031	0.030	0.940
$\sigma_u = 2.000$	0.004	0.071	0.066	0.960	−0.005	0.070	0.065	0.960
For survival								
$\gamma_0 = 0.600$	0.001	0.047	0.044	0.970	0.009	0.048	0.046	0.970
$\gamma_1 = 0.500$	0.015	0.081	0.078	0.960	0.025	0.082	0.078	0.960
$\sigma_\epsilon = 0.400$	0.003	0.030	0.028	0.940	0.009	0.030	0.028	0.940
$\eta_0 = 0$					0.002	0.042	0.043	0.920
$\eta_1 = 0$					−0.000	0.023	0.022	0.970

CP, coverage probabilities.

SE being close to *SD*, and *CP* being around the nominal value. The poor estimate of σ_ϵ (large bias and CP far from 0.95) should not be surprising because of the incorrect assumption of the distribution of ϵ_i. The parameters $\boldsymbol{a}, \boldsymbol{b}$, and σ_1 from model JM_{LN} are correctly estimated (results not shown because of the space constraint).

From the simulation study, we conclude that the joint models provide results comparable with the reduced models under independent terminal events. Under the dependent terminal mechanism, the joint models provide more accurate estimates for the MLIRT and AFT regression parameters and random effects parameters than the reduced models, while all models give reasonable estimates for outcome-specific parameters. Moreover, when the AFT model for the survival time is misspecified, the proposed joint models still provide accurate estimates for all parameters except the scale parameter σ_ϵ.

Table III. Results of fitting various reduced and joint models for setting II in which the terminal event is *dependent* on the longitudinal outcomes.

	RM_{LN}				JM_{LN}			
	BIAS	*SE*	*SD*	*CP*	*BIAS*	*SE*	*SD*	*CP*
For longitudinal outcomes								
$\beta_{10} = 1.000$	−0.322	0.093	0.100	0.110	−0.004	0.103	0.104	0.940
$\beta_{11} = -0.500$	0.110	0.123	0.115	0.860	0.016	0.132	0.125	0.950
$\rho = 0.500$	−0.043	0.034	0.034	0.780	−0.004	0.030	0.031	0.930
$\sigma_u = 2.000$	−0.292	0.065	0.061	0.010	−0.010	0.076	0.069	0.970
For survival								
$\gamma_0 = 1.500$	−0.017	0.163	0.175	0.920	0.032	0.149	0.157	0.950
$\gamma_1 = 0.500$	0.010	0.175	0.189	0.900	−0.001	0.132	0.125	0.980
$\sigma_\epsilon = 0.400$	1.348	0.120	0.132	0.000	0.040	0.072	0.061	0.940
$\eta_0 = -0.2$					−0.007	0.070	0.064	0.970
$\eta_1 = -0.8$					−0.010	0.063	0.062	0.960
	RM_{LL}				JM_{LL}			
	BIAS	*SE*	*SD*	*CP*	*BIAS*	*SE*	*SD*	*CP*
For longitudinal outcomes								
$\beta_{10} = 1.000$	−0.334	0.093	0.100	0.090	0.008	0.106	0.114	0.910
$\beta_{11} = -0.500$	0.112	0.123	0.114	0.850	0.002	0.135	0.131	0.960
$\rho = 0.500$	−0.042	0.034	0.033	0.810	−0.007	0.031	0.032	0.910
$\sigma_u = 2.000$	−0.277	0.066	0.063	0.020	−0.001	0.078	0.072	0.960
For survival								
$\gamma_0 = 1.500$	−0.250	0.140	0.136	0.580	0.001	0.155	0.149	0.970
$\gamma_1 = 0.500$	−0.037	0.169	0.175	0.940	0.001	0.146	0.154	0.950
$\sigma_\epsilon = 0.400$	0.537	0.069	0.060	0.000	−0.002	0.049	0.049	0.950
$\eta_0 = -0.2$					0.009	0.083	0.082	0.960
$\eta_1 = -0.8$					−0.005	0.066	0.064	0.940
	RM_{W}				JM_{W}			
	BIAS	*SE*	*SD*	*CP*	*BIAS*	*SE*	*SD*	*CP*
For longitudinal outcomes								
$\beta_{10} = 1.000$	−0.397	0.091	0.093	0.000	−0.029	0.103	0.120	0.930
$\beta_{11} = -0.500$	0.134	0.122	0.109	0.810	0.027	0.130	0.130	0.930
$\rho = 0.500$	−0.048	0.035	0.038	0.720	−0.007	0.031	0.035	0.900
$\sigma_u = 2.000$	−0.321	0.065	0.059	0.010	−0.013	0.078	0.074	0.950
For survival								
$\gamma_0 = 1.500$	−0.169	0.128	0.122	0.730	0.031	0.138	0.161	0.900
$\gamma_1 = 0.500$	−0.066	0.152	0.157	0.950	−0.017	0.129	0.136	0.920
$\sigma_\epsilon = 0.400$	0.602	0.071	0.062	0.000	0.034	0.052	0.045	0.920
$\eta_0 = -0.2$					−0.015	0.068	0.067	0.950
$\eta_1 = -0.8$					0.001	0.061	0.071	0.920

CP, coverage probabilities.

5. Data analysis

We apply the proposed joint models to the motivating DATATOP study. We use two parallel chains with overdispersed initial values and run each chain for 20,000 iterations. The first 10,000 iterations are discarded as burn-in; the parameter estimates are based on the remaining 10,000 iterations from each chain. The running time for each MCMC chain is around 20 h. Good mixing properties of the chains for the model parameter are observed in the trace plots.

We let $\boldsymbol{X}_{i0} = 0$ and consider the treatment assignment variable x_i (1 treatment, and 0 if placebo) as a single covariate in $\boldsymbol{X}_{i1}$. Hence, the level 2 model (4) is $\theta_{ij} = u_{i0} + [\beta_{10} + \beta_{11}x_i + u_{i1}]t_{ij}$, with visit

Table IV. Results of fitting model JM_{LN} when data are simulated from model JM_{LL} under setting II with some correlation.

	JM_{LN}			
	BIAS	*SE*	*SD*	*CP*
For longitudinal outcomes				
$\beta_{10} = 1.000$	0.002	0.106	0.114	0.910
$\beta_{11} = -0.500$	0.008	0.135	0.128	0.960
$\rho = 0.500$	−0.003	0.031	0.032	0.910
$\sigma_u = 2.000$	0.002	0.078	0.074	0.940
For survival				
$\gamma_0 = 1.500$	0.026	0.156	0.145	0.970
$\gamma_1 = 0.500$	0.002	0.147	0.152	0.920
$\sigma_\epsilon = 0.400$	0.309	0.085	0.094	0.030
$\eta_0 = -0.2$	0.013	0.084	0.082	0.970
$\eta_1 = -0.8$	−0.016	0.065	0.063	0.960

CP, coverage probabilities.

Table V. Model comparison statistics for the deprenyl and tocopherol antioxidative therapy of parkinsonism dataset. *Dbar*, the posterior mean of the deviance; *EAIC*, expected AIC; *EBIC*, expected BIC; *DIC*, deviance information criterion. Boldface indicates the preferred model.

AFT	Reduced models				Joint models			
	Dbar	*EAIC*	*EBIC*	*DIC*	*Dbar*	*EAIC*	*EBIC*	*DIC*
Log-normal	50626.6	50694.6	50853.9	51651.2	**50122.0**	**50194.0**	**50362.6**	**51302.7**
Log-logistic	50597.3	50665.3	50824.6	51620.1	50220.9	50292.9	50461.5	51340.1
Weibull	50584.2	50652.2	50811.5	51610.0	50269.3	50341.3	50509.9	51343.6

Table VI. Results of fitting the reduced log-normal model RM_{LN} and the joint log-normal model JM_{LN} in the deprenyl and tocopherol antioxidative therapy of parkinsonism dataset.

	RM_{LN}				JM_{LN}			
	Mean	*SD*	95% *CI*		Mean	*SD*	95% *CI*	
For longitudinal outcomes								
β_{10}	0.788	0.050	0.688	0.885	0.989	0.060	0.871	1.110
β_{11}	−0.382	0.058	−0.494	−0.266	−0.474	0.062	−0.600	−0.353
ρ	0.340	0.072	0.199	0.477	0.409	0.060	0.286	0.525
σ_u	0.454	0.036	0.383	0.525	0.539	0.041	0.457	0.618
For survival								
γ_0	0.111	0.048	0.018	0.204	0.142	0.047	0.051	0.237
γ_1	0.484	0.068	0.354	0.620	0.459	0.065	0.333	0.585
σ_ϵ	0.840	0.032	0.781	0.904	0.560	0.050	0.463	0.656
η_0					−0.155	0.056	−0.266	−0.045
η_1					−1.041	0.136	−1.341	−0.788

times being transformed in years as $t_{ij} = (0, 1, 3, 9, 15)/12$. The AFT model for the time to dopaminergic therapy is $\log(t_i) = \gamma_0 + \gamma_1 x_i + \eta_0 u_{i0} + \eta_1 u_{i1} + \sigma_\epsilon \epsilon_i$, with ϵ_i following normal, logistic, and extreme value distributions. Table V compares all reduced and joint models by using the model selection criteria discussed in Section 3.2. All three joint models perform significantly better than their reduced model counterparts with smaller Dbar, EAIC, EBIC, and DIC values, suggesting that the joint models are more preferable than their reduced model counterparts. Model JM_{LN} is selected as the final model because it has the best predictive ability with the smallest Dbar, EAIC, EBIC, and DIC values.

Table VI compares the results from the best predicative model JM_{LN} and its reduced model counterpart (results from other models are given in Table 1 of the supporting information). It is observed that both models give different estimates for all parameters, although the same set of parameters are identified for significance by both models. For example, both models suggest significant deprenyl effects in longitudinal outcomes ($\hat{\beta}_{11} = -0.382$ with 95% *CI* $[-0.494, -0.266]$ in model RM_{LN} vs. $\hat{\beta}_{11} = -0.474$

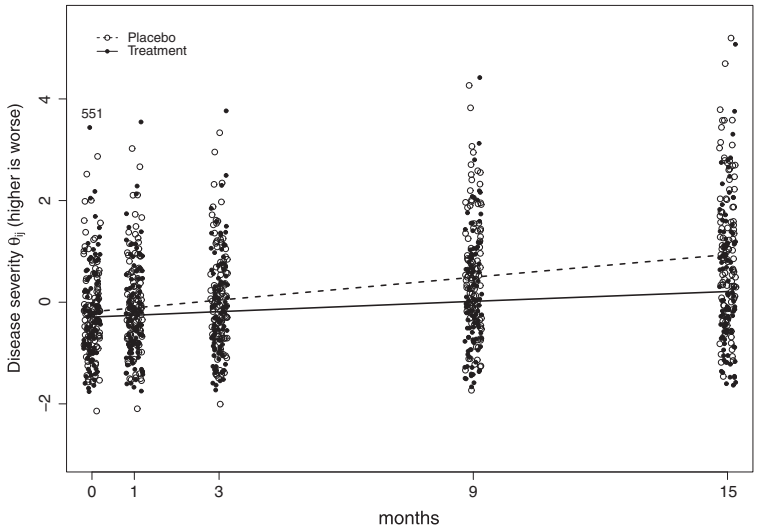

Figure 2. Estimates of the subject-specific disease severity θ_{ij} of 100 randomly selected individuals at each visit and the lowess smooth curves for the placebo and treatment groups. The number in the figure is an individual's subject ID.

with 95% *CI* $[-0.600, -0.353]$ in model JM_{LN}). Moreover, deprenyl has significant effects on the probability of having dopaminergic therapy, that is, the time to dopaminergic therapy for deprenyl individuals is estimated to be 1.622 times ($\exp(\hat{\gamma}_1)$, 95% *CI*: [1.425, 1.859]) that of nondeprenyl individuals in model RM_{LN} vs. 1.582 times (95% *CI*: [1.395, 1.795]) in model JM_{LN}. We also observe that $\hat{\eta}_0$ and $\hat{\eta}_1$ are significantly different from zero ($\hat{\eta}_0 = -0.155$, 95% *CI* $[-0.266, -0.045]$; $\hat{\eta}_1 = -1.041$, 95% *CI* $[-1.341, -0.788]$), indicating that individuals with worse baseline disease severity (higher u_{i0}) and higher disease progression rate (higher u_{i1}) tend to need dopaminergic therapy earlier.

The results from model JM_{LN} in Table VI indicate that the placebo individuals show significant deterioration across time with the disease progression rate being 0.989 units per year ($\hat{\beta}_{10}$, 95% *CI*: [0.871, 1.110]). Although the treatment individuals also show significant deterioration across time with disease progression rate being 0.515 units per year ($\hat{\beta}_{10}+\hat{\beta}_{11}$, 95% *CI*: [0.271, 0.747]), the treatment significantly slows down the disease progression rate by -0.474($\hat{\beta}_{11}$, 95% *CI*: $[-0.600, -0.353]$) units per year, suggesting the efficacy of the study drug deprenyl. The estimates of the outcome-specific parameters ($\boldsymbol{a}$ and $\boldsymbol{b}$) from models RM_{LN} and JM_{LN} are given in Table 2 of the supporting information. To visualize the difference in the disease progression rates in two groups, Figure 2 displays the posterior estimate of 100 randomly selected individuals' subject-specific latent disease severity at each visit. The lowess smooth curves [50] for the placebo and the treatment groups are denoted by the dashed and solid lines, respectively. Figure 2 shows that the placebo individuals' PD severity deteriorates at a much faster rate than the treatment individuals as manifested by the departure of two lowess curves, especially at months 9 and 15. Moreover, Figure 2 reveals one treatment individual (individual 551) who has much worse disease severity than all other individuals, because this individual has the worse SEADL and MMSE measures.

To assess model fit, Figure 3 displays the Cox-Snell residuals for models RM_{LN} and JM_{LN}. The estimated cumulative hazard for model RM_{LN} initially falls on the reference line and then falls mostly above the line, providing some evidence of poor model fit. In comparison, the estimated cumulative hazard for model JM_{LN} falls closer to the reference line, although this is some departure from the line at larger values of the cumulative hazard. Based on this figure, it appears that model JM_{LN} provides a reasonable fit to the data.

To provide some clinical insight into the regression parameters β_{10} and β_{11}, we tabulate in Table VII the change from baseline to each follow-up visit for the outcome UPDRS, and the odds ratio of the cumulative probability at a certain threshold of the outcome SEADL. At month 3, the placebo individuals are expected to increase 2.670 (95% *CI*: [2.387, 2.959]) units in UPDRS, and the odds ratio of the cumulative probability at a certain threshold of SEADL is expected to be 0.625 (95% *CI*: [0.590, 0.660]) compared with baseline, while the treatment individuals are expected to increase 1.390 (95% *CI*: [1.168, 1.615]) units in UPDRS, and the odds ratio of the cumulative probability at a certain threshold of SEADL

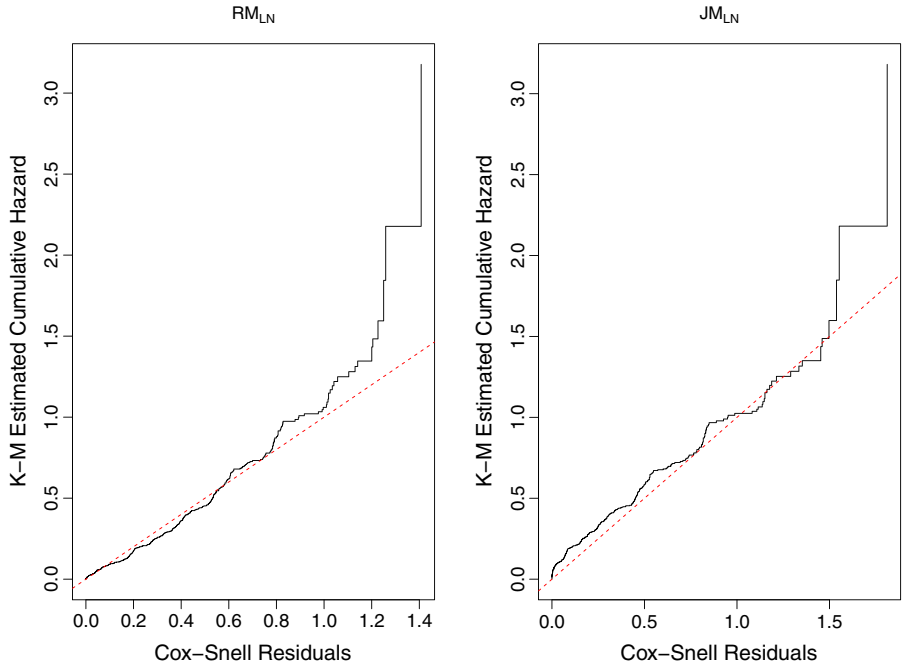

Figure 3. Cox–Snell residual plots for the reduced log-normal model RM_{LN} and the joint log-normal model JM_{LN}.

Table VII. Change from baseline to each follow-up visit for the outcome Unified Parkinson's Disease Rating Scale (UPDRS) and the *OR* of the cumulative probability at a certain threshold of the outcome Schwab and England activities of daily living (SEADL). The numbers in the subscript are the *SD*. The numbers within the square brackets are 95% equal-tailed credible intervals.

	UPDRS		$OR\{p(\text{SEADL} \leqslant l)\}$	
	Placebo	Treatment	Placebo	Treatment
Month 3	$2.670_{0.144}$ [2.387, 2.959]	$1.390_{0.113}$ [1.168, 1.615]	$0.625_{0.018}$ [0.590, 0.660]	$0.783_{0.017}$ [0.750, 0.816]
Month 9	$8.009_{0.431}$ [7.162, 8.876]	$4.170_{0.338}$ [3.503, 4.846]	$0.244_{0.021}$ [0.206, 0.288]	$0.480_{0.030}$ [0.421, 0.544]
Month 15	$13.349_{0.718}$ [11.937, 14.794]	$6.950_{0.563}$ [5.839, 8.076]	$0.096_{0.014}$ [0.072, 0.126]	$0.295_{0.031}$ [0.237, 0.362]

is expected to be 0.783 (95% *CI*: [0.750, 0.816]) compared with baseline. At month 9, the placebo individuals are expected to increase 8.009 (95% *CI*: [7.162, 8.876]) units in UPDRS, and the odds ratio of the cumulative probability at a certain threshold of SEADL is expected to be 0.244 (95% *CI*: [0.206, 0.288]) compared with baseline, while the treatment individuals are expected to increase 4.170 (95% *CI*: [3.503, 4.846]) units in UPDRS, and the odds ratio of the cumulative probability at a certain threshold of SEADL is expected to be 0.480 (95% *CI*: [0.421, 0.544]) compared with baseline. At month 15, the placebo individuals are expected to increase 13.349 (95% *CI*: [11.937, 14.794]) units in UPDRS, and the odds ratio of the cumulative probability at a certain threshold of SEADL is expected to be 0.096 (95% *CI*: [0.072, 0.126]) compared with baseline, while the treatment individuals are expected to increase 6.950 (95% *CI*: [5.839, 8.076]) units in UPDRS, and the odds ratio of the cumulative probability at a certain threshold of SEADL is expected to be 0.295 (95% *CI*: [0.237, 0.362]) compared with baseline.

In Table VI, the *SE* (σ_u) of the random intercept (u_{i1}) from model JM_{LN} is 0.539 (95% *CI*: [0.457, 0.618]), while the estimate of the correlation coefficient ρ between u_{i0} and u_{i1} is 0.409 (95% *CI*: [0.286, 0.525]). The positive correlation coefficient indicates that individuals whose baseline level of disease severity is worse than the average population tend to have faster disease progression rate and vice versa. To gain further insight into u_{i0}, u_{i1}, and ρ, we plot in Figure 4 the rankings of individuals' subject-specific baseline disease severity u_{i0} (upper panel) and disease progression rate u_{i1} (lower panel). Each

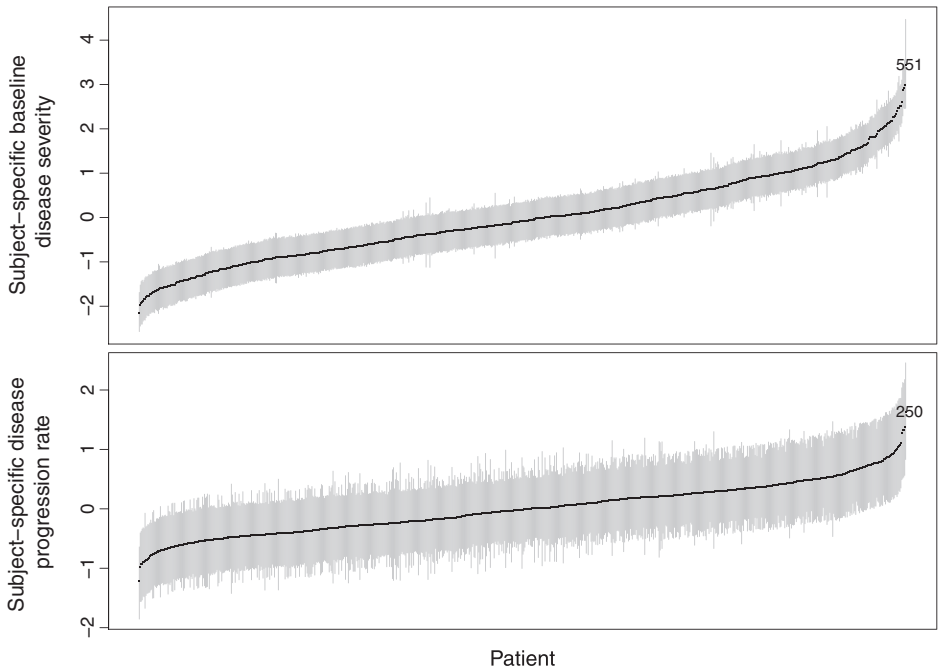

Figure 4. The ranking of the subject-specific baseline disease severity (upper panel) and the disease progression rate (lower panel) with point estimates and 95% *CI*s. The numbers in the figures are individuals' subject ID.

individual is ordered by his or her rank: individuals at the bottom left corner show milder baseline disease severity (lower rank) and slower disease progression rate (lower rank), while individuals at the upper right corner have poorer baseline disease severity (higher rank) and faster disease progression rate (higher rank). To visualize the effect of the correlation coefficient ρ, we have selected two individuals as examples. Individual 551 has the worst baseline disease severity and he/she ranks number 27 in disease progression rate. Individual 250 has the slowest disease progression rate while he/she ranks number 31 in baseline disease severity.

6. Discussion

In this article, we propose a joint modeling framework that consists of a MLIRT model for the multivariate longitudinal outcomes and a parametric AFT model for the dependent terminal event. The MLIRT model and the AFT model is linked together via shared random effects. Our simulation study shows that when there exists a dependent terminal event, the proposed joint models provide accurate estimates to all parameters while the reduced models underestimate the treatment effects and give severely biased estimates to multiple other parameters, leading to incorrect inference. When the terminal event is uncorrelated to the longitudinal outcomes, the joint models provide comparable results to the reduced models. Standard available software packages `R` and `OpenBUGS` are used to analyze the motivating DATATOP study data, and various model selection criteria are used to select between the reduced and joint models in addition to the appropriate assumption for the error term distributions. All joint models have better fit than their reduced model counterparts. The target treatment deprenyl is effective in slowing down PD progression. We have identified the positive correlation between baseline disease severity and disease progression rate. We provide subject-specific disease severity estimates for all individuals at each visit and the figure to visualize the different disease progression rates in the placebo and treatment groups.

The proposed joint modeling framework is a shared random effects model, also called a selection model in missing data problems, in which a characteristic of the longitudinal process defined as a function of the random effects is included in the survival model [34]. An alternative approach is joint models based on pattern mixture modeling (PMM) which stratify the individuals into subgroups by their censoring patterns and then model the stratified subgroups separately, for example, Hogan and Laird [35, 36], Molenberghs and Verbeke [51, chapter. 30], Zhang et al [52]. A good review of selection models and PMM models in the context of joint modeling can be found in Sousa [53]. More recently, joint latent

class models (JLCM) have been proposed as a way to relax PMM assumptions. JLCM models assume that the heterogeneous population of individuals consists of homogeneous latent subgroups of individuals that share the same marker trajectory and the same risk of the event [54, 55]. Dantan *et al*. [56] and Proust-Lima *et al*. [57] provide an excellent overview of the JLCM model and its difference from the shared random effects model model and PMM model. As a future research direction, we will investigate the application of the PMM model and JLCM model to multivariate longitudinal data in PD studies and compare them with the proposed model.

Of note, in the structural multilevel model (4), the latent disease severity θ_{ij} is determined by the random effects u_{i0}, u_{i1}, and covariate vector $\boldsymbol{X}_i$, while model (4) of Fox and Glas [24] is more flexible in allowing additional randomness by adding a random error term ϵ_{ij}. But this flexibility does not come without paying a price. As pointed out by Fox and Glas [24], their model must be identified by setting one of three constraints: (1) fixing the mean and variance of θ_{ij} to zero and one; (2) imposing $\sum_{k=1}^{K} a_k = 0$ and $\prod_{k=1}^{K} b_k = 1$; (3) fixing one discrimination parameter to one, and one difficulty parameter to zero ('the most convenient way' [24]). In contrast, the proposed model is identified by setting the mean and variance of u_{i0} to zero and one. From model (4), we have $E(\theta_{ij}) = \boldsymbol{X}_{i0}\boldsymbol{\beta}_0$ when $t_{ij} = 0$, representing the expected baseline disease severity. This identifiability constraint is much weaker than constraint (1) listed previously, which assumes $E(\theta_{ij}) = 0$. While in educational and social research with multiple similar test items, it may be reasonable to set either constraint (2) or constraint (3), in the context of the proposed PD clinical trial, it is very difficulty to justify which outcome's a_k and b_k should be set to 0 and 1, or why $\sum_{k=1}^{K} a_k = 0$ and $\prod_{k=1}^{K} b_k = 1$ should be satisfied because all outcomes are in vastly different scales. However, if one wants to add an extra random error term ϵ_{ij} to model (4), one of the aforementioned constraints is applied, and the assumption of $Var(u_{i0}) \equiv 1$ can be relaxed. The computing codes can be easily modified to incorporate this change.

Our model can be easily generalized. To model the nonlinear disease progression rate, the linear time trend assumption in model (4) can be relaxed by adding higher order terms or smoothing spline [58] terms of time. Moreover, it is assumed that there exists a single (unidimensional) latent variable θ_{ij} to measure the underlying disease severity. However, 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. Our joint models involve parametric AFT models. The models can be more flexible if the parametric assumption is relaxed by setting the intercept in model (5) to zero and assuming that the random error ϵ_i follows a mixture of Gaussian distributions. Moreover, some smoothing terms rather than linear terms for covariate modeling can be included in the AFT model. We would like to pursue these directions in our future research.

Acknowledgements

The author's research was supported in part by National Institutes of Health/National Institute of Neurological Disorders and Stroke grants U01NS043127 and U01NS43128, and National Institutes of Health/National Center for Advancing Translational Sciences grant KL2 TR000370. The author thanks the editor, associate editor, and two anonymous referees for their reading and valuable comments. The author is grateful to Bo He and Jing Huang for programming assistance. Computations were performed on the high-performance computational capabilities of the Linux cluster system at University of Texas School of Public Health. The author expresses appreciation to UTSPH information technology staff for their technical support of the cluster.

References

1. 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.
2. Fahn S, Oakes D, Shoulson I, Kieburtz K, Rudolph A, Lang A, Olanow C, 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.
3. Shoulson I. CEP – 1347 treatment fails to favorably modify the progression of Parkinson's disease (PRECEPT) study. *Neurology* 2006; **67**(1):185–185.
4. Group PS. A controlled trial of rasagiline in early Parkinson disease: the TEMPO Study. *Archives of Neurology* 2002; **59**(12):1937–43.
5. Olanow C, Hauser R, Jankovic J, Langston W, Lang A, Poewe W, Tolosa E, Stocchi F, Melamed E, Eyal E, Rascol O. A randomized, double-blind, placebo-controlled, delayed start study to assess rasagiline as a disease modifying therapy

in Parkinson's disease (the ADAGIO study): rationale, design, and baseline characteristics. *Movement Disorders* 2008; **23**(15):2194–2201.
6. O'Brien L, Fitzmaurice G. Analysis of longitudinal multiple-source binary data using generalized estimating equations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2004; **53**(1):177–193.
7. Bandyopadhyay S, Ganguli B, Chatterjee A. A review of multivariate longitudinal data analysis. *Statistical Methods in Medical Research* 2011; **20**(4):299–330.
8. O'Brien P. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984; **40**(4):1079–1087.
9. Huang P, Tilley B, Woolson R, Lipsitz S. Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens–Fisher problem. *Biometrics* 2005; **61**(2):532–539.
10. Huang P, Woolson R, O'Brien P. A rank-based sample size method for multiple outcomes in clinical trials. *Statistics in Medicine* 2008; **27**(16):3084–3104.
11. Tilley B, Marler J, Geller N, Lu M, Legler J, Brott T, Lyden P, Grotta J. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Nurological Disorders and stroke t-PA stroke trial. *Stroke* 1996; **27**(11):2136–2142.
12. Tilley B, Pillemer S, Heyse S, Li S, Clegg D, Alarćon G, *et al.* Global statistical tests for comparing multiple outcomes in rheumatoid arthritis trials. *Arthritis & Rheumatism* 1999; **42**(9):1879–1888.
13. Huang P, Woolson R, Granholm A. The use of a global statistical approach for the design and data analysis of clinical trials with multiple primary outcomes. *Experimental Stroke* 2009; **1**:100–109.
14. Huang P, Goetz C, Woolson R, Tilley B, Kerr D, Palesch Y, Elm J, Ravina B, Bergmann K, Kieburtz K, *et al.* Using global statistical tests in long-term Parkinson's disease clinical trials. *Movement Disorders* 2009; **24**(12):1732–1739.
15. Glas C, Geerlings H, van de Laar M, Taal E. Analysis of longitudinal randomized clinical trials using item response models. *Contemporary Clinical Trials* 2009; **30**(2):158–170.
16. Adams R, 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.
17. Anderson T. *An Introduction to Multivariate Statistical Analysis*, 3rd ed. John Wiley & Sons: Hoboken, New Jersey, 2003.
18. Andersen E. Latent regression analysis based on the rating scale model. *Psychology Science* 2004; **46**:209–226.
19. Christensen K, Bjorner J, Kreiner S, Petersen J. Latent regression in loglinear Rasch models. *Communications in Statistics-Theory and Methods* 2004; **33**(6):1295–1313.
20. Mislevy R. Estimation of latent group effects. *Journal of the American Statistical Association* 1985; **80**:993–997.
21. Maier K. A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics* 2001; **26**(3):307–330.
22. Kamata A. Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement* 2001; **38**(1):79–93.
23. Fox J. Multilevel IRT using dichotomous and polytomous response data. *British Journal of Mathematical and Statistical Psychology* 2005; **58**(1):145–172.
24. Fox J, Glas C. Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika* 2001; **66**(2):271–288.
25. Fox J. Applications of multilevel IRT modeling. *School Effectiveness and School Improvement* 2004; **15**(3-4):261–280.
26. Fox J. Multilevel IRT modeling in practice with the package mlirt. *Journal of Statistical Software* 2007; **20**(5):1–16.
27. Fox J. *Bayesian Item Response Modeling: Theory and Applications*. Springer Verlag: New York, New York, 2010.
28. Natesan P, Limbers C, Varni J. Bayesian estimation of graded response multilevel models using Gibbs sampling: formulation and illustration. *Educational and Psychological Measurement* 2010; **70**(3):420–439.
29. Hung L, Wang W. The generalized multilevel facets model for longitudinal data. *Journal of Educational and Behavioral Statistics* 2012; **37**:231–255.
30. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. CRC Press: Boca Raton, FL, 2004.
31. Skrondal A, Rabe-Hesketh S. Latent variable modelling: a survey. *Scandinavian Journal of Statistics* 2007; **34**(4):712–745.
32. Faucett C, Thomas D. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**(15):1663–1685.
33. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**(4):465–480.
34. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.
35. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 1997; **16**:239–257.
36. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* 1997; **16**:259–272.
37. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**(3):809–834.
38. Yu M, Law NJ, Taylor JM, Sandler HM. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* 2004; **14**(3):835–862.
39. 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.
40. 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.
41. Tseng Y, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. *Biometrika* 2005; **92**(3):587–603.
42. Schwab R, England A. Projection technique for evaluating surgery in Parkinson's disease. *Third Symposium on Parkinson's Disease*, Edinburgh: Livingstone, 1969; 152–157.

43. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer: New York, New York, 2000.
44. Curtis S. BUGS code for item response theory. *Journal of Statistical Software* 2010; **36**:1–34.
45. Tong P, Coombes K. integIRTy: a method to identify genes altered in cancer by accounting for multiple mechanisms of regulation using item response theory. *Bioinformatics* 2012; **28**(22):2861–2869.
46. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis*. CRC press: Boca Raton, FL, 2004.
47. Spiegelhalter D, Best N, Carlin B, 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. DOI: 10.1111/1467-9868.00353.
48. Carlin B, Louis T. *Bayesian methods for data analysis*. Chapman & Hall/CRC: Boca Raton, FL, 2009.
49. Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH. Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study. *Statistics in Medicine* 2001; **20**(24):3715–3728.
50. Cleveland W. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74**:829–836.
51. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer Verlag: New York, New York, 2005.
52. Zhang S, Müller P, Do KA. A Bayesian semiparametric survival model with longitudinal markers. *Biometrics* 2010; **66**(2):435–443.
53. Sousa I. A review on joint modeling of longitudinal measurements and time-to-event. *REVSTAT-Statistical Journal* 2011; **9**:57–81.
54. Lin H, Turnbull BW, McCulloch CE, Slate EH. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 2002; **97**(457):53–65.
55. Proust-Lima C, Joly P, Dartigues JF, Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics & Data Analysis* 2009; **53**(4):1142–1154.
56. Dantan E, Proust-Lima C, Letenneur L, Jacqmin-Gadda H. Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics* 2008; **4**:1–26.
57. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research* 2012.
58. Ruppert D, Wand M, Carroll R. *Semiparametric Regression*, Vol. 12. Cambridge University Press: Cambridge, UK, 2003.

