Computational and Mathematical Methods in Medicine,
Vol. 11, No. 3, September 2010, 281–295

Inferences for joint modelling of repeated ordinal scores and time to event data

Arindom Chakraborty[a] and Kalyan Das[b]*

[a]*Department of Statistics, Visva-Bharati University, Birbhum, West Bengal, India;*
[b]*Department of Statistics, University of Calcutta, 35, B.C. Road, Kolkata 700019, India*

(*Received 19 August 2009; final version received 18 March 2010*)

In clinical trials and other follow-up studies, it is natural that a response variable is repeatedly measured during follow-up and the occurrence of some key event is also monitored. There has been a considerable study on the joint modelling these measures together with information on covariates. But most of the studies are related to continuous outcomes. In many situations instead of observing continuous outcomes, repeated ordinal outcomes are recorded over time. The joint modelling of such serial outcomes and the time to event data then becomes a bit complicated. In this article we have attempted to analyse such models through a latent variable model. In view of the longitudinal variation on the ordinal outcome measure, it is desirable to account for the dependence between ordered categorical responses and survival time for different causes due to unobserved factors. A flexible Monte Carlo EM (MCEM) method based on exact likelihood is proposed that can simultaneously handle the longitudinal ordinal data and also the censored time to event data. A computationally more efficient MCEM method based on approximation of the likelihood is also proposed. The method is applied to a number of ordinal scores and survival data from trials of a treatment for children suffering from Duchenne Muscular Dystrophy. Finally, a simulation study is conducted to examine the finite sample properties of the proposed estimators in the joint model under two different methods.

Keywords: selection model; Monte Carlo EM; frailty survival model; sandwich estimate

AMS Subject Classification: 62J02; 62N01; 62P10

1. Introduction

In recent years, researchers have shown great interest to record the values of key longitudinal covariates until the occurrence of survival time (or event time) of a subject. Certainly, the data available vary across patients according to their failure time. The truncation of a patient's response can raise similar problems to those that have been considered in the event of patient dropout [12,14,18]. However the primary interest in such models is in drawing inferences for the marginal distributions of the longitudinal response in the absence of dropout. Moreover, often the longitudinal model and survival model are assumed to share some unobserved variables. In this case, separate models can result in biased estimates. This necessitates the development of a joint model. Research on this

*Corresponding author. Email: kalyanstat@gmail.com

ISSN 1748-670X print/ISSN 1748-6718 online
© 2010 Taylor & Francis
DOI: 10.1080/17486701003789096
http://www.informaworld.com

topic has received substantial attention during the last few years, primarily because of its tremendous application in biopharmaceutical and related areas.

Considerable literature [5,8,9,19] is available on models called selection model, where the joint density of repeated measures vector and failure time is obtained as the product of the conditional density of the failure time due to any event given the longitudinal outcomes and the marginal density of these outcomes. More specifically, the selection model answers the question regarding how one's response to the severity of the disease affects death (failure). In contrast to the previous work that mostly deals with continuous longitudinal data, the present study on joint modelling is based on ordinal categorical outcomes. This requires a very different approach in model building. Furthermore, unlike previous joint models, here we assume the repeated measures and failure times remain still dependent, even though the subject-specific random effect is conditionally given. This is because if the primary focus is to estimate and compare events between survival and any serial trends, one might postulate those repeated measurements that serve as covariates in the survival model. Consequently, an analyst may be able to predict the time for failure for any specific set of values of longitudinal outcome.

Our research is motivated by a data set concerning longitudinal outcomes of subjects involved in a study on muscular dystrophy among the children caused by deletion, duplication or point mutation of the Dystrophin gene located on X-chromosome. These values consist of observations on six different muscles that are responsible for walking. Two time-to-event indicators are also observed over different time points. These are time taken by a patient to walk 4 steps and to get up from lying state. Censoring occurs if a person fails to complete the four steps in 1 min or fails to stand up within 40 s. As both the causes are highly dependent, there is no harm in assuming that both the failures occur simultaneously. These scores, observed up to failure from any of the causes for each subject, are ordinal in nature. Our interest is on characterizing the relationship between failure time (due to any one of the causes) and the longitudinal outcomes. In this study only the first one (complete the four steps in 1 min) is used, which appears to be more severe according to the doctors.

Although longitudinal models in the presence of nonlinear relationships between outcomes and covariates are quite popular in practice, their use has been limited due to complexity of the likelihood function. Furthermore, ordinality in the data increases the complexity of the likelihood. Here we develop a methodology whereby a joint likelihood, based on ordered longitudinal scores and time to failure data, is maximized. The 'exact' and 'approximate' methods based on Monte Carlo EM (MCEM) are addressed. By 'exact' we mean the inference is based on the exact likelihood. Inference achieved by approximating the likelihood is considered to be an 'approximate' method. As the longitudinal outcomes and failure time data are simultaneously taken care of, this provides more efficient use of data. Since the parameters of both models are estimated at the same time, our approach uses not only the observed ordered longitudinal outcome but also failure information to get precise estimates of the strength of relationship between the longitudinal data and the risk of failure. We assume that the subject-specific (cluster-specific) effect is constant over time and is thus identical at all event times when the subject is at risk.

The rest of the paper is organized as follows: Section 2 describes the model and the likelihood. Section 3 describes a MCEM algorithm for exact inference. As a special case, a linear mixed model is discussed in this section. Section 4 presents a MCEM algorithm for approximate inference. The algorithm was carried out with the initial choice as the posterior solution of a suitable full Bayesian model shown in the Appendix. Section 5

presents a substantive example from a muscular study. Small sample properties of the proposed estimates are also examined via simulation studies in Section 6. Section 7 concludes the article with some discussion.

2. The model and the likelihood

Since our primary interest is to determine the impact of longitudinal measures (scores) on the survival outcomes, we construct the joint likelihood as the product of the time to event likelihood conditional on the longitudinal measures multiplied by the likelihood of the longitudinal outcomes. Note that for ordinal longitudinal outcomes, this likelihood is based on a latent class of mixture models where the individual trajectories for each patient are accounted for. In both models, observed covariates such as a subject's baseline variables and environmental information over time may be taken as predictors. Some of them may be time independent and some may not be. Unobserved factors enter the model as subject-specific random effects, which do not depend on the time points, so as to account for unobserved heterogeneity.

Specifically for each subject i, longitudinal q dimensional ordinal scores, taken at time point j, are denoted as $\mathbf{y}_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, m$. Thus $\mathbf{y}_{ij}$ is the longitudinal response vector for subject i at time point j.

Since the ordinal score vector $\mathbf{y}_{ij}$ can be considered as a result of the discretization of a continuous latent variable vector $\mathbf{y}^{\star}_{ij}, j = 1, \ldots, n_i$ [2], the value of each component of $\mathbf{y}_{ij}$ may be determined by a series of increasing thresholds. We express for $k = 1, 2, \ldots, q$

$$\begin{aligned} y_{ijk} &= m_k \quad \text{if} \quad \gamma_{k,m_k-1} \leq y^{\star}_{ijk} < \gamma_{k,m_k} \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

Furthermore, the survival time t_{ij} for this subject i can be right censored at some time c_{ij}. Hence $v_{ij} = \min(t_{ij}, c_{ij})$ denotes the observed time at risk for subject i on the jth time point. The failure indicator is defined as $\delta_{ij} = I_{(t_{ij} \leq c_{ij})}$. Using a Cox frailty model [4], the right-censored survival times for the particular cause are then modelled as

$$\lambda\left(t_{ij} | \mathbf{y}^{\star}_{\mathbf{ij}}(\mathbf{t_{ij}})\right) = \lambda_0(t_{ij}) \exp\left\{\beta_0' \mathbf{y}^{\star}_{\mathbf{ij}}(\mathbf{t_{ij}}) + \beta_1' \mathbf{x_{ij}}(\mathbf{t_{ij}})\right\} \tag{1}$$

where $\lambda_0(\cdot)$ is an unknown baseline hazard function and β_0, β_1 are vectors of regression parameters associated with the longitudinal variates and covariates, respectively. Note that this hazard varies with time not only because λ_0 does, but also because both $\mathbf{y}$ and $\mathbf{x}$ do. The specification of the hazard in Equation (1) leads to the following likelihood:

$$f\left(t_{ij}, \delta_{ij} | \mathbf{y}^{\star}_{ij}\right) = \lambda_0(t_{ij})^{\delta_{ij}} \exp\left[\delta_{ij}\left\{\beta_0' \mathbf{y}^{\star}_{\mathbf{ij}}(\mathbf{t_{ij}}) + \beta_1' \mathbf{x_{ij}}(\mathbf{t_{ij}})\right\}\right] S_{ij}(t_{ij}) \tag{2}$$

where

$$S_{ij}(t_{ij}) = \exp\left[-\int_0^{t_{ij}} \lambda_0(u) \exp\left\{\beta_0' \mathbf{y}^{\star}_{ij}(u) + \beta_1' \mathbf{x_{ij}}(u)\right\} \mathrm{d}u\right].$$

Then conditional on the subject-specific effects $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iq})'$, we can write the

joint distribution of the categorical variable $\mathbf{y}_{ij}$ as

$$\begin{aligned} P_{j_{m_1,\ldots,m_q}|\boldsymbol{\xi}_i} &= P(y_{ij1} = m_1, y_{ij2} = m_2, \ldots, y_{ijq} = m_q|\boldsymbol{\xi}_i) \\ &= P\Big(\gamma_{1,m_1-1} \le y^{\star}_{ij1} < \gamma_{1,m_1}, \gamma_{2,m_2-1} \le y^{\star}_{ij2} < \gamma_{2,m_2}, \\ &\quad \ldots, \gamma_{q,m_q-1} \le y^{\star}_{ijq} < \gamma_{q,m_q}|\boldsymbol{\xi}_i\Big) \\ &= \prod_{k=1}^{q} P\Big(\gamma_{k,m_k-1} \le y^{\star}_{ijk} < \gamma_{k,m_k}|\boldsymbol{\xi}_i\Big) = \prod_{k=1}^{q} p_{jk|\boldsymbol{\xi}_i} \end{aligned} \tag{3}$$

where the latent variable is subject to a nonlinear mixed model, *i.e.*

$$\mathbf{y}^{\star}_{ij} = \mathbf{g}(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i) + \mathbf{e}_{ij} \tag{4}$$

where $\mathbf{x}_{ij}$ being the vector of covariates associated with $\mathbf{y}_{ij}$, $\mathbf{e}_{ij}$ being the vector of measurement errors, $g(\cdot)$ is a known function, $\boldsymbol{\xi}_i$ is distributed with density function $h(\cdot)$ and $\mathbf{e}_{ij} \sim N_q(0, \sigma_0^2 I)$, σ_0^2 is a known positive scaler. Then the likelihood for the observed data can be expressed as

$$\begin{aligned} L &= \prod_{i=1}^{m} \int_{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} \prod_{j=1}^{n_i} l_j^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} h(\boldsymbol{\xi}_i) \mathrm{d}\boldsymbol{\xi}_i \mathrm{d}\mathbf{y}^{\star}_{ij} \\ &= \prod_{i=1}^{m} \int_{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} l_{iJ}^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} h(\boldsymbol{\xi}_i) \mathrm{d}\boldsymbol{\xi}_i \mathrm{d}\mathbf{y}^{\star}_{ij}, \end{aligned} \tag{5}$$

where

$$\begin{aligned} l_{iJ}^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} &= \prod_{j=1}^{n_i} l_j^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}}, \\ l_j^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} &= f\Big(t_{ij}, \delta_{ij}|\mathbf{y}^{\star}_{ij}\Big) \sum_{m_1=1}^{r} \sum_{m_2=1}^{r} \cdots \sum_{m_q=1}^{r} 1^{ij}_{m_1,m_2,\ldots,m_k} P_{j_{m_1,m_2,\ldots,m_q}|\xi_i, \mathbf{y}^{\star}_{ij}} \end{aligned} \tag{6}$$

and

$$\left.\begin{aligned} 1^{ij}_{m_1,m_2,\ldots,m_k} &= 1 \quad \text{if } \; y_{ij1} = m_1, y_{ij2} = m_2, \ldots, y_{ijq} = m_q, \\ &= 0 \qquad\qquad\qquad \text{otherwise} \end{aligned}\right\},$$

or equivalently

$$\begin{aligned} l_j^{\xi_i, \mathbf{y}^{\star}_{ij}} &= \sum_{u=1}^{q} \sum_{m_u=1}^{r} 1^{ij}_{m_1,m_2,\ldots,m_k} \prod_{k=1}^{q} f\Big(t_{ij}, \delta_{ij}|\mathbf{y}^{\star}_{ij}\Big) p_{jk|\xi_i} \\ &= f\Big(t_{ij}, \delta_{ij}|\mathbf{y}^{\star}_{ij}\Big) \sum_{u=1}^{q} \sum_{m_u=1}^{r} 1^{ij}_{m_1,m_2,\ldots,m_q} \prod_{k=1}^{q} \left[\Phi\left\{\frac{\gamma_{k,m_k} - g(\mathbf{x}_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i)}{\sigma_0}\right\} - \Phi\left\{\frac{\gamma_{k,m_k-1} - g(\boldsymbol{x}_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i)}{\sigma_0}\right\}\right] \\ &= f(t_{ij}, \delta_{ij}|\mathbf{y}^{\star}_{ij}) \quad \sum_{u=1}^{q} \sum_{m_u=1}^{r} 1^{ij}_{m_1,m_2,\ldots,m_q} 1^{ij*}_{m_1,m_2,\ldots,m_q} \prod_{k=1}^{q} \left[\phi\left\{\frac{g(\mathbf{x}_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i)}{\sigma_0}\right\}\right] \end{aligned} \tag{7}$$

where

$$\left.\begin{aligned} 1^{ij\star}_{m_1,m_2,\ldots,m_q} &= 1 \quad \text{if} \quad \gamma_{k,m_k-1} \le y^{\star}_{ijk} < \gamma_{k,m_k} \quad \text{for all} \quad k = 1, 2, \ldots, q, \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{aligned}\right\}$$

Our objective is to propose a suitable method for obtaining the maximum likelihood estimates of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \beta_0, \beta_1)'$.

2.1 Latent linear model

As a special case we have also considered a latent linear model. In this case, Equation (4) changes to

$$\boldsymbol{y}^{\star}_{ij} = \boldsymbol{x}'_{ij}\alpha\mathbf{1} + \boldsymbol{\xi}_i + \mathbf{e}_{ij}, \tag{8}$$

where $\mathbf{1}$ is q vector of unit elements and $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iq})'$. The likelihood remains the same except for the form of the function $g(\cdot)$.

3. Monte Carlo-based EM method

Inference for the joint model can be made using the MCEM algorithm. This MCEM approach is very effective because integral evaluation in the likelihood Equation (5) is otherwise impossible. In fact, the parameters appearing in the joint model can be estimated by performing iterations between an E-step and an M-step. The E-step is based on the complete data log likelihood which can be expressed as

$$l = \sum_{i=1}^{m} l_{ic},$$

where

$$l_{ic} = \int_{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} \prod_{j=1}^{n_i} l_j^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} h(\boldsymbol{\xi}_i) \mathrm{d}\boldsymbol{\xi}_i \mathrm{d}\mathbf{y}^{\star}_{ij},$$

where $l_j^{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}}$ is given in Equation (7).

The method is based on the relation between the unconditional score and the conditional score, *i.e.*

$$S_\theta = \sum_{i=1}^{m} \frac{\int_{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}} \frac{\partial l_{iJ}(\boldsymbol{\theta})}{\partial \theta} \mathrm{d}\xi_i \mathrm{d}\mathbf{y}^{\star}_{ij}}{\int_{\xi_i, \mathbf{y}^{\star}_{ij}} l_{iJ}(\boldsymbol{\theta}) \mathrm{d}\xi_i \mathrm{d}\mathbf{y}^{\star}_{ij}} = \sum_{i=1}^{m} E_{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}|\mathbf{y}_{ij}} \left(\frac{\partial \log l_{iJ}(\boldsymbol{\theta})}{\partial \theta} \right) = \sum_{i=1}^{m} E_{\boldsymbol{\xi}_i, \mathbf{y}^{\star}_{ij}|\text{data}} \left(S^{\star}_{\theta} \right). \tag{9}$$

where $S_{\boldsymbol{\theta}}$ is the unconditional score and $S^{\star}_{\boldsymbol{\theta}}$ is the conditional score obtained from the logarithm of the joint likelihood given by Equation (5). In fact, we apply EM algorithm where at the $(l+1)$th stage we generate $\boldsymbol{\xi}_i$, $\mathbf{y}^{\star}_{ij}$ from Equation (7), when $\boldsymbol{\theta} = \boldsymbol{\theta}^{(l)}$ (E-step) and then at the M-step we solve for $\boldsymbol{\theta}$ from

$$0 = S_{\boldsymbol{\theta}} \simeq \sum_{i=1}^{m} \frac{1}{R} \sum_{r=1}^{R} S^{\star}_{\theta^{(l)}_{ir}},$$

where $S^{\star}_{\theta^{(l)}_{ir}}$ is the value $S^{\star}_{\theta}$ for the ith individual, obtained by substituting rth set of the generated values from the conditional distribution of $(\boldsymbol{\xi}_i, \mathbf{y}_i^{\star})$ at $\theta = \theta^{(l)}$, $r = 1(1)R$, R being the number of samples. In order to solve the above, one-step Newton-Raphson method is used. This gives

$$\hat{\boldsymbol{\theta}}^{(l)} = \hat{\boldsymbol{\theta}}^{(l-1)} + (M^{-1}\hat{S}_{\boldsymbol{\theta}})|_{\hat{\boldsymbol{\theta}}^{(l-1)}}, \tag{10}$$

where for each i, $(\boldsymbol{\xi}_{i1}^{(l)}, \mathbf{y}_{ij1}^{\star(l)}), (\boldsymbol{\xi}_{i2}^{(l)}, \mathbf{y}_{ij2}^{\star(l)}), \ldots, (\boldsymbol{\xi}_{iR}^{(l)}, \mathbf{y}_{ijR}^{\star(l)})$ are generated from (7) using Metropolis-Hastings (MH) algorithm [7] and $M = (-\partial S_{\theta}/\partial \boldsymbol{\theta})$.

In summary, the foregoing EM algorithm proceeds as follows:

Step 1. Obtain an initial estimate of $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$ and generate $(\boldsymbol{\xi}_i^{(0)}, \mathbf{y}_{ij}^{\star(0)})$.

Step 2. At the $(l+1)$st iteration, obtain MC estimates of the first and second order derivatives by generating the random effects and latent variable $(\boldsymbol{\xi}_i^{(l+1)}, \mathbf{y}_{ij}^{\star(l+1)})$ by MH algorithm for the ith individual, $l = 1, 2, \ldots$; $i = 1, 2, \ldots, m$.

Step 3. At the $(l+1)$st iteration, obtain updated estimates $\boldsymbol{\theta}^{(l+1)}$ using one-step Newton-Raphson method.

Step 4. Iterate between Steps 2 and 3 until convergence.

The proposed algorithm is identical for both the models—linear and nonlinear. The foregoing MC-based EM algorithm may be slow to converge, especially when the dimension of the random effects is not small. Again the MH algorithm, within MCEM, makes things more time consuming. In Section 4, we describe a particularly useful approximate method, which converges faster by avoiding the sampling of random effects by MH algorithm.

4. Approximate method

In view of the difficulty that arises in the E-step due to high dimension of random effects, it looks worthwhile to propose a reasonable approximate method. This approximate method can at least skip a generation of samples from the posterior distribution of random effects, or is at least capable of drawing such sample observations from the normal distribution pretty easily. The method considered in this section is computationally more efficient as it converges faster and restricts other potential computational problems associated with the exact method.

There has been several studies on nonlinear mixed effect (NLME) models [15] where the exact likelihood is approximated on the basis of Taylor expansions or Laplace approximations [3,17]. Lindstorm and Bates [11] considered linear approximation through first order Taylor expansion around the current estimates and used posterior modes for the random effects. This makes the integration over the random effects in the E-step possible.

The approximate method can clearly be explained as follows. In view of (4), we write a general NLME latent model (conditional on $\boldsymbol{\xi}_i$) as

$$y_{ij}^{\star} = \mathbf{g}(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i) + \mathbf{e}_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, N \tag{11}$$

where $\mathbf{g}$ is a nonlinear function. To simplify the notations, we suppress iteration numbers and MC sample numbers. We denote the current estimates of $(\boldsymbol{\alpha}, \boldsymbol{\xi}_i)$ by $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\xi}}_i)$. Adjusting, we can write

$$\tilde{\mathbf{y}}_{ij} = \mathbf{y}_{ij}^{\star} - \mathbf{g}(x_{ij}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\xi}}_i) + \mathbf{Z}_{ij}\hat{\alpha} + \mathbf{W}_{ij}\hat{\boldsymbol{\xi}}_i, \tag{12}$$

so that the model is taken as

$$\tilde{\mathbf{y}}_{\mathbf{ij}} = \mathbf{Z}_{ij}\boldsymbol{\alpha} + \mathbf{W}_{ij}\boldsymbol{\xi}_i + \mathbf{e}_{ij}, \quad \mathbf{e}_{ij}|\boldsymbol{\xi}_i \sim N(\mathbf{0}, \sigma^2 I), \quad \boldsymbol{\xi}_i \sim N(\mathbf{0}, \mathbf{D}), \quad \mathbf{D} = \mathbf{D}(\boldsymbol{\eta}), \tag{13}$$

where $\mathbf{Z}_{ij}$ and $\mathbf{W}_{ij}$ are matrices of order $q \times p$ and $q \times s$, respectively, with $\mathbf{z}_{ij_k} = (\partial g(x_{ij}, \boldsymbol{\alpha}, \hat{\xi}_i))/(\partial \beta_0')|_{\hat{\boldsymbol{\alpha}}}$, $\mathbf{w}_{ij_k} = (\partial g(x_{ij}, \boldsymbol{\alpha}, \hat{\xi}_i))/(\partial \boldsymbol{\xi}_i')|_{\hat{\boldsymbol{\xi}}_i}$.

Under the LME model (13), it is straightforward to show that $[\boldsymbol{\xi}_i|\mathbf{y}_{ij}, \mathbf{W}_{ij}, \mathbf{Z}_{ij}, \hat{\boldsymbol{\alpha}}] \sim N(\tilde{\boldsymbol{\xi}}_i, \tilde{\Sigma}_i)$, where $\tilde{\Sigma}_i = (\hat{\sigma}^{-2}\mathbf{W}'_i\mathbf{W}_j + D^{-1}(\hat{\boldsymbol{\eta}}))^{-1}$, $\tilde{\boldsymbol{\xi}}_i$ is the best linear unbiased predictor estimates of the random effects (see Robinson 1991) and is equal to $\Sigma_i\mathbf{W}'_i(\tilde{\mathbf{Y}}_i - \mathbf{Z}_i\hat{\boldsymbol{\alpha}})/\hat{\sigma}^2$, $\tilde{\mathbf{Y}}_i = (\tilde{\mathbf{Y}}_{i1}, \tilde{\mathbf{Y}}_{i2}, .., \tilde{\mathbf{Y}}_{in_i})'$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}\mathbf{Z}_{i2}, \ldots, \mathbf{Z}_{in_i})'$, $\boldsymbol{W} = (\boldsymbol{W}_{i1}\boldsymbol{W}_{i2}, \ldots, \boldsymbol{W}_{in_i})'$.

The E-step for the ith individual can be written as

$$Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) = E_{\boldsymbol{\xi}_i, \tilde{\mathbf{Y}}_i|\mathbf{Y}_i, \mathbf{x}_i}[l_{ic}(\boldsymbol{\theta}; \tilde{\mathbf{Y}}_i, \boldsymbol{\xi}_i)|\mathbf{Y}_i, \mathbf{x}_i] = E_{\tilde{\mathbf{Y}}_i|\mathbf{Y}_i, \theta^{(l)}}[E_{\boldsymbol{\xi}_i|\tilde{\mathbf{Y}}_i, \mathbf{Y}_i, \boldsymbol{\theta}^{(l)}}\{l_{ic}(\boldsymbol{\theta}; \tilde{\mathbf{Y}}_i, \boldsymbol{\xi}_i)\}],$$

where

$$\begin{aligned}
l_{ic} &= \log L_{ic} \\
L_{ic} &= L_{i1} \times L_{i2c} \times h(\boldsymbol{\xi}_i) \\
L_{i1} &= \prod_{j=1}^{n_i} f(t_{ij}, \delta_{ij}) \\
L_{i2c} &= \prod_{j=1}^{n_i} \sum_{u=1}^{q} \sum_{m_u=1}^{r} 1^{ij}_{m_1, m_2, \ldots, m_q} \\
&\quad \times \prod_{k=1}^{q} \left[\Phi\left\{ \frac{\gamma_{k,m_k} - g(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i)}{\sigma_0} \right\} - \Phi\left\{ \frac{\gamma_{k,m_k-1} - g(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i)}{\sigma_0} \right\} \right] \\
&\approx \prod_{j=1}^{n_i} \sum_{u=1}^{q} \sum_{m_u=1}^{r} 1^{ij}_{m_1, m_2, \ldots, m_q} \\
&\quad \times \prod_{k=1}^{q} \left[\Phi\left\{ \frac{\gamma_{k,m_k} - \mathbf{Z}_{ij}\boldsymbol{\alpha} - \mathbf{W}_{ij}\boldsymbol{\xi}_i}{\sigma_0} \right\} - \Phi\left\{ \frac{\gamma_{k,m_k-1} - \mathbf{Z}_{ij}\boldsymbol{\alpha} - \mathbf{W}_{ij}\boldsymbol{\xi}_i}{\sigma_0} \right\} \right]
\end{aligned}$$

The foregoing integral generally does not have a closed form, and evaluation of the integral by numerical method is generally infeasible except for simple cases. However, this can be evaluated by using MCEM algorithm (see [16]). Specifically, we may use the Gibbs sampler to generate samples from the joint conditional distributions of $\tilde{\mathbf{y}}_{ij}$ and $\boldsymbol{\xi}_i$.

For individual i and at time t_{ij}, let $((\tilde{\mathbf{y}}^{(l)}_{ij1}, \boldsymbol{\xi}^{(l)}_{i1}), (\tilde{\mathbf{y}}^{(l)}_{ij2}, \boldsymbol{\xi}^{(l)}_{i2}), \ldots, (\tilde{\mathbf{y}}^{(l)}_{ijM_i}, \boldsymbol{\xi}^{(l)}_{iM_i}))$ be a random sample of size M_i generated from the following full conditional distributions

$$\begin{aligned}
&[\tilde{\mathbf{y}}^{(l+1)}_{ij}|\boldsymbol{\xi}^{(l)}_i, \boldsymbol{\theta}^{(l)}, \mathbf{y}^{(l)}_{ij}] \sim \text{ truncated } N(\mathbf{Z}_{ij}\boldsymbol{\alpha} + \mathbf{W}_{ij}\boldsymbol{\xi}_i, \sigma^2 I), \quad j = 1, \ldots, n_i \\
&\boldsymbol{\xi}^{(l+1)}_i|\tilde{\mathbf{Y}}^{(l+1)}_i, \quad \boldsymbol{\theta}^{(l)} \sim N(\boldsymbol{\xi}^{(l)}_i, \tilde{\Sigma}^{(l)}_i).
\end{aligned} \tag{14}$$

The E-step at the $(l+1)$st iteration can be written as

$$Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) = \frac{1}{M_i}\sum_{l=1}^{M_i} l_{ic}(\boldsymbol{\theta}; \tilde{\mathbf{Y}}_i, \boldsymbol{\xi}_{il}|\boldsymbol{\theta}^{(l)}).$$

In view of the simpler full conditional form, we do not need to take the help of MH sampling. This leads to faster convergence (less computation time). The M step of the MCEM algorithm maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) = \Sigma_i Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)})$ to produce an updated estimate $\boldsymbol{\theta}^{(l+1)}$. The score function $S(\boldsymbol{\theta})$ can be computed as

$$S(\boldsymbol{\theta}) = \frac{\partial \sum_i Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)})}{\partial \boldsymbol{\theta}} = \sum_i \frac{1}{M_i}\sum_{l=1}^{M_i} \frac{\partial l_{ic}(\boldsymbol{\theta}; \tilde{\mathbf{Y}}_i, \boldsymbol{\xi}_{il})}{\partial \boldsymbol{\theta}}$$

and hence $M = \partial S(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$.

The foregoing EM algorithm in this case may be summarized as follows:

Step 1. Consider $\mathrm{l} = 0$ and initialize $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$.

Step 2. Using Gibbs sampler, generate M_i observations $(\tilde{\mathbf{y}}_{ijw}, \boldsymbol{\xi}_{iw})$, $w = 1, \ldots, M_i$ from (14).

Step 3. At the $(l+1)$st iteration, obtain updated estimates $\boldsymbol{\theta}^{(l+1)}$ using one-step Newton-Raphson method as earlier.

Step 4. Continue till a convergence is achieved.

It is well known that the EM algorithm is highly dependent on the choice of the initial estimates. In order to obtain a good initial choice of the estimates, we develop a hierarchical Bayesian framework (see Appendix). This provides a posterior solution in a very short time span. Since our primary concern is to carry out an MCEM-based analysis, we did not emphasize on the choices of priors of parameters and hyperparameters. Just a standard Bayesian framework with conjugate priors is considered for obtaining a posterior solution.

5. Analysis of muscle score data

5.1 The models and estimation

This article is motivated by a study on muscular dystrophy among the children. The data consist of scores of six different muscles ($q = 6$), (neck, deltoid, bicep, iliopsoas, quadriceps and hamstrings), responsible for walking, observed on different time points. These scores are discrete in nature and can vary from 0 (very weak) to 10 (normal). Along with these, few important covariates such as age of the patient, on-set age and weight are recorded. To see the trajectory of the composite score (average of the muscle scores) over time, a LOWESS smoothing was done. The exploratory data analysis revealed that the composite score varies over time. This justifies the assumption of time dependency of the model.

It is natural that both the time to events are highly associated. In fact, in an earlier investigation, this was observed in a 2×2 contingency table based on a total of 80 observations. By plotting the composite score versus a model-based estimate of hazard, given in Figure 1, we see a decrease in the hazard estimates. This phenomenon is observed for both time to event measurements, and, therefore, gives us initial evidence that there is indeed some positive association between high composite score and longer relapse-free survival.

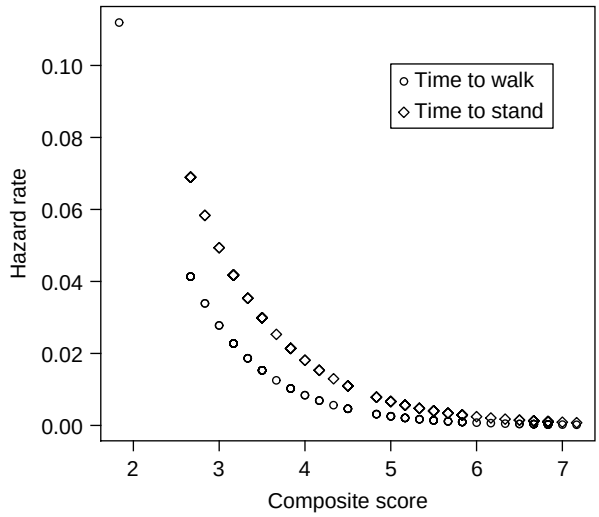


Figure 1. Model-based hazard.

The proposed model is used here to analyse the data on muscular dystrophy. The study provides the data on children (patient) ranging from 4 to 12 years. The muscle scores and time to failure from the particular causes are taken into account for each patient. The only covariate considered is the weight of each patient. Here the muscle score can vary from 0 to 10. To reduce the number of categories, we condense the data a bit. Observing the frequency distribution of the muscle scores, six categories are taken into consideration. Together with these, the time taken by a patient to walk is recorded, subject to censoring. Specifically we express the latent nonlinear mixed model (corresponding to categorical variable $\mathbf{y}_{ij}$) as

$$\mathbf{y}_{ij}^{\star} = \exp(\alpha_1 + \alpha_2 x_{0ij})\mathbf{1} + \boldsymbol{\xi}_i + \mathbf{e_{ij}}, \tag{15}$$

where $\mathbf{1}$ is a $q \times 1$ vector with each term being unity and for

$$\lambda_{ij}(t_{ij}) = \lambda_0(t_{ij})\exp\left(\sum_{k=1}^{q} y_{ijk}^{\star}\beta_k\right)$$

$$\boldsymbol{\xi}_i \sim N_q(\mathbf{0}, \Sigma), \quad \mathbf{e}_{ij} \sim N_q\left(\mathbf{0}, \sigma_0^2 I\right)$$

with

$$\Sigma = ((\sigma^2 \rho^{|s-t|})), \quad s, t = 1(1)4.$$

Note that for the latent linear mixed set-up, we consider, instead of (15), the model

$$\mathbf{y}_{ij}^{\star} = (\alpha_1 + \alpha_2 x_{0ij})\mathbf{1} + \boldsymbol{\xi}_i + \mathbf{e_{ij}}$$

for our analysis. Hazard function has been computed following the counting process formulation of Anderson and Gill [1]. This addresses the computational problem of the time-dependent covariate in the Cox proportional hazard model. For the sake of

convenience we assume the regression coefficient β_{02} that comes in the above two latent models is the same although it could be assumed to be different for latent variables associated with different muscles.

5.2 Conclusions

The performance of the MCEM-based 'exact' and 'approximate' algorithm can be seen in Figure 2. Note that in order to have stabilized values of the standard errors (SEs) of the estimates, we consider here a MC-based sandwich technique (see [6]). On the basis of the methodologies proposed in Sections 2 and 3, we carry out a numerical study. The estimated values of the parameters at convergence along with their SEs are given in Table 1. Clearly, the three methods may give different results, which is evident from Table 1. In most cases we found that the approximate estimates are closer to exact estimates than the linear mixed model estimates, and the linear mixed model method generally produces larger standard errors than the approximate estimates, possibly indicating inefficiency of the method based on a linear mixed model. The covariate effects in the model are tested based on Wald statistics. With regards to the computational time, it turns out that the standard MCEM for the exact method is quite slow for this example. All computations were done by using 'R-2.10.1', and the codes can be obtained from the authors on request. The analysis leads to some interesting conclusions. As the coefficients β_{03}, β_{05} and β_{06} are statistically significant, in all models, it indicates that the third, fifth and sixth muscle (biceps, quadriceps and hamstring, respectively) scores are strongly related to the failure

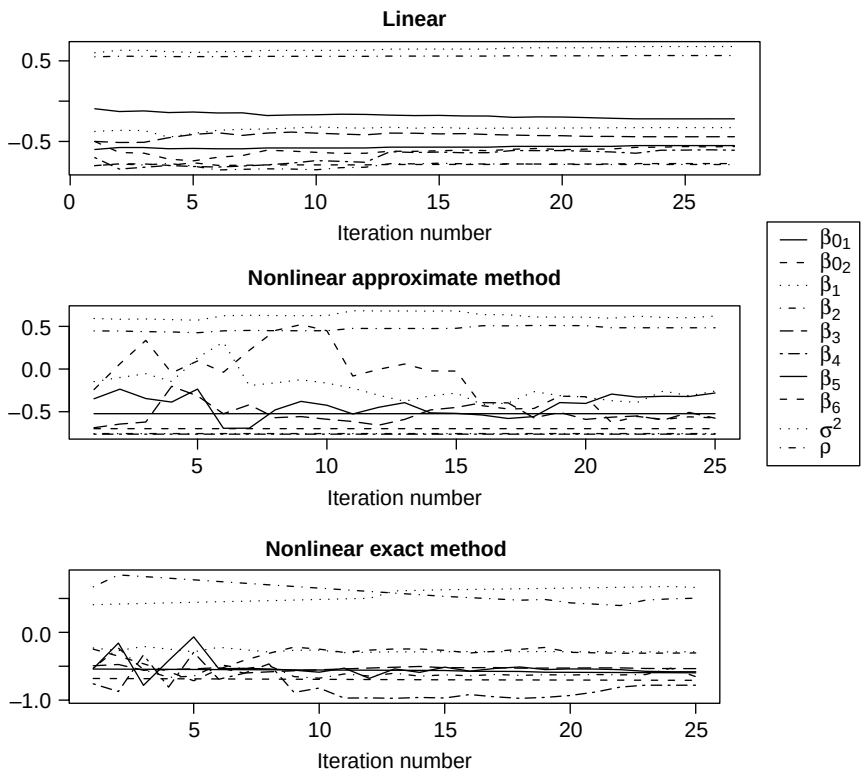


Figure 2. Performance of the MCEM-based proposed algorithm: data analysis.

Table 1. Estimates of parameters with standard errors in parentheses: data analysis.

Parameter	Linear	Nonlinear	
		Approximate method	Exact method
α_1	−0.219 (0.528)	−0.203 (0.552)	−0.323 (0.395)
α_2	−0.563 (0.792)	−0.681 (0.853)	−0.480 (0.427)
β_{01}	−0.328 (0.964)	−0.464 (0.781)	−0.279 (0.196)
β_{02}	−0.787 (0.703)	−0.953 (0.753)	−0.858 (0.486)
β_{03}	−0.441 (0.889)	−0.546* (0.104)	−0.726* (0.121)
β_{04}	−0.608 (0.752)	−0.779 (0.277)	−0.639 (0.375)
β_{05}	−0.551 (0.226)	−0.498* (0.248)	−0.601* (0.144)
β_{06}	−0.774*(0.393)	−0.741* (0.134)	−0.784* (0.224)
σ_2	0.677 (0.732)	0.661 (0.062)	0.647 (0.187)
ρ	0.565 (0.821)	0.505 (0.022)	0.556 (0.117)

*Significant effect.

time. Thus the low quadriceps, hamstring and biceps muscle scores give a strong signal to the doctors on the patients' health condition. The Akaike Information Criterion (AIC), which indicates model fit and model parsimony, is also computed for both the models. It is found to be 75.38 in latent linear models. In the case of a nonlinear model, it is found to be 29.76 and 35.49 corresponding to the exact method and approximate method, respectively. Among these models, on the basis of AIC, the nonlinear model with the exact method may be assumed to be the best fit for these data.

6. Simulations

We conducted simulation studies to compare the exact and approximate methods. Here, we consider the same model that has been used in data analysis. We take a continuous time-dependent covariate for the latent model. The survival model is expressed conditionally with longitudinal outcomes and covariates as the predictor. The longitudinal model involves subject-specific random effects. The covariate x_{ij} is simulated from a Normal distribution $N(0.65, 1)$. The latent variable $y^{\star}_{ijk}(k = 1, 2, \ldots, 6)$ are generated in view of the above model with $\alpha_1 = 0.2$, $\alpha_2 = -0.3$, $\sigma^2 = 0.7$, $\rho = 0.6$, $\sigma_0^2 = 1$ and hence the ordinal data y_{ij} are generated for $k = 1, 2, \ldots, 6$. Note that the values of the cut-off points are taken here as $\gamma_0 = -\infty$, $\gamma_1 = -2\log 3$, $\gamma_2 = -\log 3$, $\gamma_3 = 0$, $\gamma_4 = \log 3$, $\gamma_5 = 2\log 3$, $\gamma_6 = \infty$. Furthermore, based on 25% censoring proportion (*i.e.* $c = 25\%$), censoring times were generated as independent uniform random variable on $[0, c]$ for both cases. The parameters β_{01} to β_{06} appearing as coefficients associated with the observed ordinal variables are taken as $\beta_{01} = -0.3, \beta_{02} = -0.6, \beta_{03} = -0.4, \beta_{04} = -0.6, \beta_{05} = -0.5$, $\beta_{06} = -0.6$. These values are chosen on the basis of data analysis.

Note that using MH algorithm, 500 samples are generated from Equation (7) where for each sample MC average of first order and second order derivatives is taken on the basis of 500 such observations. An improved estimate is obtained using Newton-Raphson method. The iteration is continued until the desired level ($|\theta^{(k+1)} - \theta^k| < 0.0001$) of accuracy is achieved. Like the data analysis, the exact method is found to be very slow here in terms of convergence.

Simulation study is also carried out for the latent linear model. The initial choices for parameters appearing in the latent longitudinal model are obtained by replacing $\boldsymbol{y}^{\star}_{ij}$ by $\boldsymbol{y}_{ij}$ and taking the model as a mixed model. The other parameters in the survival model are estimated independently. In this case also, convergence is achieved within 25 iterations.

Table 2. Estimates of parameters with standard errors in parentheses: simulation study.

			Nonlinear	
Parameter	True value	Linear	Approximate method	Exact method
α_1	0.2	0.195 (0.509)	0.178 (0.474)	0.181 (0.905)
α_2	−0.3	−0.363 (0.485)	−0.339 (0.186)	−0.331 (0.132)
β_{01}	−0.2	−0.212 (0.434)	−0.258 (0.220)	−0.226 (0.331)
β_{02}	−0.6	−0.425 (0.652)	−0.560 (0.460)	−0.540 (0.542)
β_{03}	−0.4	−0.442 (0.411)	−0.323 (0.771)	−0.385 (0.107)
β_{04}	−0.6	−0.457 (0.436)	−0.521 (0.203)	−0.528 (0.365)
β_{05}	−0.5	−0.251 (0.354)	−0.422 (0.139)	−0.512 (0.325)
β_{06}	−0.6	−0.574 (0.352)	−0.579 (0.175)	−0.661 (0.139)
$\sigma 2$	0.7	0.677 (0.218)	0.693 (0.163)	0.624 (0.143)
ρ	0.6	0.615 (0.315)	0.664 (0.271)	0.658 (0.144)

The simulation results are presented in Table 2. The presence of longitudinal model has been assured through their SEs. The SE values have been improved by stabilizing through the sandwich approach. The proposed MCEM algorithm converges in all cases within 25 iterations. It is evident from Figure 3 that the parameters associated with the longitudinal model initially show somewhat erratic behaviour in the convergence, although at a later stage they ultimately converge. In contrast, the estimates of survival parameters in the survival model exhibit a very steady convergent behaviour. Table 2 shows a comparison between exact and alternative methods. It shows that the approximate

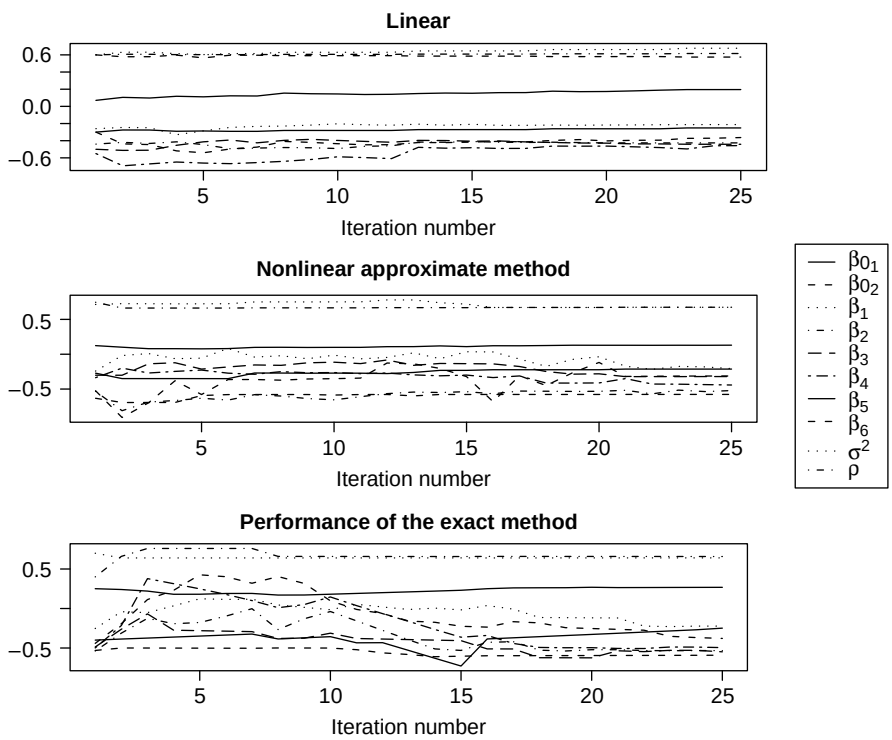


Figure 3. Performance of the MCEM-based proposed algorithm: simulation.

method performs quite well. There seems to be very little difference between the exact and approximate estimates. Thus in this context, the approximate method may offer important advantages because it iteratively solves linear models with less computing time. Clearly our exact method exhibits a better performance in the sense of smaller bias. Here, we also found that the approximate estimates are closer to the exact estimates except for the estimate of β_{03}, which is underestimated by the approximate method, whereas β_{02} is overestimated by the exact method, although the magnitude of biases in both cases is very small.

7. Discussion

We have proposed a joint model for the observed longitudinal and survival data that link the responses via a common frailty. Observe that in our joint model, the survival model part depends on the longitudinal outcomes. This arises from the fact that the muscle score has a big role to play for the time to failure. Furthermore, we have assumed that there may be a nonlinear relationship between weight and the muscle scores through a latent vector. It is important to check the appropriateness of this covariate model and the sensitivity of parameters under alternative covariate models. Sensitivity analysis may be carried out by fitting various covariate models. There is scope for further extension in the direction of nonlinear relationship. However, the linear mixed model is also fitted. We have considered the exact and approximate methods for NLME models in joint modelling. When the exact method exhibits computational problems such as slow convergence, the approximate method may be preferable, because it is computationally more appealing. A good idea may be to use the approximate method for model selection while the exact method is for final selection. This may be done by using the approximate estimates as the starting values for the exact method. Furthermore, the estimates obtained by fitting the linear mixed model may act as the initial estimates for the approximate method. Note that the nonlinear function is expanded by Taylor series and only the linear part is retained. On the basis of the AIC value computed thereafter, one can think of considering higher order terms in the expression to get improved results in subsequent stages. For the sake of simplicity, under both simulation and data analysis, we have assumed the error dispersion matrix in the latent model to be an identity matrix up to a scaler multiple. Albeit, this could be carried out easily with an unknown value of the error variance–covariance matrix. To avoid any subjectivity in the choice of the cut-off points, one can also estimate them along with the parameters. This will just enhance the computational complexity. In our investigation, we have assumed the time to failure data for the two causes are independent. Further extension with the Bivariate survival set-up would be possible at the cost of computational simplicity. This investigation is currently being carried out and is being planned to communicate soon.

Acknowledgements

This work was supported by the UGC through the award of a minor project (F 37-547/2009 (SR)) awarded to the first author. We would like to thank National Neurosciences Centre for providing us the data. The authors are grateful to the editor, associate editor and two referees for their valuable suggestions and insights.

References

[1] P.K. Andersen and R.D. Gill, *Cox's regression model for counting processes: A large sample study*, Ann. Stat. 10 (2000), pp. 1100–1120.

[2] R.D. Bock, *Multivariate Statistical Methods in Behavioral Research*, Wiley, New York, 1975.
[3] N.E. Breslow and D.G. Clayton, *Approximate inference in generalized linear mixed models*, J. Am. Stat. Assoc. 88 (1993), pp. 9–25.
[4] D.G. Clayton, *A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence*, Biometrika 65 (1978), pp. 141–151.
[5] V. DeGruttola and X.M. Tu, *Modelling progression of CD-4 lymphocyte count and its relationship to survival time*, Biometries 50 (1994), pp. 1003–1014.
[6] H. Friedl and G. Kauermann, *Standard errors for EM estimates in generalized linear models with random effects*, Biometrics 56 (2000), pp. 761–767.
[7] W.R. Gilks and P. Wild, *Adaptive rejection sampling for Gibbs sampling*, Appl. Stat. 41 (1992), pp. 337–348.
[8] X. Guo and B.P. Carlin, *Separate and joint modelling of longitudinal and event time data using standard computer packages*, Am. Stat. 58 (2004), pp. 16–24.
[9] R. Henderson, P.J. Diggle, and A. Dobson, *Joint modelling of longitudinal measurements and event time data*, Biostatistics 1 (2000), pp. 465–480.
[10] J.G. Ibrahim, M.H. Chen, and S.R. Lipsitz, *Monte Carlo EM for missing covariates in parametric regression models*, Biometrics 55(2) (1999), pp. 591–596.
[11] M.J. Lindstrom and D.M. Bates, *Nonlinear mixed effects models for repeated measurements data*, Biometrics 46 (1990), pp. 673–687.
[12] R.J.A. Little, *Modelling the dropout mechanism in repeated-measures studies*, J. Am. Stat. Assoc. 90 (1995), pp. 1112–1121.
[13] G.K. Robinson, *That BLUP is a good thing: the estimation of random effects*, Statistical Sciences 6 (1991), pp. 15–32.
[14] T.R. Ten Have, M.E. Miller, B.A. Reboussin, and M.K. James, *Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from longitudinal study of aging*, Biometrics 56 (2000), pp. 279–287.
[15] E.F. Vonesh, V.M. Chinchilli, and K. Pu, *Goodness-of-fit in generalized nonlinear mixed-effects models*, Biometrics 52 (1996), pp. 572–587.
[16] G.C.G. Wei and M.A. Tanner, *A Monte Carlo implementation of the EM algorithm and the poor man data augmentation algorithms*, J. Am. Stat. Assoc. 85 (1990), pp. 699–704.
[17] R.D. Wolfinger, *Laplace's approximation for nonlinear mixed models*, Biometrika 80 (1993), pp. 791–795.
[18] M.C. Wu and K.R. Bailey, *Estimation and comparison of change in the presence of informative right censoring: conditional linear model*, Biometrics 45 (1989), pp. 939–955.
[19] M.S. Wulfson and A.A. Tsiatis, *A joint model for survival and longitudinal data measured with error*, Biometrics 53 (1997), pp. 330–339.

Appendix: Choice of the initial estimates

It is known that the proposed EM algorithm may be trapped in local maxima area. To overcome this difficulty, we used a fully Bayesian approach to get 'good' initial choice of the parameters. Let us denote $S = \{\gamma : \gamma_{k,m_k}, k = 1, 2, \ldots, q; -\infty = \gamma_{k,m_{k-1}} < \gamma_{k,m_{k-1}+1} < \cdots < \gamma_{k,m_k} = \infty\}$.

Consider the following priors

$\lambda_r \sim \Gamma(a_r, b_r), r = 1, 2, \ldots, R_0$ (assuming the hazard function to be piecewise continuous)

$\sigma^2 \sim IG(a, b)$

$\beta_0 \sim N(\mathbf{0}, \boldsymbol{A})$, (eigenvalues of $\boldsymbol{A}$ are large enough to justify the improper prior)

$\boldsymbol{\alpha} \sim N(d, \mathrm{diag}(\sigma)), \mathrm{diag}(\sigma) = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$

$\boldsymbol{\xi}_i \sim N(\mathbf{0}, \Sigma)$

$\Sigma^{-1} \sim W(c, C^{-1})$

$\gamma \sim kI_S(\gamma)$, $I_S(\gamma)$ is an indicator function defined on S

$d_l|\sigma_b^2 \sim N(0, \sigma_b^2); l = 1(1)p$

$\sigma_l^{-2}|s_1, s_2 \sim^{\text{ind}} \mathrm{Gamma}(s_1, s_2); l = 1(1)p$

Note that for applying Gibbs sampler, we need the full conditionals given the data and the latent observations.

First of all, the latent observations are generated from

$$\boldsymbol{y}_{ijk}^{\star}|\boldsymbol{\beta}_0, \boldsymbol{\xi}_i, \gamma, \boldsymbol{y}_i \propto N(\mathbf{g}(x_{ij}, \boldsymbol{\alpha}, \boldsymbol{\xi}_i), \sigma^2)\mathbf{1}_{[\gamma_{u_k,ijk-1} < y_{ijk}^{\star} < \gamma_{u_k,ijk}]}$$

Full conditionals can then be formed as follows:

$$\boldsymbol{\alpha}|\boldsymbol{d}, \boldsymbol{\beta}_0, \sigma, \boldsymbol{\xi}, \boldsymbol{y}^{\star}, \boldsymbol{y} \propto \prod_{i=1}^{m}\prod_{j=1}^{n}\left[\sum_{u=1}^{q}\sum_{m_u=1}^{r_u}\mathbf{1}^{ij}_{m_1,m_2,\ldots,m_q}\prod_{k=1}^{q}p_{jk}^{\xi_i}\right]N(\boldsymbol{d}, \mathrm{diag}(\sigma)),$$

$$\boldsymbol{\beta}_0|\boldsymbol{d}, \boldsymbol{\alpha}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{y}^{\star}, \boldsymbol{y} \propto \prod_{i=1}^{m}\prod_{j=1}^{n}\sum_{u=1}^{q}\sum_{m_u=1}^{r_u}\mathbf{1}^{ij}_{m_1,m_2,\ldots,m_q}\prod_{k=1}^{q}p_{jk}^{\xi_i}N(\mathbf{0}, A),$$

$$\boldsymbol{d}|\boldsymbol{\alpha}, \sigma \sim N\left(\Delta^{\star}\mathrm{diag}\left(\sigma_1^{-2}, \sigma_2^{-2}, .., \sigma_p^{-2}\right)\boldsymbol{\alpha}, \Delta^{\star}\right),$$

$$\Delta^{\star} = \mathrm{diag}\left(\sigma_1^{-2}, \sigma_2^{-2}, .., \sigma_p^{-2}\right) + \sigma^{-2}I,$$

$$\boldsymbol{\sigma}^{-2}|\boldsymbol{\alpha}, \boldsymbol{d} \sim^{\mathrm{ind}} \mathrm{Gamma}\left(s_1 + \frac{1}{2}, s_2 + \frac{(\beta_0 - d)^2}{2}\right),$$

$$\boldsymbol{\xi}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, \sigma, \mathbf{y}_i, \mathbf{y}_i^{\star} \propto l_{ij}^{\xi_i}N(\mathbf{0}, \Sigma),$$

$$\Sigma^{-1} \sim W_q\left(\left[C + \sum_{i=1}^{m}\boldsymbol{\xi}_i\boldsymbol{\xi}_i'\right]^{-1}, c + q\right),$$

$$\lambda_r|\alpha, \beta_0, d, \sigma, \xi, y^{\star}, y \sim \mathrm{Gamma}\left(\delta_{ij} + a_r, b_r + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\exp\left[y_{ij}'\beta + z_{ij}'\xi_i\right]S_{ijr}\right),$$

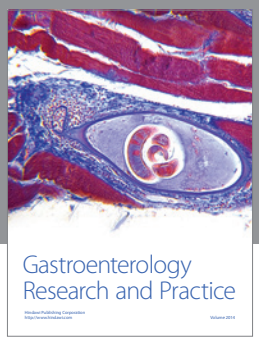
$$S_{ijr} = \exp(\exp[y_{ij}'\beta + z_{ij}'\xi_i]I(t_{ij} > u_{r-1}))\gamma|\alpha, \beta_0, d, \sigma, \xi, y^{\star}, y$$

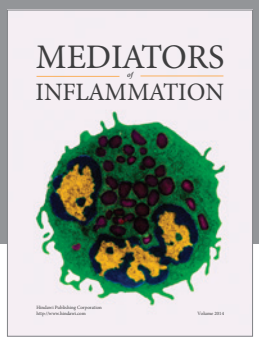
$$\sim U\left(\max\left\{y_i : y_i^{\star} = k\right\}, \min\left\{y_i : y_i^{\star} = k+1\right\}\right).$$

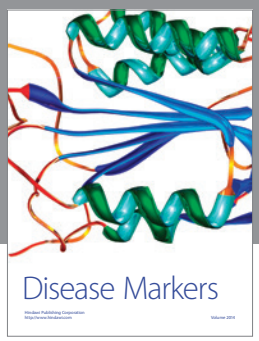
$$\sigma^{-2}|\alpha, \beta_0, d, \xi, y^{\star}, y \sim \mathrm{Gamma}\left(a + \frac{q}{2}\sum_{i=1}^{m}n_i, b + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[y_{ij} - g(x_{ij}, \alpha, \xi_i)\right]\right).$$

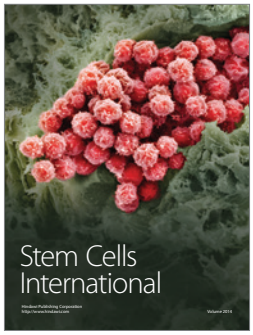
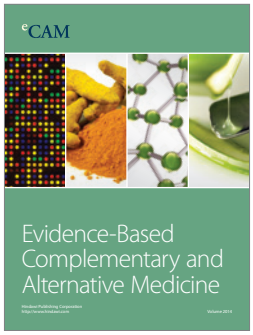
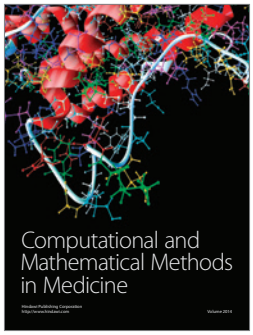
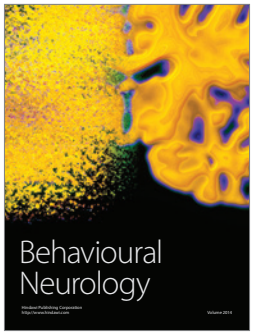
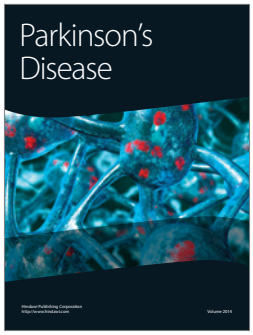
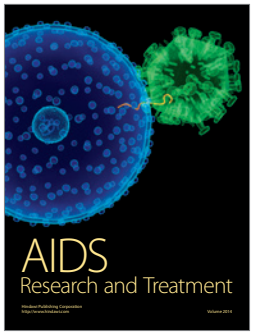
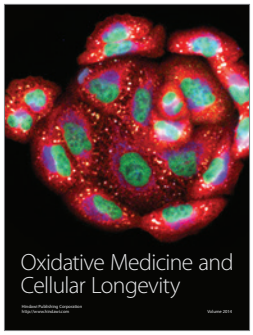
We have considered the following set of values for the hyperparameters:

$$a = 2, b = 1, \sigma_b^2 = 10, s_1 = 1, s_2 = 0.005, c = 13, A, C = \text{diagonal matrix}$$







Hindawi

Submit your manuscripts at
http://www.hindawi.com