Published in final edited form as:
Stat Med. 2010 February 28; 29(5): 546–557. doi:10.1002/sim.3798.

Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial

Ning Li[1,*], **Robert M. Elashoff**[2], **Gang Li**[2], and **Jeffrey Saver**[3]

[1]Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, Florida, USA

[2]Department of Biostatistics, School of Public Health, University of California at Los Angeles, Los Angeles, California, USA

[3]Department of Neurology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California, USA

SUMMARY

Existing joint models for longitudinal and survival data are not applicable for longitudinal ordinal outcomes with possible non-ignorable missing values caused by multiple reasons. We propose a joint model for longitudinal ordinal measurements and competing risks failure time data, in which a partial proportional odds model for the longitudinal ordinal outcome is linked to the event times by latent random variables. At the survival endpoint, our model adopts the competing risks framework to model multiple failure types at the same time. The partial proportional odds model, as an extension of the popular proportional odds model for ordinal outcomes, is more flexible and at the same time provides a tool to test the proportional odds assumption. We use a likelihood approach and derive an EM algorithm to obtain the maximum likelihood estimates of the parameters. We further show that all the parameters at the survival endpoint are identifiable from the data. Our joint model enables one to make inference for both the longitudinal ordinal outcome and the failure times simultaneously. In addition, the inference at the longitudinal endpoint is adjusted for possible non-ignorable missing data caused by the failure times. We apply the method to the NINDS rt-PA stroke trial. Our study considers the modified Rankin Scale only. Other ordinal outcomes in the trial, such as the Barthel and Glasgow scales can be treated in the same way.

1. INTRODUCTION

In clinical trials longitudinal ordinal outcomes are commonly encountered and quite often some observations are missing due to dropout or death. If the probability of dropout or death is related to the unobserved observations, the missing mechanism is often called missing not at random (MNAR) or non-ignorable [1]. One example is the clinical trial of intravenous recombinant tissue-plasminogen activator (rt-PA) in patients with acute stroke [2]. In this study, patients treated with rt-PA were compared with those given placebo to look for an improvement from baseline in the score on the Modified Rankin Scale, an ordinal measure of degree of disability with categories ranging from no symptoms, no significant disability to severe disability or death. During the follow-up patients could dropout, die or experience treatment failure. A treatment failure occurs if the patient remains in severe disability after

*Correspondence to: Ning Li, Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida, P.O. Box 100231, Gainesville, FL 32610-0231. nli@phhp.ufl.edu, Tel: (352) 273-5822

treatment initiation. Both death and dropout could result in non-ignorable missing values in the Modified Rankin Scale because these events are highly related to the disease condition of the patients. The problem is further complicated by the fact that treatment failure, death and dropout are potentially correlated. It is suggested by the clinicians to use treatment failure and death to provide additional information on the treatment efficacy. In this trial we are interested in estimating the treatment effects on both the longitudinal measurements of the Modified Rankin Scale and the risk of treatment failure or death. The estimates need to be adjusted for possible non-ignorable missing data in Modified Rankin Scale and informative censoring of treatment failure or death by dropout.

Non-ignorable missing data problem in longitudinal studies has motivated a growing literature on joint analysis of the repeated measurements and the missing data mechanism. A great body of work exists for normal-distributed longitudinal measurements in the setting of linear mixed effects models or marginal models [3-9]. These were also extended to generalized longitudinal measurements with exponential family distributions [10,11,12]. However, the approaches cannot be used for longitudinal ordinal outcomes which are encountered very often in medical studies. There have been very limited efforts to extend the joint analysis to longitudinal ordinal measurements. Molenberghs, Kenward, and Lesaffre proposed a model for longitudinal ordinal data with nonrandom drop-out, which linked the multivariate Dale model for longitudinal ordinal data to a logistic regression model for drop-out [13]. A pattern-mixture model was developed by Kaciroti et. al to analyze clustered longitudinal ordinal data with non-ignorable missing values [14]. These methods assume finite, discrete missing data patterns and thus are not applicable to the aforementioned NINDS rt-PA stroke trial where the death time is continuous and there are multiple reasons leading to non-ignorable missing data. For the NINDS rt-PA stroke trial, a competing risks framework is essential to distinguish treatment failure/death from dropout because failure or death is an important clinical endpoint to evaluate the treatment efficacy in addition to the longitudinal measurements of Modified Rankin Scale. To the best of our knowledge, we are the first to consider competing risks failure times to deal with possible non-ignorable missing values in the longitudinal ordinal measurements.

In this article we formulate a joint model which consists of the following two components: (1) a partial proportional odds model for the longitudinal ordinal outcome, which extends the model proposed by Peterson and Harrell [15] to correlated ordinal observations. Such extensions have been studied by Hedeker and Mermelstein [16,17]. The partial proportional odds model is built upon the popular proportional odds model for ordinal data [18], but allows non-proportional odds for a subset of the predictors. It is a more flexible approach and at the same time provides a useful tool to test the proportional odds assumption. (2) a cause-specific hazards model for the competing risks failure times data to allow for multiple risks at the survival endpoint [19], in which we incorporate a frailty to take into account correlations between the failure times. We further show that the frailty can be identified from the data. The two sub-models are associated through the joint distribution of the random effects in (1) and (2) so that the event time processes (e.g., missing data mechanism) can depend on both observed and missing measurements in the longitudinal endpoint. Our joint model not only enables one to make inference for both the longitudinal ordinal outcome and the failure times simultaneously, but also adjusts estimated quantities of the longitudinal measurements for possible non-ignorable missing data caused by the failure times. Our model further extends the previous methods in that it considers multiple failure types with potential correlations at the survival endpoint.

This paper is organized as follows. Section 2 describes the joint model and its likelihood function, and further shows that all the components at the survival endpoint, especially the frailty, are identifiable. Section 3 proposes an EM algorithm for the maximum likelihood

estimates of the joint model and a profile likelihood approach for standard error estimation. Section 4 contains an application of the method to the the NINDS rt-PA stroke trial. Some simulation studies are provided in Section 5. The final section contains a discussion.

2. THE JOINT MODEL AND THE LIKELIHOOD FUNCTION

Our joint model consists of two linked sub-models: (1) a partial proportional odds model for the longitudinal ordinal repeated measurements; (2) a cause-specific hazards model for the competing risks failure time data. Sub-model (1) is an extension of the partial proportional odds model proposed by Peterson and Harrell [15] to allow for multiple observations on each study subject by incorporating subject-specific random effects. If we have n subjects under study, each with n_i observations, $i = 1, \ldots, n$, let Y_{ij} denote the jth response for subject i, where Y_{ij} takes values in $\{1, \ldots, K\}$ for some integer $K \geq 2$, X_{ij} a $p \times 1$ vector of predictors, $\tilde{X}_{ij}$ a $s \times 1$ vector, $s \leq p$, containing a subset of the p predictors for which the proportional odds assumption may not be satisfied, and W_{ij} a $q \times 1$ vector of predictors for the random effects. The partial proportional odds model for Y_{ij} is written as:

$$P\left(Y_{ij} \leq k | X_{ij}, \tilde{X}_{ij}, W_{ij}, \theta, \beta, \alpha, b_i\right) = \frac{1}{1+\exp\left(-\theta_k - X_{ij}^T\beta - \tilde{X}_{ij}^T\alpha_k - W_{ij}^T b_i\right)} \quad (1)$$

for $k = 1, \ldots, K - 1$, where $\theta = (\theta_1, \ldots, \theta_{K-1})^T$ with $\theta_1 < \theta_2 < \cdots \theta_{K-1}$, $\beta = \beta_1, \ldots, \beta_p)^T$ are fixed effects of X_{ij}, $\alpha_k = \alpha_{k1}, \ldots, \alpha_{ks})^T$ is a $s \times 1$ vector of regression coefficients and $\alpha_1 = 0$, so that $\tilde{X}_{ij}^T\alpha_k$ is an increment associated with the logit of probability $Y_{ij} \leq k$ comparing to that of $Y_{ij} \leq 1$, and $b_i \sim N_q(0, \Sigma_b)$ is a vector of random effects for subject i. Let the vector $\alpha = \left(\alpha_2^T, \ldots, \alpha_{K-1}^T\right)^T$.

We assume a proportional cause-specific hazards sub-model for the competing risks failure time data. Let $Z_i(t)$ denote the associated $l \times 1$ vector of time-dependent predictors and $C_i = (T_i, D_i)$ denote the survival data on subject i, where T_i is the failure time or censoring time, and D_i takes value from $\{0, 1, \ldots, g\}$, with $D_i = 0$ indicating a censored event and $D_i = d$ showing that subject i fails from the dth type of failure, where $d = 1, \ldots, g$. The sub-model for C_i is specified as

$$\begin{aligned}\lambda_d(t; Z_i(t), u_i, \gamma, \nu) &= \lim_{h\to 0} h^{-1} P(t \leq T_i < t+h, D_i = d | T_i \geq t, Z_i(t), u_i, \gamma, \nu) \\ &= \lambda_{0d}(t) \exp\left(Z_i(t)^T\gamma_d + \nu_d u_i\right)\end{aligned} \quad (2)$$

for $d = 1, \ldots, g$, where $\lambda_d(t; Z_i(t), u_i, \gamma, \nu)$ is the instantaneous failure rate due to type d at time t given $Z_i(t)$ and the frailty u_i and in the presence of all other failure types, $\lambda_{0d}(t)$ is a completely unspecified baseline hazard function for risk d, $\gamma = \left(\gamma_1^T, \ldots, \gamma_g^T\right)^T$ is a vector of fixed unknown regression coefficients, and $\nu = (\nu_1, \ldots, \nu_g)^T$ collects the coefficients of the frailty u_i for the g competing risks. This model is an extension of the cause-specific hazards model for competing risks survival data [19] by including subject-specific random effects u_i. The random effects u_i can be interpreted as unobservable traits that are shared by all the g event processes on the same subject and induce correlations among different failure types. Note that we do not assume the latent failure times are independent conditional on u_i and the covariates, and allow existence of other sources of correlations among the failure times that

are not accounted for in (2). Throughout, the censoring mechanism is assumed to be independent of the survival time. Dependent (or informative) censoring can be treated as one of the g types of failures. The association between Y and C is modeled by the assumption that the random effects u_i and b_i jointly have a multivariate normal distribution:

$$a_i = \begin{pmatrix} b_i \\ u_i \end{pmatrix} \sim N_{(q+1)} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_b & \Sigma_{bu}^T \\ \Sigma_{bu} & \sigma_u^2 \end{pmatrix} \right).$$

The parameter v_1 is set to 1 to ensure identifiability. A Wald's test can be used to test the null hypothesis H_0: $\Sigma_{bu} = 0$ for the association between Y and C. It is easily seen that the joint model reduces to separate analysis of the two endpoints if $\Sigma_{bu} = 0$. The correlation between b_i and u_i can also be derived from Σ_{bu} and this, together with the magnitude of v, can be used used to determine the strength of association between Y and C. We further assume Y and C are independent given the latent random effects a_i and the covariates. Note that our joint model allows measurements in Y after event times, which is necessary in the rt-PA stroke trial since the Modified Rankin scale can be observed after treatment failure.

For competing risks failure data, it is well known that the distribution of (T, D) is the identified minimum and that the joint distribution of the underlying failure times is not identifiable from the data [20]. Under the assumptions that variation with observed regressors $\{exp(Z(t)^T \gamma_d), d = 1, \ldots, g\}$ contains a non-empty open set in $\mathbb{R}^g$ and that the expectation of the frailty term $exp(u)$ is finite, Abbring and van den Berg proved that the parameters of a mixed proportional cause-specific hazards model are identifiable based on competing risks survival data [21]. Their arguments can be applied to establish the identiability of the parameters in our Model (2). They also established the identifiability of the joint distribution of the latent failure times by further assuming independence between the latent failure times conditional on the covariates and random effects. In the paper we only need to be concerned with identifying the parameters of the mixed proportional cause-specific hazards model, rather than the joint distribution of the latent failure times, from the observed competing risks survival data. Therefore, we do not require the independence assumption between the latent failure times.

Let $\Psi = (\theta, \beta, \alpha, \gamma, v, \Sigma, \lambda_{01}(t), \ldots, \lambda_{0g}(t))$ collects all the parameters in (1) and (2), where Σ is the variance-covariance matrix of a_i. We assume that the missing values in the longitudinal measurements caused by reasons other than the events are missing at random.

For the notation, we write $Y_i = (Y_{i1}, \ldots, Y_{in_i})^T$, $Y = \left(Y_1^T, \ldots, Y_n^T\right)^T$, and $C = (C_1, D_1, \ldots, C_n, D_n)^T$. Let $\pi_{ij}(k)$ stand for the probability that $Y_{ij} \leq k$ given the covariates and the random effects, and thus $\pi_{ij}(K) = 1$ and $\pi_{ij}(0) = 0$ for all i and j. The observed-data likelihood function for Ψ is therefore

$$
\begin{aligned}
L(\Psi;Y,C) &\propto \prod_{i=1}^{n} f(Y_i, C_i|\Psi) \\
&= \prod_{i=1}^{n} \int_a f(Y_i|C_i, a, \Psi)\, f(C_i|a, \Psi)\, f(a|\Psi)\, da \\
&= \prod_{i=1}^{n} \int_a f(Y_i|a, \Psi)\, f(C_i|a, \Psi)\, f(a|\Psi)\, da \\
&= \prod_{i=1}^{n} \int_a \left[\prod_{j=1}^{n_i} \prod_{k=1}^{K} \left\{ \pi_{ij}(k) - \pi_{ij}(k-1) \right\}^{I(Y_{ij}=k)} \right] \\
&\qquad \left\{ \prod_{d=1}^{g} \lambda_d(T_i; Z_i(T_i), u, \gamma_d, \nu_d)^{I(D_i=d)} \right\} \\
&\qquad \times exp\left[-\int_0^{T_i} \left\{ \sum_{d=1}^{g} \lambda_d(t; Z_i(t), u, \gamma_k, \nu_k) \right\} dt \right] \\
&\qquad \times \frac{1}{\sqrt{(2\pi)^{q+1}|\Sigma|}} exp\left(-\tfrac{1}{2} a^T \Sigma^{-1} a \right) da.
\end{aligned} \tag{3}
$$

Here we rely on the assumption that Y_i and C_i are independent conditional on the covariates and the random effects.

3. ESTIMATION AND INFERENCE

The observed-data likelihood is difficult to maximize directly because of integration with respect to the latent random effects a_i. The procedure can be simplified using the complete-data likelihood conditional on the random effects:

$$
L(\Psi;Y,C,a) \propto \prod_{i=1}^{n} \left[\prod_{j=1}^{n_i} \prod_{k=1}^{K} \left\{ \pi_{ij}(k) - \pi_{ij}(k-1) \right\}^{I(Y_{ij}=k)} \right]
$$

$$
\begin{aligned}
&\left\{ \prod_{d=1}^{g} \lambda_d(T_i; Z_i(T_i), u_i, \gamma_d, \nu_d)^{I(D_i=d)} \right\} \\
&\times exp\left[-\int_0^{T_i} \left\{ \sum_{d=1}^{g} \lambda_d(t; Z_i(t), u_i, \gamma_k, \nu_k) \right\} dt \right] \\
&\times \frac{1}{\sqrt{(2\pi)^{q+1}|\Sigma|}} exp\left(-\tfrac{1}{2} a_i^T \Sigma^{-1} a_i \right).
\end{aligned} \tag{4}
$$

The maximum likelihood estimates of Ψ can be obtained by an EM-algorithm which iterates between an E-step in which the expected logarithm of the complete-data likelihood (4) is computed conditional on the observed data and the current estimates of the parameters, and an M-step in which the new parameter estimates are calculated by maximizing this expected log-likelihood. The cumulative hazards of the baseline functions in λ_d are chosen to be step functions with jumps at observed event times. We need to solve score equations in the maximization step. There are no closed-form solutions for θ, β, α, γ, and ν, for which we use a one-step Newton-Raphson algorithm. These parameter estimates in the M-step depend on the conditional expectations of functions of a_i, which are evaluated in the E-step in each iteration. The algorithm iterates between the E-step and the M-step until the estimates converge. Please refer to the Appendix for more detail.

The dimension of our maximum likelihood estimates of Ψ increases with the sample size due to the non-parametric feature of the baseline hazard function λ_{0d}, which motivates a profile likelihood approach for the standard error estimates of the parametric components θ,

β, α, γ, ν, and Σ, in which the baseline hazards functions have been profiled. We propose to approximate its variance-covariance matrix of $\Omega = (\theta, \beta, \alpha, \gamma, \nu, \Sigma)$ by inverting the empirical Fisher information obtained from the profile likelihood. Let $l^{(i)}(\hat{\Omega}; Y, C)$ denote the observed score vector from the profile likelihood on the ith subject evaluated at $\hat{\Omega}$. The observed information matrix of Ω can be approximated by $\sum_{i=1}^{n} l^{(i)}\left(\widehat{\Omega}; Y, C\right) l^{(i)}\left(\widehat{\Omega}; Y, C\right)^{T}$.

4. ANALYSIS OF THE NINDS rt-PA TRIAL

The NINDS rt-PA trial of intravenous recombinant tissue plasminogen activator (rt-PA) in patients with acute ischemic stroke compares rt-PA with placebo using a randomized double-blind design. A total of 624 patients entered the study and were randomized to one of the two groups of 312 patients each. Among other measures of efficacy, the modified Rankin scale was recorded at baseline, 7–10 days, 3 months, 6 months, and 12 months post stroke onset. The measure is in an ordinal scale and some of the categories were pooled using the following: 1 = no symptoms or no significant disability despite symptoms, 2 = slight disability, 3 = moderate disability or moderately severe disability, 4 = severe disability or dead. Although death is in one of the levels, we do not impute missing data after death in the modified rankin scale, but take care of the time to death in the survival endpoint. Out of the 624 patients, 25 dropped out before 12 months (14 in rt-PA group and 11 in the placebo) and 168 died (78 in rt-PA group and 90 in the placebo group). A treatment failure occurs if the patient remains in severe disability in two consecutive observations after randomization. We observed 54 treatment failures, of which 17 died later. The average number of visits is 4.25, and the percent of missing data in the modified Rankin Scale at 12 months is 30%. The missing data after death or dropout could be non-ignorable since patients with a higher Rankin score would be more likely to die or drop out of the study because of low efficacy of the treatment.

In this example we illustrate the application of the joint model using a subset of the patients whose disease subtypes are small vessel occlusive disease, large vessel atherosclerosis / cardioembolic stroke, or unknown reasons. The following covariates were considered in modeling the longitudinal ordinal modified Rankin scale post stroke onset: treatment group (rt-PA or placebo), the three subtypes of acute stroke (small vessel occlusive disease, large vessel atherosclerosis or cardioembolic stroke, and unknown reasons), modified Rankin scale prior stroke onset (based on the original definition without collapsing categories), and time since randomization. We adopt unstructured time trend using three dummy variables, time3, time6, time12, for 3, 6, and 12 months respectively, so that the measure at 7 – 10 days serves as reference. There are 587 patients included in the analysis and their baseline characteristics and changes in modified rankin scale over time are summarized in Table 1. The disease subtypes and the baseline modified rankin scale are distributed evenly between the groups, but the rt-PA group has significantly lower modified rankin scale after treatment initiation. For both groups, the modified rankin scale decreased over time. The Kendall's tau correlations among the modified rankin scale at 7–10 days, 3, 6, and 12 months are in between 0.65 – 0.87 (p-values < 0.0001).

In the joint model, we are interested in modeling two competing risks at the survival endpoint, the time to dropout (risk 1) and the time to death or remaining in severe disability (risk 2). We combine death and remaining in severe disability in one risk because both of the events are strong evidence of low treatment efficacy. One dummy variable "group" is created treating the placebo treatment arm as the reference group, and another two dummy variables "small vessel" and "large vessel or cardioembolic stroke" are generated to represent the two blocks, treating the block "unknown reasons" as the reference. We carried out the likelihood ratio test to assess the fit of the proportional odds assumption by

expanding the vector $\tilde{X}$ and testing $\alpha = 0$, and identified divergence of the block effects from the proportional odds assumption.

The results of the joint model are shown in Table 2 where at the longitudinal endpoint we have $\tilde{X}$ =(small vessel, large vessel or cardioembolic stroke)^T. Since we could not identify significant interaction effects between group and the time trend, these terms were not considered in our final model. As shown in Table 2, there are significant effects of treatment, modified Rankin scale prior onset, time3, time6, time12, and the interaction between large vessel or cardioembolic stroke and the treatment. The Rankin scale has a decreasing trend over time, given that conditional on other covariates and the random effects, the cumulative odds ratio for $Y \leq k$, $k = 1, 2, 3$, is exp(2.12) = 8.33 at 3 months compared to 7–10 days post stroke onset (the 95% confidence interval is (5.63, 12.33)), and is 9.68 and 11.59 at 6 months and 12 months, with the 95% confidence intervals (6.54, 14.32) and (7.53, 17.84), respectively. The block effects do not satisfy the proportional odds assumption. Compared to the patients with unknown reasons (as stated in the database), the small vessel patients have lower Rankin scales, and the conditional cumulative odds ratio is exp(3.49) for $Y \leq 1$, and is exp(3.77) and exp(6.14) for $Y \leq 2$ and $Y \leq 3$. Note that the estimate 6.14 may not be reliable due to the fact that there are only 4 patients with $Y = 4$ in the stratum small vessel. The patients with large vessel and cardioembolic stroke tend to have higher Rankin scales than the patients with unknown reasons. The treatment is not as effective for the large vessel or cardioembolic stroke patients as for the patients with unknown reasons, and there is no significant difference in the treatment effects between the patients with small vessel and those with unknown reasons. For the patients with unknown reasons, the conditional cumulative odds ratio is exp(1.48) = 4.39 for $Y \leq 1, 2$, and 3 comparing rt-PA group to the placebo group (the 95% confidence interval (2.30, 8.39)). In contrast, in the large vessel or cardioembolic stroke patients, the conditional cumulative odds ratio is exp(1.48–2.27) = 0.45 (the 95% confidence interval (0.09, 2.31)).

We are not able to observe significant treatment effects at the survival endpoint for either the time to dropout or the time to death or remaining in severe disability. There appears to be a higher risk of death or remaining in severe disability in the patients with a higher prior onset Rankin scale. On the other hand, the patients with small vessel tend to have a lower risk for this event than those with unknown reasons. The estimate of v_2 is positive, suggesting that the two risks are positively correlated, i.e., the patients with a higher risk of dropout are more likely to experience death or remaining in severe disability. There is a negative correlation ($\rho_{bu} < 0$) between the random intercept b_i in sub-model (1) and the frailty u_i in sub-model (2), which indicates that patients with higher Rankin scales tend to have a higher risk of dropout, death or remaining in severe disability. We also carried out separate analysis of the longitudinal Rankin scale measurements using either Wilcoxon two sample test or a partial proportional odds model assuming ignorable missing data mechanism. Significant treatment effects were found in both methods and the partial proportional odds model produced similar estimates as our joint model. However, this is not always the case in the presence of non-ignorable missing data. In the next section we show using simulation studies that the separate analysis of the longitudinal ordinal data could give rise to biased estimates and poor inference when there are non-ignorable missing data.

5. SIMULATION STUDIES

Tables 3 and 4 summarize the simulation results on 200 Monte Carlo samples with the sample size $n = 200$ and 500, respectively. We generated data from Model (1)-(2) with $K = 3$. For each simulated dataset, we apply both the joint model and the separate analysis of the longitudinal outcome and the competing risks survival times using Model (1) and (2), respectively. The covariate vector $X_{ij} = (t_{ij}, x_i, t_{ij}x_i)^T$, where $t_{ij} = 0, 0.5, \ldots,$ up to 4 (may be

censored by the failure times) is the visit time, $x_i \sim Bernoulli(0.5)$ is the treatment group indicator, and $t_{ij}x_i$ is the interaction between the two. We further set $\tilde{X}_{ij}=x_i$ and $W_{ij} = 1$, so that b_i is the random intercept for subject i and its variance is σ_b^2. The true values of the parameters β, α, θ and σ_b^2 are given in Tables 3 and 4. We simulated two competing risks with $\lambda_{01} = 0.15$, $\lambda_{02} = 0.25$, $Z_i(t) = (z_i, x_i)^T$ with $z_i \sim N(2, 1)$, and $u_i \tilde{} N\left(0, \sigma_u^2\right)$. The random intercept b_i in Model (1) and u_i have a bivariate normal distribution with correlation ρ_{bu}. The censoring time τ_i for subject i was generated from an exponential distribution with mean 10. We could only observe one failure type on each study subject, depending on which happens first. Furthermore, censoring could occur if τ_i is smaller than both failure times. In our simulation the rate of risk 1 is around 44%, risk 2 is around 37%, and the censoring rate is around 19%. The simulated bias, standard error and the 95% confidence interval coverage probability (CP) are given in Tables 3 and 4.

Compared to the joint model, the separate analysis produces relatively large bias in the time trend β_1 and the interaction with the treatment β_3. With the negative correlation between b_i and u_i ($\rho_{bu} < 0$)and the positive correlation between the two failure times ($v_2 > 0$), the subject with a higher ordinal outcome tends to have a higher risk of experiencing both failures and thus leave the study early, so that that the observed time trend is under-estimated (note that we model the probability of $Y_{ij} \leq k$), which results in a low confidence interval coverage for β_1. Because the treatment lowers the ordinal outcome, there would be unbalanced event rates between the two groups, the estimated difference in the time trend (β_3) is also biased. These biases will not vanish as we increase the sample size to 500 and the CP for β_1 is even poorer. The separate analysis of the competing risks data also underestimates v_2 when $n = 200$ and produces larger empirical standard errors for γ. In the joint analysis the missing data mechanism has been modeled together with the longitudinal measurements so that we are able to obtain almost unbiased estimates of β. Furthermore, by combining information from the longitudinal endpoint, it is more efficient in estimating γ and v_2. Overall the joint model performs better asymptotically ($n = 500$) with smaller mean square errors for all the parameters. At last, we observe that estimation of σ_u^2 requires a relatively large sample size in both the joint model and the separate analysis.

To further compare the performances of the joint model and the separate analysis under a more general senario, we conducted a second set of simulations in which the random effects a_i were generated from a multivariate t-distribution with degrees of freedom $d.f. = 5$, but the data were analyzed on the basis of the assumptions specified in Models (1)-(2). We know that the t-distribution has longer tails than the normal distribution, and the latter is included as a special case as the $d.f.$ goes to infinity. The results of the simulations are given in Tables 5 and 6. We do not show the estimates for the random components since the estimates and the true parameter values are no longer comparable under model misspecification. Similar to the results in Tables 3 and 4, bias in the estimates of β_1, β_3 and v_2 is identified in the separate analysis, but now v_2 tends to be over-estimated, and its bias does not vanish as the sample size increases to 500. The joint model in general produces more accurate point estimates than the separate analysis. The standard error estimation methods in both approaches are not robust to model misspecification as some parameters show poor confidence interval coverage probabilities, especially as the sample size gets large. The separate analysis tends to have larger variances (or SE) in the parameter estimates for the fixed effects at the longitudinal endpoint than that in Table 3. However, the impact of the model misspecification on the variances of the estimates in the joint model is minimal. Comparisons of the mean square errors between the two approaches again suggest that the joint model performs superior to the separate analysis.

6. DISCUSSION

The proposed model extends existing methods to handle longitudinal ordinal data with possible non-ignorable data using a partial proportional odds model, adopts competing risks framework for the missing data mechanism, and therefore is more general in terms of distinguishing different events that cause missing data in the study. On the basis of the arguments given in Abbring and van den Berg [21], it is easy to show that all the parameters at the survival endpoint are identifiable. Our joint model enables one to make inferences for both endpoints simultaneously, while at the same time adjusting estimated quantities of the longitudinal measurements for possible non-ignorable missing data caused by the failures. Using simulations we show that the joint analysis performs better than the separate analysis, even under model misspecification where the underlying distribution of the random effects a_i has longer than normal tails. Employment of the partial proportional odds model also enables us to test the fit of the proportional odds assumption for the ordinal measures. If the sample size permits, one could start with the full partial proportional odds model by setting $\tilde{X}_{ij}=X_{ij}$ and backward eliminate the non-significant covariates from $\tilde{X}_{ij}$. In our joint analysis settings, the correlations among the longitudinal ordinal data are modeled through the random effects, which makes it difficult to obtain the fitted correlations as one of the outputs of the model fitting process. If it is of interest to the investigator, marginal models for multivariate ordinal data could be used instead. Because our joint model involves infinite dimensional parameters in the baseline hazard functions, a rigorous treatment of the asymptotic properties of the maximum likelihood estimates warrants future research.

Model selection, as in any regression setting, is an important problem in joint analysis. However, this issue has not been fully addressed in the joint modeling literature. In the application to the stroke study, we use likelihood ratio test to assess the fit of proportional odds assumption by expanding the covariate vector $\tilde{X}$. This problem can not be easily tackled by the popular model selection criteria, such as the Akaike information criterion (AIC), since it is difficult in the presence of nonparametric baseline functions of the cause-specific hazards. In particular, in semiparametric models like our joint model, the number of nuisance parameters in the baseline hazard functions increases with the sample size. The partial likelihood approach for Cox models is also inapplicable due to the correlation with longitudinal measurements introduced by the frailty term. Some authors have presented profile likelihood methods for model selection in the context of frailty models [22], which can be extended to the joint model. It is possible to extend their work to joint models. Future research in this direction is warranted.

Acknowledgments

The publicly available dataset for the NINDS rt-PA Stroke Trials was downloaded through the National Technical Information Service website. We are grateful to NIH and the NINDS rt-PA Study Group for making this dataset available as a public resource.

Appendix. The EM Algorithm

E-step

In the E-step of the $(m + 1)$th iteration, conditional on the observed data and the parameter estimates from the mth iteration, we evaluate

$$\begin{aligned} E_{a_i|Y_i,C_i,\Psi^{(m)}}(h(a_i)) &= \int h(a_i) f\left(a_i|Y_i, C_i, \Psi^{(m)}\right) da_i \\ &= \frac{\int h(a_i) f\left(Y_i, C_i, a_i|\Psi^{(m)}\right) da_i}{f\left(Y_i, C_i|\Psi^{(m)}\right)} \\ &= \frac{\int h(a_i) f\left(Y_i|a_i, \Psi^{(m)}\right) f\left(C_i|a_i, \Psi^{(m)}\right) f\left(a_i|\Psi^{(m)}\right) da_i}{\int f\left(Y_i|a_i, \Psi^{(m)}\right) f\left(C_i|a_i, \Psi^{(m)}\right) f\left(a_i|\Psi^{(m)}\right) da_i}. \end{aligned} \tag{5}$$

The integrals can be evaluated using Gaussian-Hermite quadrature.

M-step

Use E to stand for $E_{\alpha_i|Y_i, C_i, \Psi(m)}$. We have, for Σ,

$$\Sigma_b^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n} E\left(b_i b_i^T\right), \tag{6}$$

$$\sigma_u^{2(m+1)} = \frac{1}{n}\sum_{i=1}^{n} E\left(u_i^2\right), \tag{7}$$

and

$$\Sigma_{bu}^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n} E\left(b_i u_i\right). \tag{8}$$

Suppose there are q_d distinct failure times due to the dth cause and write $t_{d1} \le \ldots \le t_{dqd}$ for $d = 1, \ldots, g$. Let $R(t_{dj})$ be the risk set at time t_{dj}, and n_{dj} be the number of failures due to cause d at time t_{dj}. The cumulative baseline hazard function for cause d is

$$\begin{aligned} H_{0d}^{(m+1)}\left(t_{dq}\right) &= \sum_{j=1}^{q} \lambda_{0d}^{(m+1)}\left(t_{dj}\right) \\ &= \sum_{j=1}^{q} \frac{n_{dj}}{\sum_{r\in R\left(t_{dj}\right)} \exp\left(Z_r\left(t_{dj}\right)^T \gamma_d^{(m)}\right) E\left(\exp\left(v_d^{(m)} u_r\right)\right)}. \end{aligned} \tag{9}$$

No closed-form solutions exist for θ, β, α, γ, and v, which are updated by a one-step Newton-Raphson algorithm in each iteration:

$$\theta_k^{(m+1)} = \theta_k^{(m)} - S_{\theta_k}^{(m)} / I_{\theta_k}^{(m)} \tag{10}$$

where $k = 1, \ldots, K - 1$, with $I_{\theta_k}^{(m)}$ and $S_{\theta_k}^{(m)}$ being

$$I_{\theta_k}^{(m)}=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij}=k\right)E\left\{\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)\left(1-2\pi_{ij}(k)\right)}{\pi_{ij}(k)-\pi_{ij}(k-1)}-\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)^2}{\left(\pi_{ij}(k)-\pi_{ij}(k-1)\right)^2}\right\}-I\left(Y_{ij}=k+1\right)E\left\{\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)\left(1-2\pi_{ij}(k)\right)}{\pi_{ij}(k+1)-\pi_{ij}(k)}+\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)^2}{\left(\pi_{ij}(k+1)-\pi_{ij}(k)\right)^2}\right\}\right], \tag{11}$$

$$S_{\theta_k}^{(m)}=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij}=k\right)E\left\{\frac{\pi_{ij}(k)-\pi_{ij}^2(k)}{\pi_{ij}(k)-\pi_{ij}(k-1)}\right\}-I\left(Y_{ij}=k+1\right)E\left\{\frac{\pi_{ij}(k)-\pi_{ij}^2(k)}{\pi_{ij}(k+1)-\pi_{ij}(k)}\right\}\right], \tag{12}$$

$$\beta^{(m+1)}=\beta^{(m)}+I_{\beta}^{(m)-1}S_{\beta}^{(m)}, \tag{13}$$

with $I_{\beta}^{(m)}$ and $S_{\beta}^{(m)}$ being

$$I_{\beta}^{(m)}=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{K}I\left(Y_{ij}=k\right)\left[E\left\{\pi_{ij}(k)-\pi_{ij}^2(k)\right\}+E\left\{\pi_{ij}(k-1)-\pi_{ij}^2(k-1)\right\}\right]X_{ij}X_{ij}^T, \tag{14}$$

$$S_{\beta}^{(m)}=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{K}I\left(Y_{ij}=k\right)\left[1-E\left\{\pi_{ij}(k)\right\}-E\left\{\pi_{ij}(k-1)\right\}\right]X_{ij}, \tag{15}$$

$$\alpha_k^{(m+1)}=\alpha_k^{(m)}-I_{\alpha_k}^{(m)-1}S_{\alpha_k}^{(m)} \tag{16}$$

where $k = 2, \ldots, K - 1$, with $I_{\alpha_k}^{(m)}$ and $S_{\alpha_k}^{(m)}$ being

$$I_{\alpha_k}^{(m)}=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij}=k\right)E\left\{\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)\left(1-2\pi_{ij}(k)\right)}{\pi_{ij}(k)-\pi_{ij}(k-1)}-\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)^2}{\left(\pi_{ij}(k)-\pi_{ij}(k-1)\right)^2}\right\}\tilde{X}_{ij}\tilde{X}_{ij}^T-I\left(Y_{ij}=k+1\right)E\left\{\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)\left(1-2\pi_{ij}(k)\right)}{\pi_{ij}(k+1)-\pi_{ij}(k)}+\frac{\left(\pi_{ij}(k)-\pi_{ij}^2(k)\right)^2}{\left(\pi_{ij}(k+1)-\pi_{ij}(k)\right)^2}\right\}\tilde{X}_{ij}\tilde{X}_{ij}^T\right], \tag{17}$$

$$S_{\alpha_k}^{(m)} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij}=k\right) E\left\{\frac{\pi_{ij}(k)-\pi_{ij}^2(k)}{\pi_{ij}(k)-\pi_{ij}(k-1)}\right\}\tilde{X}_{ij} - I\left(Y_{ij}=k+1\right) E\left\{\frac{\pi_{ij}(k)-\pi_{ij}^2(k)}{\pi_{ij}(k+1)-\pi_{ij}(k)}\right\}\tilde{X}_{ij}\right], \tag{18}$$

$$\gamma_d^{(m+1)} = \gamma_d^{(m)} + I_{\gamma d}^{(m)-1} S_{\gamma d}^{(m)} \tag{19}$$

where $d = 1, \ldots, g$, with $I_{\gamma_d}^{(m)}$ and $S_{\gamma_d}^{(m)}$ being

$$I_{\gamma d}^{(m)} = \sum_{i=1}^{n}\sum_{t_{dj}\le T_i} \lambda_{0d}^{(m+1)}\left(t_{dj}\right)\exp\left(Z_i\left(t_{dj}\right)^T\gamma_d^{(m)}\right)\times E\left(\exp\left(\nu_d^{(m)}u_i\right)\right) Z_i\left(t_{dj}\right) Z_i\left(t_{dj}\right)^T, \tag{20}$$

$$S_{\gamma d}^{(m)} = \sum_{i=1}^{n}\left\{ I\left(D_i=d\right) Z_i\left(T_i\right) - \sum_{t_{dj}\le T_i} \lambda_{0d}^{(m+1)}\left(t_{dj}\right)\exp\left(Z_i\left(t_{dj}\right)^T\gamma_d^{(m)}\right)\times E\left(\exp\left(\nu_d^{(m)}u_i\right)\right) Z_i\left(t_{dj}\right)\right\}, \tag{21}$$

$$\nu_d^{(m+1)} = \nu_d^{(m)} + S_{\nu_d}^{(m)} / I_{\nu_d}^{(m)} \tag{22}$$

where $d = 2, \ldots, g$, with $I_{\nu_d}^{(m)}$ and $S_{\nu_d}^{(m)}$ being

$$I_{\nu_d}^{(m)} = \sum_{i=1}^{n}\sum_{t_{dj}\le T_i} \lambda_{0d}^{(m+1)}\left(t_{dj}\right)\exp\left(Z_i\left(t_{dj}\right)^T\gamma_d^{(m+1)}\right)\times E\left(u_i^2\exp\left(\nu_d^{(m)}u_i\right)\right), \tag{23}$$

$$S_{\nu_d}^{(m)} = \sum_{i=1}^{n}\left\{ I\left(D_i=d\right) E\left(u_i\right) - \sum_{t_{dj}\le T_i} \lambda_{0d}^{(m+1)}\left(t_{dj}\right)\exp\left(Z_i\left(t_{dj}\right)^T\gamma_d^{(m+1)}\right)\times E\left(u_i\exp\left(\nu_d^{(m)}u_i\right)\right)\right\}. \tag{24}$$

Because the model requires that the elements in θ satisfy $\theta_1 < \theta_2 < \cdots < \theta_{K-1}$, we start the EM algorithm by setting the initial values of θ in the increasing order. In each M-step, we monitor the order of the updated θ and switch the values of some components to maintain the monotonicity. However, in our simulations and the real data analysis we have not encountered situations where we need to switch values.

REFERENCES

1. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2nd edn. Wiley; Hoboken: 2002.

2. The national institute of neurological disorders and stroke rt-PA stroke study group. Tissue plasminogen activator for acute ischemic stroke. The New England Journal of Medicine. 1995; 333:1581–1587. [PubMed: 7477192]

3. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis (with discussion). Applied Statistics. 1994; 43:49–93.

4. Little RJA. A class of pattern-mixture models for normal incomplete data. Biometrika. 1994; 81:471–483.

5. Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics. 2000; 4:465–480. [PubMed: 12933568]

6. Hogan JW, Daniels MJ. A hierarchical modelling approach to analysing longitudinal data with drop-out and non-compliance, with application to an equivalence trial in paediatric acquired immune defficiency syndrome. Applied Statistics. 2002; 51:1–21.

7. Zeng D, Cai J. Simultaneous modelling of survival and longitudinal data with an application to repeated quality of life measures. Lifetime Data Analysis. 2005; 11:151–174. [PubMed: 15940822]

8. Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. Statistics in Medicine. 2007; 26:2813–2835. [PubMed: 17124698]

9. Elashoff RM, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics. 2008; 64:762–771. [PubMed: 18162112]

10. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. Biometrics. 1995; 51:151–168. [PubMed: 7766771]

11. Ibrahim JG, Chen M-H, Lipsitz SR. Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. Biometrika. 2001; 88:551–564.

12. Roy J, Daniels MJ. A general class of pattern mixture models for nonignorable dropout with many possible dropout times. Biometrics. 2008; 64:538–545. [PubMed: 17900312]

13. Molenberghs G, Kenward MG, Lesaffre E. The analysis of longitudinal ordinal data with nonrandom drop-out. Biometrika. 1997; 84:33–44.

14. Kaciroti NA, Raghunathan TE, Schork MA, Clark NM, Gong M. A Bayesian approach for clustered longitudinal ordinal outcome with nonignorable missing data: evaluation of an asthma education program. Journal of the American Statistical Association. 2006; 101:435–446.

15. Peterson B, Harrell FE. Partial proportional odds models for ordinal response variables. Applied Statistics. 1990; 39:205–217.

16. Hedeker D, Mermelstein RJ. A multilevel thresholds of change model for analysis of stages of change data. Multivariate Behavioral Research. 1998; 33:427–455.

17. Hedeker D, Mermelstein RJ. Analysis of longitudinal substance use outcomes using ordinal random-effects regression models. Addiction. 2000; 95(Supplement 3):S381–S394. [PubMed: 11132364]

18. McCullagh P. Regression models for ordinal data. Journal of the Royal Statistical Society, Series B. 1980; 42:109–142.

19. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. Biometrics. 1978; 34:541–554. [PubMed: 373811]

20. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences of the United States of America. 1975; 72:20–22. [PubMed: 1054494]

21. Abbring JH, van den Berg GJ. The identifiability of the mixed proportional hazards competing risks model. The Journal of the Royal Statistical Society, Series B. 2003; 65:701–710.

22. Xu, R.; Gamst, A.; Donohue, M.; Vaida, F.; Harrington, DP. Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. Harvard University Biostatistics Working Paper Series. 2006. http://www.bepress.com/harvardbiostat/paper43

Table 1

Baseline characteristics of study subjects and changes in modified rankin scale over time (we show the mean (standard deviation) and the frequency (%) for modified rankin scale and the disease subtypes, respectively)

	rt-PA group (n = 292)	Placebo group (n = 295)	p-value
Disease subtypes			
small vessel	31 (10.62%)	30 (10.17%)	
large vessel or cardioembolic stroke	181 (61.99%)	184 (62.37%)	0.9804[a]
Modified rankin scale			
modified rankin scale prior onset	0.27 (0.73)	0.29 (0.80)	0.9872[b]
modified rankin scale (7–10 days)	2.55 (1.20)	2.90 (1.07)	0.0006[b]
modified rankin scale (3 months)	1.97 (1.07)	2.27 (1.02)	0.0017[b]
modified rankin scale (6 months)	1.91 (1.04)	2.18 (1.04)	0.0060[b]
modified rankin scale (12 months)	1.81 (1.01)	2.13 (1.04)	0.0015[b]

[a] The p-values are calculated using Chi-square test

[b] The p-values are calculated using Wilcoxon rank-sum test

Table 2

Results from the joint analysis for the NINDS study

	Estimate (SE)
Longitudinal outcome	
proportional odds (PO) (*cumulative prob of* $Y \leq k$, $k = 1, 2, 3$)	
group	$1.48^{\dagger}$ (0.33)
modified Rankin scale prior onset	$-1.67^{\dagger}$ (0.27)
time3	$2.12^{\dagger}$ (0.20)
time6	$2.27^{\dagger}$ (0.20)
time12	$2.45^{\dagger}$ (0.22)
small vessel × group	−0.74 (1.26)
large vessel or cardioembolic stroke × group	$-2.27^{\dagger}$ (0.76)
partial PO (cumulative prob of $Y \leq 1$)	
small vessel	$3.49^{\dagger}$ (0.68)
large vessel or cardioembolic stroke	$-1.04^{\dagger}$ (0.44)
partial PO (cumulative prob of $Y \leq 2$)	
small vessel	$3.77^{\dagger}$ (0.68)
large vessel or cardioembolic stroke	$-1.36^{\dagger}$ (0.39)
partial PO (cumulative prob of $Y \leq 3$)	
small vessel	$6.14^{\dagger}$ (1.15)
large vessel or cardioembolic stroke	−0.64 (0.49)
Cause-specific hazards	
Risk 1: dropout	
group	0.23 (0.47)
modified Rankin scale prior onset	0.06 (0.42)
small vessel	0.55 (0.57)
large vessel	−0.29 (0.51)
small vessel × group	0.04 (1.14)
large vessel or cardioembolic stroke	0.30 (1.02)
Risk 2: death or remaining in severe disability	
group	−0.46 (0.27)
modified Rankin scale prior onset	$0.53^{\dagger}$ (0.17)
small vessel	$-2.07^{\dagger}$ (0.79)
large vessel	0.37 (0.27)
small vessel × group	0.35 (1.49)
large vessel or cardioembolic stroke	0.81 (0.54)
Random effects	
σ_b^2	34.66 (3.94)

	Estimate (SE)
σ_u^2	0.51 (0.07)
ρ_{bu}	$-0.997^{\dagger}$ (0.19)
v_2	$3.12^{\dagger}$ (0.51)

†p-value < 0.05

Table 3

Comparison of the joint model and the separate analysis of the longitudinal outcome (n = 200)

		Separate			Joint			
Parameter	**True**	**Bias**	**SE**	**CP**	**Bias**	**SE**	**CP**	MSE_S/MSE_J
Longitudinal								
Fixed effects								
β_1	−1	**0.322**	0.301	**0.795**	−0.020	0.155	0.970	7.954
β_2	1.5	0.030	0.352	0.940	0.014	0.294	0.960	1.441
β_3	0.8	**−0.216**	0.528	0.935	0.039	0.287	0.990	3.879
α_2	0	−0.023	0.285	0.950	−0.008	0.236	0.975	1.466
θ_1	−0.5	0.017	0.161	0.960	−0.010	0.165	0.950	0.959
θ_2	1	0.023	0.178	0.955	0.015	0.153	0.950	1.363
Random effects								
σ_b^2	1	−0.059	0.589	0.900	0.021	0.346	0.955	2.916
Survival								
Fixed effects								
γ_{11}	0.8	−0.015	0.152	0.950	0.062	0.163	0.965	0.767
γ_{12}	−1	0.014	0.308	0.950	−0.079	0.297	0.955	1.006
γ_{21}	0.5	0.009	0.170	0.970	0.041	0.159	0.965	1.075
γ_{22}	−1	−0.024	0.301	0.970	−0.072	0.269	0.975	1.176
Random effects								
v_2	0.5	**−0.192**	0.894	0.985	0.044	0.419	0.960	4.711
σ_u^2	0.5	−0.112	0.418	0.910	0.203	0.473	0.930	0.707
Covariance								
ρ_{bu}	−0.9				0.048	0.127	0.990	

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

Table 4

Comparison of the joint model and the separate analysis of the longitudinal outcome (n = 500)

		Separate			Joint			
Parameter	**True**	**Bias**	**SE**	**CP**	**Bias**	**SE**	**CP**	MSE_S/MSE_J
Longitudinal								
Fixed effects								
β_1	−1	**0.311**	0.189	**0.565**	−0.007	0.101	0.960	12.92
β_2	1.5	−0.005	0.228	0.950	0.014	0.193	0.960	1.389
β_3	0.8	**−0.169**	0.340	0.900	0.012	0.176	0.980	4.632
α_2	0	−0.015	0.169	0.960	−0.014	0.154	0.940	1.204
θ_1	−0.5	0.023	0.100	0.945	−0.009	0.102	0.945	1.004
θ_2	1	0.014	0.112	0.950	0.008	0.106	0.930	1.127
Random effects								
σ_b^2	1	−0.078	0.337	0.920	0.030	0.225	0.960	2.322
Survival								
Fixed effects								
γ_{11}	0.8	0.008	0.137	0.960	0.026	0.109	0.950	1.450
γ_{12}	−1	−0.007	0.224	0.955	−0.038	0.195	0.945	1.273
γ_{21}	0.5	0.004	0.139	0.965	0.022	0.094	0.955	2.075
γ_{22}	−1	−0.015	0.212	0.940	−0.036	0.180	0.955	1.340
Random effects								
v_2	0.5	−0.046	0.859	0.970	0.027	0.268	0.950	10.20
σ_u^2	0.5	−0.082	0.423	0.940	0.125	0.281	0.925	1.963
Covariance								
ρ_{bu}	−0.9				0.061	0.088	0.970	

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

Table 5

Comparison of the joint model and the separate analysis of the longitudinal outcome when the underlying distribution of a_i is multivariate t with $d.f.$ = 5 (n = 200)

		Separate			Joint			
Parameter	**True**	**Bias**	**SE**	**CP**	**Bias**	**SE**	**CP**	MSE_S/MSE_J
Longitudinal								
Fixed effects								
β_1	−1	**0.398**	0.266	**0.740**	−0.023	0.153	0.945	9.573
β_2	1.5	0.036	0.371	0.980	0.064	0.338	0.980	1.174
β_3	0.8	**−0.228**	0.565	0.910	−0.041	0.291	0.960	4.298
α_2	0	−0.035	0.314	0.930	0.024	0.229	0.950	1.883
θ_1	−0.5	0.027	0.178	0.945	−0.029	0.156	0.970	1.287
θ_2	1	0.037	0.195	0.955	−0.022	0.154	0.970	1.628
Survival								
Fixed effects								
γ_{11}	0.8	−0.051	0.190	**0.870**	0.020	0.165	0.980	1.401
γ_{12}	−1	0.079	0.291	0.915	−0.042	0.283	0.970	1.111
γ_{21}	0.5	−0.021	0.171	0.970	0.027	0.150	0.970	1.278
γ_{22}	−1	0.004	0.301	0.975	−0.034	0.300	0.970	0.994
Random effects								
v_2	0.5	**0.114**	0.878	0.955	0.049	0.387	0.960	5.151

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

Table 6

Comparison of the joint model and the separate analysis of the longitudinal outcome when the underlying distribution of a_i is multivariate t with $d.f.$ = 5 (n = 500)

		Separate			Joint			
Parameter	**True**	**Bias**	**SE**	**CP**	**Bias**	**SE**	**CP**	MSE_S/MSE_J
Longitudinal								
Fixed effects								
β_1	−1	**0.400**	0.183	**0.405**	−0.027	0.104	**0.895**	16.76
β_2	1.5	0.029	0.233	0.950	0.029	0.201	0.975	1.337
β_3	0.8	**−0.275**	0.362	**0.825**	−0.065	0.182	**0.895**	5.533
α_2	0	0.005	0.182	0.945	−0.001	0.149	0.960	1.493
θ_1	−0.5	0.037	0.109	0.935	−0.011	0.107	0.930	1.145
θ_2	1	0.036	0.122	0.940	−0.003	0.097	0.950	1.718
Survival								
Fixed effects								
γ_{11}	0.8	−0.031	0.144	**0.800**	0.016	0.109	0.970	1.788
γ_{12}	−1	0.034	0.245	**0.865**	−0.020	0.194	0.945	1.609
γ_{21}	0.5	−0.002	0.138	0.925	0.011	0.094	0.965	2.127
γ_{22}	−1	−0.031	0.234	0.925	−0.025	0.178	0.935	1.725
Random effects								
v_2	0.5	**0.103**	0.909	**0.870**	0.019	0.214	0.950	18.13

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