

# Generalized Linear Mixed Models for Longitudinal Data with Missing Values: A Monte Carlo EM Approach

Mohamed Y. Sabry, Rasha B. El Kholy, Ahmed M. Gad\*

Statistics Department, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt

**Abstract** Longitudinal data have vast applications in medicine, epidemiology, agriculture and education. Longitudinal data analysis is usually characterized by its complexity due to inter-correlation between repeated measurements within each subject. One way to incorporate this inter-correlation is to extend the generalized linear model (GLM) to the generalized linear mixed model (GLMM) via including the random effects component. When the distribution of the random effects is far from the usual features of the Gaussian density, such as the student's t-distribution, this increases the complexity of the analysis as well as introducing critical features of subject heterogeneity. Thus, statistical techniques based on relaxing normality assumption for the random effects distribution are of interest. This paper aims to find the maximum likelihood estimates of the parameters of the GLMM when the normality assumption for the random effects is relaxed. Estimation is done in the presence of missing data. We assume a selection model for longitudinal data with a dropout pattern. The proposed estimation method is applied to both simulated and real data sets.

**Keywords** Breast Cancer, Longitudinal data, Missing data, Mixed models, Monte Carlo EM

## 1. Introduction

Longitudinal studies are not uncommon in biomedical research and social sciences. In such studies each individual (experimental unit) is measured repeatedly over time for the same response variable (outcome) either under different conditions, or at different times, or both. Longitudinal studies are subject to missing values. The missing values are termed dropout if the subject withdraws from the study at certain time without any observed value for it after that specified time. This is called monotone pattern for missing data. If the subject withdraws occasionally, then we have intermittent missing process and a non-monotone pattern.

In the presence of missing data, the validity of any estimation method requires certain assumptions about the reasons behind the missingness; which are the missing data mechanism. Incorporating the missing data mechanism in statistical model means including an indicator variable,  $R$ , that takes the value 1 if an item is missing and 0 otherwise. Following Rubin's taxonomy [1], the missing data mechanism is said to be missing not at random (MNAR) if  $R$  depends on missing data and may depend on the observed data. In likelihood context this mechanism is labeled as non-ignorable.

The likelihood estimation, especially in the presence of

missing values, is intractable as integration over the random effects is required. Moreover, assuming non-normal distribution for the random effects makes the likelihood more complex and it cannot be expressed in closed form. Various approximation methods to evaluate the likelihood integrations were introduced. These include analytical approximation such as the penalized quasi-likelihood (PQL) and numerical integration such as Gauss-Hermite quadrature. Also a lot of iterative simulation techniques based on the EM algorithm have been used to optimize the likelihood function. The Monte Carlo EM algorithm was introduced to estimate likelihood parameters specially when sampling from conditional distribution of random effects or missing data is difficult.

Violation of the normality assumption of random effects has been studied in literature. [2] uses a mixture of normals for the random effects density. [3] and [4] assume that the random effects distribution has a smooth but unspecified density, that can be estimated using semi-nonparametric representation. [5] replaces the normal distribution in linear mixed models by a skew-t distribution. [6] propose replacing the normal distribution with a mixture of normal distributions where the weights of the mixture components are estimated using a penalized approach. They also use the log-normal and gamma distribution for the random effects density. [7] estimate the maximum likelihood parameters for the logistic model when random effects follow a log normal distribution using a Monte Carlo EM (MCEM) algorithm.

Many studies have been emerged focusing on estimating

\* Corresponding author:

dr\_ahmedgad@yahoo.co.uk (Ahmed M. Gad)

Published online at <http://journal.sapub.org/ijps>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

the GLMM parameters in the presence of missing data. [8] use a shared parameter model where an individual's random slope has been used as a covariate in a probit model for the censoring process. [9] propose conditioning on the time to censoring, and use censoring time as a covariate in random effects model. [10] approximate the generalized linear model using a generalized linear mixed model by conditioning the response variable on the missingness data. [11] develop a Monte Carlo EM algorithm to estimate parameters of GLMM with non-ignorable missing response data, with a normal random effects model. [12] propose a latent class model, which is an alternative approach to a pattern mixture model, by assuming that the dropout time belongs to unobserved (latent) class membership. [13] develop the stochastic EM-algorithm (SEM) algorithm for parameter estimation of a GLM in the presence of intermittent non-random missing values. [14] propose a pseudo likelihood method to estimate parameters for longitudinal data with binary outcome and categorical covariates. They assume missing not at random mechanism and non-monotone pattern. [15] propose a class of unbiased estimating equations using the pairwise conditional technique to estimate the GLMM parameters under non ignorable missingness with a normal random effects model. [16] propose a modification to the weights in the weighted generalized estimating equations to obtain unbiased estimates. [17] propose a heterogeneous random effects covariance matrix, which depends on covariates, using the modified Cholesky decomposition. [18] develop the hierarchical-likelihood method for nonlinear and generalized linear mixed models with arbitrary non-ignorable missing pattern and measurement errors in covariates.

The maximum likelihood estimation of GLMM models in presence of missing data and non-normal distribution of the random effects is still a hot area of research. The aim of this paper is to propose a Monte Carlo EM algorithm to obtain the maximum likelihood estimates of GLMM model for longitudinal data. The focus is on longitudinal data with non-ignorable missingness. A selection model for longitudinal is used, where the responses are not normally distributed.

The rest of the paper is organized as follows. Section 2 gives a description of the used model and notation. Section 3 is devoted to the novel estimation procedure. The performance of the proposed technique is evaluated via simulation studies in Section 4. In Section 5 the proposed technique is applied to a real data example. Finally, Section 6 presents the discussion.

## 2. Model and Notation

The dependent outcome  $y_{ij}$ ,  $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, n_i$  represent the  $j^{\text{th}}$  measurement on the  $i^{\text{th}}$  subject. Their distribution belongs to the exponential family and it can be expressed in a generalized linear model (GLM) as

follows:

$$f(y_{ij}; \theta_{ij}) = \exp\left\{\frac{[y_{ij}\theta_{ij} - a(\theta_{ij})]}{\tau} + d(y_{ij}, \tau)\right\}, \quad (1)$$

where  $a(\cdot)$  and  $d(\cdot)$  are functions defined according to the chosen distribution of exponential family and  $\theta_{ij}$  is called the natural parameter of the distribution. For simplification we assume a balanced design, that is  $n_i = n$ . Let  $\mathbf{y}_i = (\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis})$  where  $\mathbf{y}_i, \mathbf{y}_{i,obs}$  and  $\mathbf{y}_{i,mis}$  are complete, observed and missing outcomes vectors of subject  $i$  respectively. Assuming that we have  $s_i$  missing values for each subject  $i$ , then  $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{in})$ ,  $\mathbf{y}_{i,obs} = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{in-s_i})$  and  $\mathbf{y}_{i,mis} = (\mathbf{y}_{in-s_i+1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{in})$ .

The conditional distribution of  $\mathbf{y}_{ij}$  given the random effects  $\mathbf{b}_i$  has a GLM form and follows the exponential family form as in Eq. (1) with a canonical identity link function, i.e.,  $a(\theta_{ij}) = \theta_{ij}^2/2$ . Thus  $\theta_{ij}$  can be written in the form:  $\theta_{ij} = \mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}_i$ , where  $\mathbf{x}_{ij} = (\mathbf{x}_{ij1}, \mathbf{x}_{ij2}, \dots, \mathbf{x}_{ijp})$  is the  $j^{\text{th}}$  row of the  $n \times p$  design matrix,  $\mathbf{X}_i$ , for vector of fixed effect parameters  $\boldsymbol{\beta}$  and  $\mathbf{z}_{ij} = (\mathbf{z}_{ij1}, \mathbf{z}_{ij2}, \dots, \mathbf{z}_{ijq})$  is the  $j^{\text{th}}$  row in  $\mathbf{Z}_i$ , the design matrix of  $i^{\text{th}}$  subject's random effect,  $\mathbf{b}_i$ . The scale parameter is  $\tau = \theta_{ij}^2$  and

$$d(y_{ij}, \tau) = \frac{-1}{2} \left( \frac{y_{ij}^2}{\tau} + \log(2\pi\tau) \right).$$

The GLMM can be written in the following form:

$$f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \tau) = \exp\left[\left(\frac{y_{ij}\theta_{ij} - a(\theta_{ij})}{\tau}\right) + d(y_{ij}, \tau)\right] \quad (2)$$

$$E(y_{ij}|\mathbf{b}_i) = \hat{a}(\theta_{ij}) = \theta_{ij} = \mu_{ij} \quad (3)$$

$$\mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}_i$$

$$\mathbf{b}_i \sim k(\mathbf{b}_i|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu).$$

We assume that the scale parameter  $\tau$  is constant and equals to one. Hence we can write  $f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \tau) = f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta})$  and  $d(y_{ij}, \tau) = d(y_{ij})$  in Eq. (2). The function  $k(\cdot)$  is the density function of the random effects  $\mathbf{b}_i$  with non-central mean  $\boldsymbol{\mu}_0$ , variance covariance matrix  $\boldsymbol{\Sigma}$  and  $\nu$  degrees of freedom.

Note that the responses at all occasions  $j=1, \dots, n$  for any subject  $i$  are conditionally independent given the common random effect  $\mathbf{b}_i$ . Thus, the likelihood for  $N$  subjects with complete data ( $\mathbf{s}_i = \mathbf{0}$ ) can be written as follows:

$$L(\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu | \mathbf{Y}, \mathbf{b}) = \prod_{i=1}^N \prod_{j=1}^n f(y_{ij}|\boldsymbol{\beta}, \mathbf{b}_i) k(\mathbf{b}_i|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu) \quad (4)$$

The marginal likelihood of parameters  $\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu$  is proportional to the marginal density function  $f(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu)$ . This can be obtained by integrating out the random effect, i.e.

$$f(\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu) = \int f(\mathbf{Y}, \mathbf{b}|\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \nu) d\mathbf{b} \quad (5)$$

The integration in Eq. (5) is of high dimension ( $N \times q$ ). Thus evaluating the likelihood is cumbersome even with complete data. Moreover, in most cases this integration is intractable, specially, when the normality assumption is violated.

When the missing data mechanism is nonignorable, it is required to incorporate the parametric model for the missingness mechanism into the log-likelihood of complete data. Let the random vector  $\mathbf{r}_i \in \mathbf{R}^n$  be the missing data mechanism indicator whose  $j$ th component  $r_{ij}$  equals to 1 when the response  $y_{ij}$  is missing, otherwise it equals to zero. Under the selection model [19], the distribution of  $\mathbf{r}_i$  given the response  $\mathbf{y}_i$  is indexed with a parameter vector  $\boldsymbol{\phi}$  and has a multi-nominal distribution with  $2^n$  cell probabilities, i.e.

$$\Pr(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\phi}) = \prod_{j=1}^n [\Pr(r_{ij} = \mathbf{1} | \boldsymbol{\phi})]^{r_{ij}} [1 - \Pr(r_{ij} = \mathbf{1} | \boldsymbol{\phi})]^{1-r_{ij}}. \quad (6)$$

Thus, the complete density for subject  $i$  is denoted by

$$\begin{aligned} f(\mathbf{y}_i, \mathbf{b}_i, \mathbf{r}_i | \boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \boldsymbol{\phi}) \\ = f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) k(\mathbf{b}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \Pr(\mathbf{r}_i | \boldsymbol{\phi}, \mathbf{y}_i), \end{aligned} \quad (7)$$

where  $f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i) = \prod_{j=1}^n f(y_{ij} | \boldsymbol{\beta}, \mathbf{b}_i)$ .

The classical normality assumption of the random effects is not always realistic. The t-distribution offers a more viable alternative with respect to real-world data. Assuming a central multivariate t-distribution ( $\boldsymbol{\mu}_0 = \mathbf{0}$ ) for the random effect,  $\mathbf{b}_i$ , the density function  $k(\cdot)$  is of the form

$$k(\mathbf{b}_i | \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \frac{[\Gamma(\boldsymbol{\nu}+q)/2] [1 + \frac{1}{\boldsymbol{\nu}} \mathbf{b}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{b}_i]^{-(\boldsymbol{\nu}+q)/2}}{\Gamma(\frac{\boldsymbol{\nu}}{2}) \boldsymbol{\nu}^{q/2} \pi^{q/2} |\boldsymbol{\Sigma}|^{1/2}} \quad (8)$$

The degrees of freedom is assumed to be known or unknown.

From Eq. (7), the complete-data log-likelihood can be defined as:

$$\ell(\boldsymbol{\gamma}) = \sum_{i=1}^N \log[f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i)] + \log[k(\mathbf{b}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu})] + \log[\Pr(\mathbf{r}_i | \boldsymbol{\phi}, \mathbf{y}_i)], \quad (9)$$

where  $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu}, \boldsymbol{\phi})$  is the set of all parameters.

### 3. Likelihood Estimation

In this section, the fixed effect parameters,  $\boldsymbol{\beta}$ , the random effects parameters,  $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu})$  and the missing data model's parameters,  $\boldsymbol{\phi}$  are estimated. A monotone pattern of missing data for the outcome variable  $\mathbf{y}_i$  is assumed, where  $\mathbf{y}_i = (\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis})$ . The random effects of the  $i$ th subject,  $\mathbf{b}_i$ , are unobserved, so they are also considered as missing data. Thus, in E-step of EM algorithm we need to integrate the complete-data log-likelihood in Eq. (9) over the missing data  $\mathbf{y}_i$  and  $\mathbf{b}_i$ . In other words, in the E step the expectation of the log-likelihood function of complete data given the observed data and current estimate of parameters is obtained. The expectation is with respect to the joint distribution of missing data  $\mathbf{y}_i$  and  $\mathbf{b}_i$ . Starting with initial values  $(\boldsymbol{\beta}^{(0)}, \boldsymbol{\mu}_0^{(0)}, \boldsymbol{\Sigma}^{(0)}, \boldsymbol{\nu}^{(0)}, \boldsymbol{\phi}^{(0)})$ , the  $(t+1)$ th iteration of the EM algorithm proceeds as follows:

- **The E-step:** Obtain the expectations with respect to the conditional distribution of  $f(\mathbf{b}_i, \mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i)$  given in items 1 and 2 below. Also, obtain the conditional expectation with

respect to conditional distribution of  $f(\mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{b}_i, \mathbf{r}_i)$  given in item 3 below.

1.  $E[\log f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}) | \mathbf{y}_{i,obs}, \boldsymbol{\beta}^{(t)}]$
  2.  $E[\log k(\mathbf{b}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu}) | \mathbf{y}_{i,obs}, \boldsymbol{\mu}_0^{(t)}, \boldsymbol{\Sigma}^{(t)}, \boldsymbol{\nu}^{(t)}]$
  3.  $E[\log \Pr(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\phi}) | \mathbf{y}_i, \boldsymbol{\phi}^{(t)}]$
- **The M step:** Find the values  $\boldsymbol{\beta}^{(t+1)}$  that maximize the function in item 1 above. Also, the values  $\boldsymbol{\mu}_0^{(t+1)}, \boldsymbol{\Sigma}^{(t+1)}, \boldsymbol{\nu}^{(t+1)}$  that maximize the function in item 2 above and  $\boldsymbol{\phi}^{(t+1)}$  that maximize the function in item 3 above.

If convergence is achieved, the current values are the maximum likelihood estimates. Otherwise, let  $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t+1)}$ ,  $\boldsymbol{\Sigma}^{(t)} = \boldsymbol{\Sigma}^{(t+1)}$ ,  $\boldsymbol{\mu}_0^{(t)} = \boldsymbol{\mu}_0^{(t+1)}$ ,  $\boldsymbol{\nu}^{(t)} = \boldsymbol{\nu}^{(t+1)}$ ,  $\boldsymbol{\phi}^{(t)} = \boldsymbol{\phi}^{(t+1)}$  and return to the E-step.

The expectation of the log-likelihood for the  $i$ th subject at iteration  $(t+1)$  is

$$\begin{aligned} Q_i(\boldsymbol{\gamma} | \boldsymbol{\gamma}^{(t)}) &= E[\ell(\boldsymbol{\gamma}; \mathbf{y}_i, \mathbf{b}_i, \mathbf{r}_i | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i)] \\ &= \iint \log[f(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{b}_i)] f(\mathbf{b}_i, \mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i) d\mathbf{b}_i d\mathbf{y}_{i,mis} + \\ &\quad \iint \log[k(\mathbf{b}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\nu})] f(\mathbf{b}_i, \mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i) d\mathbf{b}_i d\mathbf{y}_{i,mis} + \\ &\quad \iint \log[\Pr(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\phi})] f(\mathbf{b}_i, \mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i) d\mathbf{b}_i d\mathbf{y}_{i,mis} \end{aligned} \quad (10)$$

where  $f(\mathbf{b}_i, \mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i)$  is the conditional distribution of missing data  $\mathbf{b}_i, \mathbf{y}_{i,mis}$  given the observed data and the current parameter estimates. The integral in Eq. (10) is intractable. Therefore, expectations are obtained via Monte Carlo simulation where sampling from  $f(\mathbf{b}_i, \mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i)$  is required.

The following two relations describe the complete conditional distributions

$$f(\mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{b}_i, \mathbf{r}_i) \propto f(\mathbf{r}_i | \boldsymbol{\phi}^{(t)}, \mathbf{y}_i) f(\mathbf{y}_i | \boldsymbol{\beta}^{(t)}, \mathbf{b}_i) \quad (12)$$

$$\begin{aligned} f(\mathbf{b}_i | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}, \mathbf{r}_i) \\ \propto f(\mathbf{b}_i | \boldsymbol{\Sigma}^{(t)}, \boldsymbol{\mu}_0^{(t)}, \boldsymbol{\nu}^{(t)}) f(\mathbf{y}_i | \boldsymbol{\beta}^{(t)}, \mathbf{b}_i) \end{aligned} \quad (13)$$

Sampling from the first complete conditional distribution in Eq. (11) is equivalent to sampling from  $f(\mathbf{y}_{i,mis} | \boldsymbol{\gamma}^{(t)}, \mathbf{y}_{i,obs}, \mathbf{r}_i)$  because the missingness mechanism doesn't depend on the random effects. A single draw  $\mathbf{y}_{i,mis}^*$  for the first dropout of the  $i$ th subject is generated from the predictive distribution  $P(\mathbf{y}_{i,mis}^* | \mathbf{y}_{i,obs}, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\eta}^{(t)})$  instead of  $P(\mathbf{y}_{i,mis}^* | \mathbf{y}_{i,obs}, \mathbf{b}_i, \boldsymbol{\eta}^{(t)})$  where  $\boldsymbol{\eta}^{(t)}$  are the current parameter estimates. Then an accept-reject sampling procedure is used to sample  $\mathbf{y}_{i,mis}$ . For more details, interested readers are referred to [20]. From properties of a multivariate normal,  $P(\mathbf{y}_{i,mis}^* | \mathbf{y}_{i,obs}, \mathbf{b}_i, \boldsymbol{\eta}^{(t)})$  is a normal distribution with mean  $\boldsymbol{\mu}_{i,m.o}$  and covariance matrix  $\mathbf{V}_{i,m.o}$ , where

$$\boldsymbol{\mu}_{i,m.o} = \boldsymbol{\mu}_{i,m} - \mathbf{V}_{i,m,o} \mathbf{V}_{i,o,o}^{-1} (\mathbf{y}_{i,obs} - \boldsymbol{\mu}_{i,o}) \quad (13)$$

$$\mathbf{V}_{i,m.o} = \mathbf{V}_{i,m,m} - \mathbf{V}_{i,m,o} \mathbf{V}_{i,o,o}^{-1} \mathbf{V}_{i,o,m} \quad (14)$$

The  $\boldsymbol{\mu}_{i,o}$ ,  $\boldsymbol{\mu}_{i,m}$ ,  $\mathbf{V}_{i,m,o}$ ,  $\mathbf{V}_{i,m,m}$  and  $\mathbf{V}_{i,o,o}$  are suitable

partitions from mean  $\mu$  and covariance matrix  $V$  respectively.

Sampling from the second complete conditional distribution in Eq. (12) is conducted using the Metropolis algorithm. A candidate value  $\mathbf{b}_i^*$  is generated from a proposal distribution which is chosen to be the density function of the random effects (the multivariate t). The generated value is accepted as opposed to keep the previous value with the probability:

$$A(\mathbf{b}_i, \mathbf{b}_i^*) = \min\left(1, \frac{\prod_{j=1}^n f(y_{ij} | \mathbf{b}_i^*, \boldsymbol{\beta}, \sigma)}{\prod_{j=1}^n f(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \sigma)}\right), \quad (15)$$

where  $f(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \sigma)$  is a univariate normal and could be easily evaluated. The acceptance probability can be simplified as

$$A(\mathbf{b}_i, \mathbf{b}_i^*) = \min\left(1, \frac{\prod_{j=1}^n \sigma_{ij}^{*-0.5} \times \exp\left(-\frac{\delta^*}{2\sigma_{ij}^{*2}}\right)}{\prod_{j=1}^n \sigma_{ij}^{-0.5} \times \exp\left(-\frac{\delta}{2\sigma_{ij}^2}\right)}\right), \quad (16)$$

where

$$\delta^* = \sum_{j=1}^n (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta} - \mathbf{z}_{ij}\mathbf{b}_i^*)^2$$

and

$$\delta = \sum_{j=1}^n (y_{ij} - \mathbf{x}_{ij}\boldsymbol{\beta} - \mathbf{z}_{ij}\mathbf{b}_i)^2$$

The  $y_{ij}$  is  $j$ th row of  $\mathbf{y}_j$  and  $\mathbf{x}_{ij}$  is the  $j$ th row in matrix  $\mathbf{X}_i$  of dimension  $1 \times n$  and  $\mathbf{z}_{ij}$  is the  $j$ th row in matrix  $\mathbf{Z}_i$  of dimension  $1 \times q$ ,

$$\sigma_{ij}^2 = \mathbf{I}_{jj} + \mathbf{z}_{ij}\boldsymbol{\Sigma}_i\mathbf{z}_{ij}' \quad (17)$$

$$\sigma_{ij}^{*2} = \mathbf{I}_{jj} + \mathbf{z}_{ij}\boldsymbol{\Sigma}_i^*\mathbf{z}_{ij}', \quad (18)$$

where  $\mathbf{I}$  is an identity matrix of dimension  $n \times n$  and  $\boldsymbol{\Sigma}_i$  is the variance covariance matrix of random effects component  $\mathbf{b}_i$  of dimension  $q \times q$ .

**Table 1.** Parameter Estimates and their Corresponding Standard Errors (simulation study)

Par.	Actual	Estimate	Relative Bias%	SE
$\beta_1$	-0.7	-0.615	-12.12	0.302
$\beta_2$	1	1.011	1.18	0.123
$\beta_3$	1.2	1.241	3.47	0.125
$\rho_t$	0.5	0.726	45.24	0.063
$\sigma_t$	1.4	1.181	-15.62	0.310
$\phi_1$	-1.8	-2.639	16.62	0.348
$\phi_2$	-0.24	-0.273	13.82	0.109
$\phi_3$	0.15	0.129	-13.83	0.091

### 4. Simulation Study

The aim of this simulation study is to evaluate the performance of the proposed techniques in Section (3). The number of subjects is assumed to be  $N = 30$  subjects. The

number of time points is assumed to be  $m = 7$  time points. These numbers are chosen arbitrary for the sake of time. Hence, the response values are  $y_{ij}, i = 1, 2, \dots, 30, j = 1, 2, \dots, 7$ . The random effects  $\mathbf{b}_i, i = 1, 2, \dots, 30$ , were generated from the multivariate t-distribution with dimension  $q = 4$  and known degrees of freedom  $\nu = 4$ . The response values were generated, conditionally on the random effect, with variance covariance matrix  $\boldsymbol{\Sigma}_i$  which is assumed to have a first order autoregressive correlation structure AR(1).

The responses of the  $i$ th are modeled using the GLMM as

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \mathbf{z}_i \mathbf{b}_i + e_{ij}, \quad (19)$$

where  $\mathbf{b}_i = (b_1, b_2, b_3, b_4)'$  is a subject random effects. The  $\mathbf{z}_i$  is a vector of ones  $\in R^4$ . The noise component  $e_{ij}$  follows a multivariate normal  $MVN(\mathbf{0}, \mathbf{I}_7)$ . The  $x_{kij}, k = 1, 2, 3; i = 1, 2, \dots, N; j = 1, 2, \dots, m$  are covariates (design matrix) associated with each subject. The values of  $x_{kij}$  are chosen as follows:

$$\begin{aligned} \mathbf{X}_{1i} &= \begin{cases} 2 & \text{for odd observations, and} \\ 1 & \text{for even observations} \end{cases} \\ \mathbf{X}_{2i} &= \begin{cases} (1, 0, 1, 0, 1, 0, -1)' & \text{for } i = 1, 3, 6, 8 \\ (0, 1, 0, 1, 0, 1, -1)' & \text{for } i = 2, 4, 5, 7 \\ (-1, 0, -1, 1, 1, 1, 0)' & \text{for } i = 9, 10, 11, 12 \\ (0, 0, 0, 0, 0, 0, 0)' & \text{for } i > 12 \end{cases} \\ \mathbf{X}_{3i} &= \begin{cases} (0, 1, 0, 1, 0, 1, -1)' & \text{for } i = 1, 3, 6, 8 \\ (1, 0, 1, 0, 1, 0, -1)' & \text{for } i = 2, 4, 5, 7 \\ (-1, 1, -1, 0, 0, 0, 1)' & \text{for } i = 9, 10, 11, 12 \\ (0, 0, 0, 0, 0, 0, 0)' & \text{for } i > 12 \end{cases} \end{aligned}$$

The missing data process is modeled as a logistic model, where the missing data mechanism depends on the current and previous outcome as:

$$\text{logit}\{\Pr(r_{ij} = 1 | \boldsymbol{\phi})\} = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 y_{ij-1} + \boldsymbol{\phi}_2 y_{ij}. \quad (12)$$

It is assumed that  $\mathbf{y}_{i0} = \mathbf{0}$  for all  $i$ . Also,  $r_{ij}$  are assumed to be independent for all  $(i, j)$  given  $\mathbf{y}_{ij-1}$  and  $\mathbf{y}_{ij}$ . From Eq. (6) the missing data model for all observations can be written as:

$$\Pr(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\phi}) = \prod_{i=1}^{30} \prod_{j=1}^{pos_i} [\Pr(r_{ij} = 1 | \boldsymbol{\phi})]^{r_{ij}} [1 - \Pr(r_{ij} = 1 | \boldsymbol{\phi})]^{1-r_{ij}} \quad (21)$$

where  $pos_i$  is the occasion at the first dropout for subject  $i$ .

The parameter values are fixed as  $\beta_1 = -0.7, \beta_2 = 1, \beta_3 = 1, \rho_t = 0.7, \sigma_t = 1.4, \phi_1 = -1.8, \phi_2 = -0.24$  and  $\phi_3 = 0.15$ .

Table 1 shows the maximum likelihood estimates obtained using the MCEM algorithm, and the corresponding standard errors. 3000 samples were generated from missing data (missing outcome and random effect) with a burn-in period 500 iterations. Percentage of complete subjects is around 50%. As can be seen from Table 1, all parameters estimates are highly significant. All parameter estimates do not exceed a relative bias of 30% except  $\rho_t$  its bias percentage is 43.14%, where the relative bias is defined as the ratio of bias with respect to actual values. Small values for standard errors show that we have good results, which are lying in confidence interval 95%.

All parameters of missing mechanism are significant, while parameter estimates of missing-mechanism for previous and current outcome are very significant, which support that the missing data-mechanism is non-ignorable.

### 5. Breast Cancer Data

This data set is related to quality of life among premenopausal women diagnosed with breast cancer. It is prepared via a clinical trail taken by the International Breast Cancer Study Group (IBCSG) [21]. The patients are followed for death, relapse and quality of life. The patients were randomized to one of four chemotherapy regimes. The patients were asked to answer a quality of life questionnaire at baseline (before starting treatment) and every three months for fifteen months. Therefore, each questionnaire should be answered 6 times in the specified period.

**Table 2.** Fixed parameters estimates and standard errors at different initial values (breast cancer data)

Par.	Initial values	Estimate	SE	P-value
$\beta_1$	0.4	-0.13	0.3308	< 0.0001
$\beta_2$	0.6	0.88	0.3144	< 0.0001
$\beta_3$	0.9	-0.54	0.3342	< 0.0001
$\beta_1$	0.8	-0.13	0.3309	< 0.0001
$\beta_2$	1	1.02	0.3145	< 0.0001
$\beta_3$	1.3	-0.77	0.3343	< 0.0001
$\beta_1$	0.6	-0.18	0.3294	< 0.0001
$\beta_2$	0.8	0.99	0.3131	< 0.0001
$\beta_3$	1.1	-0.71	0.3328	< 0.0001
$\beta_1$	0.3	-0.19	0.3308	< 0.0001
$\beta_2$	0.5	0.85	0.3144	< 0.0001
$\beta_3$	0.8	-0.83	0.3342	< 0.0001

One of the effective determinants of quality of life was the Perceived Adjustment to Chronic Illness Scale (PACIS). This measures the amount of efforts the patient could bear to cope with her illness. The larger the score of the measure the greater the effort is done by patient to cope with her illness. It has a continuous scale from 0 as a best quality of life to 100 as the worst one. Assessment has been established for patients who remained alive during the study, so the missing data is not due to the death. Ten patients were died within the period are excluded from the study. Four hundreds and forty six patients have survived the period. Also, 64 cases have been removed as their missing responses were at first assessment. Thus, the number of patients in the trial is 382 subjects. A patient may refuse to answer the questionnaire on occasion and she is asked to answer the questionnaire in the next follow-up occasion. Therefore, the pattern of missingness in the trial is a combination of dropout and intermittent. The patient may refuse to answer the questionnaire, if her mood is poor, so the missingness mechanism is nonrandom. The percentage of complete subjects is 23%. Regarding the missing data, the percentage

of patients with dropout pattern is 63%, while the percentage with intermittent pattern is 37%.

These data have been used in previous researches as follows: the first 9 months of the study were analyzed with only complete responses (complete case analysis) by [21]. An analysis for the first 6 months, taking into consideration the missing data, has been conducted by [22]. [11] have analysed the patient's self mood for 18 months of the study. They used normal random effects model with AR(1) covariance structure and logistic model for missing data mechanism. [24] used both AR(1) and unstructured covariance structure with hybrid pattern of missing data (intermittent and dropout missing).

In this paper, we assume a central multivariate t with known degrees of freedom (assumed to be 4) with dropout pattern of missing and AR(1) covariance structure. The intermittent missing values have been deleted, therefore, the number of patients in the trail becomes 284 patients. Also, following [21] a square root transformation for assessments is used to normalize data. A GLMM model is used to link the response variable to the treatments and the random effects with  $i = 1, 2, \dots, 284, j = 1, 2, \dots, 6$  and treatments  $x_1, x_2, x_3$  are defined as follows:

$$(x_1, x_2, x_3) = \begin{cases} (1, 0, 0) & \text{for treatment A,} \\ (0, 1, 0) & \text{for treatment B,} \\ (0, 0, 1) & \text{for treatment C,} \\ (-1, -1, -1) & \text{for treatment D.} \end{cases}$$

The MCEM is implemented using 3000 samples with a bur-in period of 500 iterations.

**Table 3.** Parameters estimates and standard errors at different initial values (breast cancer data)

Par.	Initial values	Estimate	SE	P-value
$\beta_1$	0.6	-0.18	0.3294	< 0.0001
$\beta_2$	0.8	0.99	0.3131	< 0.0001
$\beta_3$	1.1	-0.71	0.3328	< 0.0001
$\rho_t$	0.7	-0.13	0.3309	< 0.0001
$\sigma_t$	1	1.21	0.1300	< 0.0001
$\phi_1$	-1.6	-1.88	0.2006	< 0.0001
$\phi_2$	-0.24	-0.18	0.3294	< 0.0001
$\phi_3$	0.18	-0.25	0.0280	< 0.0001
$\beta_1$	0.3	-0.29	0.3308	< 0.0001
$\beta_2$	0.5	0.85	0.3144	< 0.0001
$\beta_3$	0.8	-0.83	0.3342	< 0.0001
$\rho_t$	0.4	0.99	0.0001	< 0.0001
$\sigma_t$	0.8	1.23	0.1326	< 0.0001
$\phi_1$	-1.2	-1.91	0.2002	< 0.0001
$\phi_2$	-0.4	0.26	0.0319	< 0.0001
$\phi_3$	0.5	-0.24	0.0274	< 0.0001

The proposed algorithm is implemented using several starting values of the parameters. The maximum likelihood estimates of the parameters are presented in Tables 2 where only different starting points are used for fixed effect parameters. The starting points were  $\rho = 0.7, \sigma = 1,$

$\phi_1 = -1.6$ ,  $\phi_2 = -0.2$  and  $\phi_3 = 0.2$ . The estimates of  $\rho$ ,  $\sigma$ ,  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  were around 0.99, 1.21, -1.90, 0.26, and -0.24 respectively. The results in Table 3 are the maximum likelihood estimates using different starting points for all parameters.

Results show that the algorithm converges properly as the parameter estimates are similar for the different starting values. Also, all estimates are highly significant. Highly significant coefficients of treatments are indication of strong relationship between the response variable (PACIS) and the treatments. Given, that any patient receives only one treatment during the trail, the higher doze from treatment A or D, the smaller PACIS value i.e. a better quality of life the patient gets. On the other hand, the higher doze from treatment B or C, the higher PACIS value, i.e. patient endure hardship to cope with her life. Also, the significance of  $\phi_2$  and  $\phi_3$  supports that the missing data mechanism is non-ignorable.

## 6. Discussion

In this paper, a Monte Carlo EM algorithm is proposed to estimate the likelihood parameters of a longitudinal model in GLMM framework. We used a selection model for longitudinal data with non-ignorable missing mechanism in the context of dropout pattern. In this case the obtained likelihood function is intractable and it is difficult to integrate over the missing data (random effect and missing values of outcomes). Therefore a Monte Carlo EM algorithm is implemented. In the context of the proposed model, direct simulation is not possible because there is no defined mathematical form for the joint density of missing data (random effect and missing outcomes) given the observed data and current estimate of parameters. Thus, Monte Carlo Sampling techniques are used; Metropolis-Hastings and reject-accept techniques. Hessian matrix is used to estimate standard errors. The proposed algorithm is validated using a simulation study. Also, the proposed algorithm is applied to a real data set of breast cancer.

## REFERENCES

- [1] Rubin, D. B., 1976, Inference and missing data, *Biometrics*, 63, 581-592.
- [2] Caffo, B., An, M. W. and Rohde, C., 2007, Flexible random intercept models for binary outcomes using mixtures of normals, *Computational Statistics and Data Analysis*, 51, 5220-5235.
- [3] Zhang, D. and Davidian, M., 2001, Linear mixed models with flexible distributions of random effects for longitudinal data, *Biometrics*, 57(3), 795-802.
- [4] Chen, J., Zhang, D. and Davidian, M., 2002, A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution, *Biostatistics*, 3, 347-360.
- [5] Zhou, T. and He, X., 2008, Three-step estimation in linear mixed models with skew t-distributions, *Journal of Statistical Planning and Inference*, 138, 1542-1555.
- [6] Komarek, A. and Lesaffre, E., 2008, Generalized linear mixed model with a penalized Gaussian mixture as random effects distribution, *Computational Statistics and Data Analysis*, 52, 3441-3458.
- [7] Gad, A. M. and El Kholy R. B., 2012, Generalized linear mixed models for longitudinal data, *International Journal of Probability and Statistics*, 1(3), 41-47.
- [8] Wu, M. C. and Carroll, R. J., 1988, Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process, *Biometrics*, 44, 175-188.
- [9] Wu, M. C. and Bailey, K. R., 1989, Estimation and comparison of changes in the presence of informed censoring: Conditional linear model, *Biometrics*, 45, 939-955.
- [10] Follmann, D. and Wu, M., 1995, An approximate generalized linear model with random effects, *Biometrics*, 51, 151-168.
- [11] Ibrahim, J. G., Lipstiz, S. R. and Chen, M. H., 2001, Missing Covariates in generalized linear models when the missing data mechanism is non-ignorable, *Journal of the Royal Statistical Society B*, 61(1), 173-190.
- [12] Roy, J., 2003, Modeling longitudinal data with non-ignorable dropouts using a latent dropout class model, *Biometrics*, 59, 829-836.
- [13] Gad, A. M. and Ahmed, A. S., 2006, Analysis of longitudinal data with intermittent missing values using the stochastic EM algorithm, *Computational Statistics and Data Analysis*, 50, 2702-2714.
- [14] Parzen, M., Lipstiz, S. R., Fitzmaurice, J. G., Ibrahim, J. G. and Andrea T., 2006, Pseudo-likelihood methods for longitudinal binary data with non-ignorable responses and covariates, *Statistics in Medicine*, 25, 2784-2796.
- [15] Zhang, H. and Paik M. C., 2008, Handling missing responses in generalized linear mixed model without specifying missing mechanism., *Journal of Biopharmaceutical statistics*, 19, 6--15.
- [16] Brajendra, C. S. and Taslim, S. M., 2010, Modified weights based generalized quasi-likelihood inferences in incomplete longitudinal binary models, *The Canadian Journal of Statistics*, 38, 217-231.
- [17] Lee, K., Lee, J., Hogan, J. and Yoo, K., 2011, Modeling the random effects covariance matrix for generalized linear mixed models, *Computational Statistics and Data Analysis*, 56, 1545-1551.
- [18] Noh, M., Wu, L. and Lee, Y., 2012, Hierarchical likelihood methods for nonlinear and generalized linear mixed models with missing data and measurement errors in covariates, *Journal of Multivariate Analysis*, 109, 42-51.
- [19] Heckman, J. J., 1976, The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of Economics and Social Measurement*, 5, 475-492.
- [20] Gad, A. M. and Youssef, N. A., 2006, Linear mixed models for longitudinal data with nonrandom dropouts, *Journal of*

Data Science, 4, 447 - 460.

- [21] Hurny, C., Bernard, J. G., Gelber, R. D., Coates, A., Gastiglione, M., Isley, M., Dreher, D., Peterson, H., Goldhirsch, A., and Senn, H. J., 1992, Quality of life measures for patients receiving adjuvant therapy for breast cancer: an international trial, *European Journal of Cancer*, 28, 118-124.
- [22] Troxel, A. B., Harrington, D. P. and Lipstiz, S. R., 1998, Analysis of longitudinal data with non-ignorable non-monotone missing values, *Applied Statistics*, 47, 425-438.
- [23] Gad, A. M., 2011, A selection model for longitudinal data with non-Ignorable non-monotone missing values, *Journal of Data Science*, 9, 171-180.