

Fitting Multivariate Linear Mixed Model for Multiple Outcomes Longitudinal Data with Non-ignorable Dropout

Ahmed M. Gad*, Niveen I. EL-Zayat

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

Abstract Measurements made on several outcomes for the same unit, implying multivariate longitudinal data, are very likely to be correlated. Therefore, fitting such a data structure can be quite challenging due to the high dimensioned correlations exist within and between outcomes over time. Moreover, an additional challenge is encountered in longitudinal studies due to premature withdrawal of the subjects from the study resulting in incomplete (missing) data. Incomplete data is more problematic when missing data mechanism is related to the unobserved outcomes implying what so-called non-ignorable missing data or missing not at random (MNAR). Obtaining valid estimation under non-ignorable assumption requires that the missing-data mechanism be modeled as a part of the estimation process. The multiple continuous outcome-based data model is introduced via the Gaussian multivariate linear mixed models while the missing-data mechanism is linked to the data model via the selection model such that the missing-data mechanism parameters are fitted using the multivariate logistic regression. This article proposes and develops the stochastic expectation-maximization (SEM) algorithm to fit MLMM in the presence of non-ignorable dropout. In the M-step maximizing likelihood function is implemented via a new proposed Quasi-Newton (QN) algorithm that is of EM type, while maximizing the multivariate logistic regression is implemented via Newton-Raphson (NR) algorithm. A simulation study is conducted to assess the performance of the proposed techniques.

Keywords Non-ignorable missing, Selection models, Multivariate linear mixed models, Stochastic EM algorithm, Newton-Raphson, Fisher-Scoring, Quasi-Newton

1. Introduction

Longitudinal studies are very common in many fields such as medicine, public health, psychology, biology and more. In the simplest design of univariate longitudinal data a single response is collected repeatedly over time, or possibly under changing experimental conditions, on each subject. A fundamental feature of longitudinal data is that the repeated measurements of the same subject tend to be highly correlated, and even more, such correlations often change over time resulting a complicated covariance structure (Diggle et al, 2002). In many applications it is common to measure a set of several responses, m , at each occasion for subjects which result in multivariate longitudinal data. The relationship among those outcomes is crucial. It requires addressing special methods of statistical analysis that can properly account for the intra-subject correlation as well as the cross-correlation between outcomes since ignoring such

association structure will lead to invalid inferences. Modeling repeated measures on multivariate continuous responses is often implemented via multivariate linear mixed models. The multivariate linear mixed models are more applicable in this area due their flexibility in allowing (i) unbalanced data where a number of repeated measures might differ within subjects per outcome, (ii) using different design matrix across responses, and (iii) modeling distinct and more complex covariance structures. For a comprehensive monograph of multivariate linear mixed models see for example, Demidenko(2004) and Kim and Tim (2007).

Missing data are very common in longitudinal data studies. Dealing with missing data depends on the process that generates the missing values; the missing data mechanism (MDM). A general taxonomy of MDMs, originally, introduced by Rubin (1976) and later explained by Little and Rubin (2002), differ in terms of assumptions about whether missingness is related to observed and/or unobserved responses. When dropout mechanism does not affected by the neither observed nor unobserved response values, it is called missing completely at random (MCAR). Missing data is said to be missing at random (MAR) when the missingness depends on the observed responses. Finally, missing data is missing not at random (MNAR) when missingness depends on the unobserved response. Nevertheless, it might be necessary to accommodate such type of missingness in the

* Corresponding author:

Ahmed.gad@feps.edu.eg (Ahmed M. Gad)

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

Copyright © 2018 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

modeling process to reduce estimation biases (Little and Rubin, 2002).

Likelihood-Based approach has been established in literature to be one of the most widely-used modern approaches of handling missing data. Parameter estimation has received a serious attention in the missing data literature. However, with non-ignorable assumption, such estimates cannot be obtained unless approaches are developed to model the missing data mechanism (Endres, 2010; Graham, 2012). Therefore, a number of selection model have been proposed to model longitudinal data in the presence of missing values aiming to characterize the joint distribution of data and the probability of missingness (Stubbendick, 2003; Little and Rubin, 2002). Direct maximization of the likelihood function, in this case, is not feasible (Jasson, and Xihong, 2002). So, numerical and iterative computations are exceedingly desired. Three generic iterative approaches such as Newton-Raphson (NR), Fisher Scoring (FS), Quasi-Newton (QN) and Expectation-Maximization (EM). EM algorithm can be used to maximize the likelihood function. The EM algorithm is a generic iterative approach to find the maximum likelihood estimates (MLEs) for the model parameters when there is missing data or when the model contains unobserved latent variables (Demidenko, 2004).

The multivariate linear mixed models (MLMM) have received more attention in literature particularly under non-ignorable MD. Reinsel (1982, 1984) considered fitting the MLMM by a closed form with complete and balanced data. While Shah *et al.* (1997) extended the work of Laird and Ware (1982) to the bivariate setting ($m=2$) using the EM algorithm to obtain the estimates of the model parameters under unbalanced and ignorable missing data. Schafer and Yucel (2002) and later Yucel (2015) proposed a variant technique of EM algorithm again for estimating the MLMM with ignorable missing data. Their EM algorithm considers the fisher scoring procedure in the M-step which fastens the convergence. Under non-ignorable assumption Roy and Lin (2002) consider fitting the MLMM under non-ignorable assumption covering the missingness in covariates as well as the multiple responses. They assumed that the dropout process depends on the latent variable and applied the selection model in order to account for non-ignorable missing data. Luwanda and Mwambi (2016) considered fitting nonlinear mixed-effects models to fit the multivariate longitudinal data in the presence of non-ignorable dropout. They proposed the stochastic approximation EM (SAEM) algorithm and apply the proposed technique to estimate parameters characterising human immunodeficiency virus (HIV) disease dynamics.

The aim of this article is to address statistical modeling of the multivariate longitudinal data with continuous responses via the Gaussian multivariate linear mixed models (MLMM) under non-ignorable incomplete data. The missing-data mechanism is linked to the data model via the selection model such that the missing-data mechanism parameters are fitted using the multivariate logistic regression. The article

proposes a stochastic EM algorithm to estimate the parameters of the model.

The paper is organized as follow. The Multivariate linear mixed model for multivariate longitudinal data is presented in Section 2. This section also presents the dropout model for multivariate longitudinal data and formulates the joint distribution of the dropout mechanism and multivariate longitudinal response in the form of the full likelihood function. Section 3 introduces the proposed method for estimating the multivariate linear mixed model parameters in the presence of non-ignorable dropout. The estimation process involves estimating the dropout parameters and the response model parameters. An SEM algorithm combined with Newton-Raphson approach is proposed to estimate the dropout parameters. Also, a new proposed Quasi-Newton algorithm that is of EM type to estimate the response model parameters. The required equations for the estimation are derived. In section 4, simulation studies are conducted to evaluate the performance of the proposed technique. Finally, concluding remarks are provided in Section 5.

2. Models and Likelihood Function

In this section we introduce the used models; the multivariate linear mixed model and the dropout model. Also, the needed likelihood functions are defined.

Multivariate linear mixed model (MLMM)

Assume that the number of subjects are N , $i = 1, 2, \dots, N$ and there are n_i repeated measures on each subject. Also, assume that there are m outcomes on each subject, $k = 1, 2, \dots, m$. The responses for the subject i are collected in the response matrix $\mathbf{Y}_i = [y_{ijk}]$. The y_{ijk} denotes the measurement of the i^{th} subject of the k^{th} outcome at the j^{th} occasion. Assume that $\mathbf{Y}_i = (\mathbf{y}_{i1} \mathbf{y}_{i2} \dots \mathbf{y}_{im})$ is the $n_i \times m$ response matrix of the i^{th} subject where each of $\mathbf{y}_{ik} (k = 1, 2, \dots, m)$ is the k^{th} response n_i -vector measurements of subject i . It is very common that some subjects may withdraw from the study prematurely which results in dropout.

The multivariate longitudinal data can be modeled using the multivariate linear mixed model (MLMM), where for the i^{th} subject the response is modeled as:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N \quad (1)$$

where \mathbf{X}_i is the $n_i \times p$ fixed between-subject design matrix, $\boldsymbol{\beta}$ is the $p \times m$ matrix of fixed effects assumed to be common for all subjects, \mathbf{Z}_i is the $n_i \times q$ random within-subject design matrix, \mathbf{b}_i is the $q \times m$ matrix of random effects, and $\boldsymbol{\varepsilon}_i$ is the $n_i \times m$ matrix of the measurements errors associated with the response matrix \mathbf{Y}_i . The distributional assumptions that are often related to the MLMM are:

- The $\tilde{\boldsymbol{\varepsilon}}_i = \text{vec}(\boldsymbol{\varepsilon}_i) \sim N_{n_i m}(\mathbf{0}_{n_i m}, \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \otimes \mathbf{I}_{n_i})$; the rows of $\boldsymbol{\varepsilon}_i$ are normally distributed with mean $\mathbf{0}$ and $m \times m$ unstructured covariance matrix $\boldsymbol{\Sigma}$. The vec is the vectorized operator that stacks all columns of the

matrix ε_i vertically.

- The random effects $\tilde{\mathbf{b}}_i = \text{vec}(\mathbf{b}_i) \sim N_{qm}(\mathbf{0}_{qm}, \mathbf{\Psi})$, where $\mathbf{\Psi}$ is $qm \times qm$ unstructured matrix. Moreover, \mathbf{b}_i and ε_i are assumed to be independent.
- The marginal distribution of \mathbf{Y}_i has normal distribution with expectation $E(\mathbf{Y}_i) = \mathbf{X}_i\beta$ and dispersion matrix \mathbf{V}_i of dimension $n_i m \times n_i m$;

$$\mathbf{V}_i = (I_m \otimes \mathbf{Z}_i) \mathbf{\Psi} (I_m \otimes \mathbf{Z}_i)' + (\mathbf{\Sigma} \otimes \mathbf{I}_{n_i}). \quad (2)$$

The responses \mathbf{Y}_i can be expressed as $\tilde{\mathbf{y}}_i = \text{vec}(\mathbf{Y}_i)$ where $\tilde{\mathbf{y}}_i \sim N_{n_i m}(\mu_i = (I_m \otimes \mathbf{X}_i) \tilde{\beta}, \mathbf{V}_i)$, and can be modelled as:

$$\tilde{\mathbf{y}}_i = (I_m \otimes \mathbf{X}_i) \tilde{\beta} + (I_m \otimes \mathbf{Z}_i) \tilde{\mathbf{b}}_i + \tilde{\varepsilon}_i, \quad (3)$$

where $i = 1, 2, \dots, N$, and $\tilde{\beta} = \text{vec}(\beta)$. The p.d.f. of $\tilde{\mathbf{y}}_i$ is given by

$$\begin{aligned} f_{\tilde{\mathbf{y}}_i}(\tilde{\mathbf{y}}_i) &= (2\pi)^{-\frac{n_i m}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{\mathbf{y}}_i - \mu_i)' \mathbf{V}_i^{-1} (\tilde{\mathbf{y}}_i - \mu_i)} \\ &= (2\pi)^{-\frac{n_i m}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} e^{-\frac{1}{2} \text{tr}\{[\tilde{\mathbf{y}}_i - \mathbf{X}_i \beta] \mathbf{V}_i^{-1} [\tilde{\mathbf{y}}_i - \mathbf{X}_i \beta]'\}}. \end{aligned} \quad (4)$$

It can be verified that

$$\tilde{\mathbf{y}}_i | \tilde{\mathbf{b}}_i \sim N_{n_i m}((I_m \otimes \mathbf{X}_i) \tilde{\beta} + (I_m \otimes \mathbf{Z}_i) \tilde{\mathbf{b}}_i, \mathbf{\Sigma}_i). \quad (5)$$

The complete data likelihood function

Given that the m outcomes are fully observed, the complete-data likelihood function of the parameters, $\theta = (\beta, \mathbf{\Psi}, \mathbf{\Sigma})$, can be obtained as

$$\begin{aligned} L_1(\theta) &= \prod_{i=1}^N f(\mathbf{Y}_i | \mathbf{b}_i, \beta, \mathbf{\Sigma}) \times f(\mathbf{b}_i | \mathbf{\Psi}) \\ &= \prod_{i=1}^N f(\mathbf{Y}_i | \beta, \mathbf{\Sigma}) \times f(\mathbf{b}_i | \mathbf{\Psi}) \\ &= \prod_{i=1}^N f(\mathbf{Y}_i | \theta). \end{aligned} \quad (6)$$

The MLE of θ can be obtained by maximizing the likelihood function in Eq. (6). Note that this function is based on the marginal distribution function of the response \mathbf{Y}_i in the absence of the random effects. The random effects can be considered as nuisance parameters as long as the major interest is on the variance structure introduced by $(\mathbf{\Sigma}, \mathbf{\Psi})$. Hence the log-likelihood function, based on Eq. (4) for the whole sample is

$$\begin{aligned} \ell(\theta) &= \log\left[\prod_{i=1}^N f_{\tilde{\mathbf{y}}_i}(\tilde{\mathbf{y}}_i | \theta)\right] = \sum_{i=1}^N \log f_{\tilde{\mathbf{y}}_i}(\tilde{\mathbf{y}}_i | \theta) \\ &= -\frac{mM}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^N \log |\mathbf{V}_i| - \frac{1}{2} \sum_{i=1}^N \mathbf{r}_i' \mathbf{V}_i^{-1} \mathbf{r}_i, \end{aligned} \quad (7)$$

where $\theta = (\beta, \mathbf{\Psi}, \mathbf{\Sigma})$, is considered the parameter matrix of interest, $M = \sum_{i=1}^N n_i$, while $\mathbf{r}_i = \text{vec}(\mathbf{Y}_i - \mathbf{X}_i \beta) = \tilde{\mathbf{y}}_i - (I_m \otimes \mathbf{X}_i) \tilde{\beta}$.

A simpler form of the log-likelihood function is given by (Schafer and Yucel, 2002)

$$\begin{aligned} \ell(\theta) &= -\frac{mM}{2} \log(2\pi) + \frac{M}{2} \log |\mathbf{\Sigma}^{-1}| + \frac{N}{2} \log |\mathbf{\Psi}^{-1}| \\ &\quad - \frac{1}{2} \sum_{i=1}^N \log |\mathbf{U}_i^{-1}| - \frac{1}{2} \sum_{i=1}^N \mathbf{r}_i' \mathbf{W}_i \mathbf{r}_i, \end{aligned} \quad (8)$$

where $\mathbf{U}_i = (\mathbf{\Psi}^{-1} + (\mathbf{\Sigma}^{-1} \otimes \mathbf{Z}_i' \mathbf{Z}_i))^{-1}$, and $\mathbf{W}_i^{-1} = \mathbf{V}_i$.

We assume that the variance-covariance matrices, $\mathbf{\Psi}$ and $\mathbf{\Sigma}$, are unstructured. Moreover, we represent these matrices in terms of their precision form, $\mathbf{\Psi}^{-1}$ and $\mathbf{\Sigma}^{-1}$. These forms are obtained as linear functions as.

$$\begin{aligned} \mathbf{\Psi}^{-1}(\omega_{ij}) &= \sum_{i \geq j=1}^{mq} \omega_{ij} \xi_{ij} = \sum_{l=1}^h \omega_l^* \xi_l, \\ \mathbf{\Sigma}^{-1}(\sigma_{ij}) &= \sum_{i \geq j=1}^m \sigma_{ij} \xi_{ij} = \sum_{l=1}^g \sigma_l^* \xi_l, \end{aligned} \quad (9)$$

where $[\omega_l^*] = \text{vec}(\psi^{-1}) = (\omega_1^*, \omega_2^*, \dots, \omega_h^*)'$ and $h = \frac{mq(mq+1)}{2}$ is the number of diagonal and off-diagonal elements of ψ^{-1} . Thus $[\omega_l^*] = [\omega_{ij}]$, such that $\{(i, j); i \geq j, l = [(j-1)(mq-j/2) + i, i, j = 1, 2, \dots, mq]\}$. Similarly, $[\sigma_l^*] = \text{vec}(\Sigma^{-1}) = (\sigma_1^*, \sigma_2^*, \dots, \sigma_g^*)'$, $g = \frac{m(m+1)}{2}$, and $[\sigma_l^*] = [\sigma_{ij}]$, such that $\{(i, j); i \geq j, l = [(j-1)(m-j/2) + i, i, j = 1, 2, \dots, m]\}$ (Saber, 2008).

Further, each of ξ_{ij} and ξ_{ij} are square matrices of order mq and m , respectively, with one on the $(i, j)^{th}$ position and zero elsewhere. Hence, the response model parameter vector $\theta = (\text{vec}(\beta)' \sigma_1^* \sigma_2^* \dots \sigma_g^* \omega_1^* \omega_2^* \dots \omega_h^*)'$ and the total number of parameters is $mp + g + h$.

Missing data model and likelihood function

Assume that, for subject i , \mathbf{Y}_i is subject to non-random dropout at a certain occasion $d_i > 1$; $i = 1, 2, \dots, N$. The responses \mathbf{Y}_i can be partitioned into two components; the observed part \mathbf{Y}_i^o of dimension $(d_i - 1) \times m$ and missing part \mathbf{Y}_i^m of dimension $(n_i - d_i + 1) \times m$. Also, the response matrix using the vec operator can be partitioned into two parts; $\tilde{\mathbf{y}}_i = (\tilde{\mathbf{y}}_i^o \tilde{\mathbf{y}}_i^m)'$, where $\tilde{\mathbf{y}}_i^o = \text{vec}(\mathbf{Y}_i^o)$, and $\tilde{\mathbf{y}}_i^m = \text{vec}(\mathbf{Y}_i^m)$. Assume that the missing data indicator $\mathbf{R}_i = [r_{ijk}]$ of dimension $n_i \times m$ where r_{ijk} is defined as

$$r_{ijk} = \begin{cases} 1 & \text{if } y_{ijk} \text{ is observed,} \\ 0 & \text{if } y_{ijk} \text{ is missing.} \end{cases}$$

For monotone missing data we can define $\mathbf{R}_i = (\mathbf{1} \mathbf{0})'$, where $\mathbf{1}$ is the matrix of ones of dimension $(d_i - 1) \times m$ and $\mathbf{0}$ is the matrix of zeros of dimension $(n_i - d_i + 1) \times m$.

The Diggle and Kenward (1994) model has been proposed for univariate longitudinal data setting. We extend this model to the multivariate longitudinal data of m outcomes. Assume that the dropout of subject i occurs at time t_{d_i} , $\mathbf{H}_{id_i} = (\mathbf{h}_{i1} \mathbf{h}_{i2} \dots \mathbf{h}_{im})'$ denotes the $m \times [d_i - 1]$ matrix of observed responses history of the m outcomes up to the time t_{d_i} , where each of $\mathbf{h}_{ik} = (\mathbf{y}_{i1k} \mathbf{y}_{i2k} \dots \mathbf{y}_{i(d_i-1)k})'$ corresponds the row vector of length $d_i - 1$ of observed responses of k^{th} outcome. Diggle and Kenward (1994) modeled the probability of dropout in terms of the history $\mathbf{H}_{ij'}$ up to occasion j' and the possibly missing current response, $\mathbf{y}_{ij'}$ as $Pr(D_i = j' | \mathbf{H}_{ij'}, \mathbf{y}_{ij'}) = P_{ij'}(\mathbf{H}_{ij'}, \mathbf{y}_{id_i}, \phi)$, $j' = 2, \dots, n_i + 1$. This probability is computed for each subject across all occasions as follows:

$$Pr(D_i = j' | \mathbf{H}_{ij'}, \mathbf{y}_{ij'}) = \begin{cases} \prod_{j'=2}^{d_i-1} (1 - \mathbf{P}_{ij'}) \times \mathbf{P}_{id_i} & d_i \leq n_i \\ \prod_{j'=2}^{d_i-1} (1 - \mathbf{P}_{ij'}) \times \mathbf{P}_{id_i} & d_i = n_i + 1, \end{cases} \quad (10)$$

where the logit transformation that is used to model that probability, $P_{ij'}$ can be obtained by

$$\begin{aligned} \eta_{ij'} &= \text{logit}(\mathbf{P}_{ij'}) = \log\left(\frac{\mathbf{P}_{ij'}}{1 - \mathbf{P}_{ij'}}\right) \\ &= \phi_0 + \phi_1 \mathbf{y}_{ij'} + \phi_2 \mathbf{H}_{ij'}. \end{aligned} \quad (11)$$

Hence, the dropout probability is obtained via the inverse of the logit transformation as:

$$\mathbf{P}_{ij'} = \frac{\exp \eta_{ij'}}{\exp \eta_{ij'} + 1}. \quad (12)$$

The probability \mathbf{P}_{id_i} can be computed as

$$\mathbf{P}_{id_i} = \Pr(\mathbf{D}_i = d_i | \mathbf{y}_{id_i}) = \int P(\mathbf{H}_{d_i} | \mathbf{y}_{id_i}, \phi) f(\mathbf{y}_{id_i} | \mathbf{H}_{d_i}, \theta) d\mathbf{y}_{id_i}, \quad (13)$$

where $f(\mathbf{y}_{id_i} | \mathbf{H}_{d_i}, \theta)$ is the conditional probability of the missing given the observed responses. Hence the integral in Eq. (13) can be approximated using Monte Carol expectation (Gad and Ahmed, 2006).

The complete-data log-likelihood for the N subjects can be given by

$$\begin{aligned} \ell_c(\gamma) &= \log \mathbf{L}_c(\gamma) = \log \prod_{i=1}^N (f(\mathbf{Y}_i, \mathbf{R}_i | \gamma)) \\ &= \log \prod_{i=1}^N (f(\mathbf{Y}_i | \theta) \times f(\mathbf{R}_i | \mathbf{Y}_i, \phi)), \end{aligned} \quad (14)$$

where $\gamma = (\theta, \phi) = (\beta, \Psi, \Sigma, \phi)$. Also, it can be re-expressed as

$$\ell_c(\gamma) = \ell(\theta) + \ell(\phi) \quad (15)$$

The MLEs of the parameters γ can be obtained maximizing each of the two components in Eq. (15) separately. Direct maximization is not feasible because it requires multidimensional integration and maximization (Jason and Xihong, 2002). Numerical and iterative computations are exceedingly required.

3. The Proposed SEM Approach

Maximization of the log-likelihood function can be carried using iterative approaches to obtain the maximum-likelihood estimates (MLEs) such as Newton-Raphson (NR), Fisher Scoring algorithm (FS), Quasi-Newton (QN) and Expectation-Maximization (EM). Expectation-Maximization algorithm proposed by Dempster et al. (1977). It overcomes the difficulties involved with obtaining MLEs in the presence of missing values. The EM algorithm generally contains of two main steps; an Expectation step (E-step) and a Maximization step (M-step). For more details see McLachlan and Krishnan (2008). The E-step is often difficult to be evaluated. Celuex (1985) proposes the SEM algorithm as an alternative to the EM by replacing the E-step by a simulation step (S-step). Gad and Ahmed (2006) proposed SEM algorithm to fit the univariate longitudinal data in the presence of non-ignorable dropout.

In this article we propose an SEM algorithm to fit the multivariate linear mixed model for multivariate longitudinal data with non-ignorable dropout. The M-step is implemented via two sub-steps. In the first sub-step the Newton-Raphson algorithm is developed to obtain the MLEs of the dropout parameters ϕ . In the second sub-step a new type of Quasi-Newton algorithm is proposed to obtain the MLEs of the response parameter θ .

The S-Step (Simulation step)

Let $\mathbf{Y}_i^m = (\mathbf{y}_{id_i} | \mathbf{Y}_i^{m+})$ denotes the missing components

of the response matrix \mathbf{Y}_i over the m outcomes such that $\mathbf{y}_{id_i} = (\mathbf{y}_{id_i1}, \mathbf{y}_{id_i2}, \dots, \mathbf{y}_{id_im})'$, indicates the first $m_i \times 1$ missing vector, while the second matrix of order $m \times [n_i - d_i]$ is introduced by $\mathbf{Y}_i^{m+} = (\mathbf{y}_{i(d_i+1)}, \mathbf{y}_{i(d_i+2)}, \dots, \mathbf{y}_{i(n_i)})$. In this step (the S-step), we simulate m -vector from $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{R}_i, \theta^{(t)})$ to represent the responses of the m outcomes at the time of dropout d_i . This conditional distribution can be factorized, up to a constant of proportionality, as

$$f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{R}_i, \theta^{(t)}) \propto (\mathbf{y}_i^m | \mathbf{y}_i^o, \theta^{(t)}) \times f(\mathbf{R}_i | \mathbf{y}_i, \phi^{(t)}). \quad (16)$$

This conditional distribution has no closed form, so direct simulation is not possible. Following the idea of Gad and Ahmed (2006), we suggest using an accept-reject procedure. A vector of length m is simulated from $f(\mathbf{y}_i^m | \mathbf{y}_i^o; \theta^{(t)})$, then it is passed through the following accept-reject procedure;

1. Simulate a candidate m -vector $\mathbf{y}_{d_i}^* = (\mathbf{y}_{d_i1}^*, \mathbf{y}_{d_i2}^*, \dots, \mathbf{y}_{d_im}^*)'$, from the conditional distribution $f(\mathbf{y}_i^m | \mathbf{y}_i^o; \theta^{(t)})$ which is m -multivariate normal distribution with parameters $\mu^{(t)}$ and $\Omega^{(t)}$; the $m \times 1$ location vector and the $m \times m$ dispersion matrix respectively, are given by

$$\begin{aligned} \mu^{(t)} &= \mu_m^{(t)} + \mathbf{V}_{mo}^{(t)} (\mathbf{V}^{(t)})_{oo}^{-1} (\mathbf{y}_i^o - \mu_o^{(t)}), \\ \Omega^{(t)} &= \mathbf{V}_{mm}^{(t)} - \mathbf{V}_{mo}^{(t)} (\mathbf{V}^{(t)})_{oo}^{-1} \mathbf{V}_{om}^{(t)}, \end{aligned} \quad (17)$$

where $(\mathbf{V}^{(t)})_{oo}^{-1}$ is the variance matrix corresponds of the observed response matrix \mathbf{Y}_i^o ;

$$(\mathbf{V}^{(t)})_{oo} = (\mathbf{I}_m \otimes \mathbf{Z}_{io}) \Psi_o^{(t)} (\mathbf{I}_m \otimes \mathbf{Z}_{io}') + (\Sigma_o^{(t)} \otimes \mathbf{I}_{n_i}).$$

2. Compute the m -vector probability of dropout for the candidate vector $\mathbf{y}_{d_i}^*$ based on the dropout model

$$\begin{aligned} \text{logit}[\Pr(\mathbf{D}_i = d_i | \mathbf{y}_{d_i}^*, \mathbf{y}_{d_{i-1}}, \phi)] \\ = \phi_0^{(t)} + \phi_1^{(t)} \mathbf{y}_{d_i}^* + \phi_2^{(t)} \mathbf{y}_{d_{i-1}}, \end{aligned} \quad (18)$$

where $\mathbf{y}_{d_{i-1}}$ comprises the m -vector of responses at $(d_i - 1)$; the occasion precedes the dropout occasion. The dropout probability vector, $\mathbf{P}_i = (P_{i1}, P_{i2}, \dots, P_{im})'$, is evaluated using the inverse transformation in Eq. (12). Then define $P_i^* = \|\mathbf{P}_i\|$.

3. Simulate a random m -vector \mathbf{U} from uniform distribution over the interval $[0,1]$, then take $\mathbf{y}_i^m = \mathbf{y}_{d_i}^*$ if $\|\mathbf{U}\| \leq P_i^*$, otherwise repeat step 1.

We argue that the remaining missing values, following the first missing value, can be considered as missing at random.

The M-Step (Maximization step)

The joint probability function of the pseudo-observed data is

$$\begin{aligned} f(\mathbf{y}_i^o, \mathbf{y}_{id_i}, \mathbf{R}_i | \theta, \phi) = \\ \int f(\mathbf{y}_i^o, \mathbf{y}_{id_i}, \mathbf{Y}_i^{m+} | \theta) f(\mathbf{R}_i | \mathbf{y}_i^o, \mathbf{y}_{id_i}, \mathbf{Y}_i^{m+}; \phi) d\mathbf{Y}_i^{m+}, \end{aligned} \quad (19)$$

where $\mathbf{Y}_i^{m+} = \text{vec}(\mathbf{Y}_i^{m+})$. Since \mathbf{R}_i does not depend on \mathbf{Y}_i^{m+} by construction, the above equation can be simplified to

$$f(\mathbf{y}_i^o, \mathbf{y}_{id_i}, \mathbf{R}_i | \theta, \phi) = f(\mathbf{y}_i^o, \mathbf{y}_{id_i} | \theta) f(\mathbf{R}_i | \mathbf{y}_i^o, \mathbf{y}_{id_i}; \phi). \quad (20)$$

Thus the likelihood function for θ and ϕ will be any function of θ and ϕ proportional to $f(\mathbf{y}_i^0, \mathbf{y}_{id_i}, \mathbf{R}_i | \theta, \phi)$. Moreover, based on the selection model, θ and ϕ are assumed to be distinct by definition. Hence, $L(\theta, \phi | \mathbf{y}_i^0, \mathbf{y}_{id_i}, \mathbf{R}_i)$ is the same as likelihood $L(\theta | \mathbf{y}_i^0, \mathbf{y}_{id_i})$.

The M-step of the algorithm comprises two substeps; the logistic step (M1-Step), and the normal step (M2-Step):

The M1-Step (the logistic step):

In this substep the MLEs of the dropout parameter, $\phi^{(t)}$, is estimated numerically using Newton-Raphson approach. A step halving algorithm is adopted to guarantee that the log-likelihood function increasing toward the maxima over the iterations. Moreover, in case when the Hessian matrix of $\phi^{(t)}$ is not negative definite, the Newton direction $[-\mathbf{H}^{-1}(\phi_t)\mathbf{g}(\phi_t)]$ may not point in an ascent direction. Therefore, a modification is adopted using the Levenberg-Marquardt technique, to ensure that the search direction is an ascent direction, hence the algorithm guides the process to the maximum point (Chong and Zak, 2013).

The basic Newton-Raphson algorithm is used to obtain the MLEs for the dropout parameters of the multivariate logistic regression model. The algorithm seeks for optimizing, the log-likelihood function

$$\begin{aligned} \ell(\phi) = & \sum_{i=1}^{N-N^*} [\sum_{j=2}^{d_i-1} \ln(1 - \mathbf{P}_{ij}) \\ & + \ln \mathbf{P}_{id_i}] + \sum_{i=1}^{N^*} \sum_{j=2}^{n_i} \ln(1 - \mathbf{P}_{ij}), \end{aligned} \quad (21)$$

where N^* refers to the subjects who complete the study (the completers), hence, $N - N^*$ refers to dropout subjects. The probabilities vector \mathbf{P}_{ij} is computed using Eq. (11) for any occasion $j = 1, 2, \dots, d_i - 1$, and from Eq. (18) at the occasion of dropout $j = d_i$. Implementing the algorithm needs computing the gradient $\mathbf{g}(\phi)$ vector and the Hessian matrix $\mathbf{H}(\phi)$. First, the elements of the gradient vector $\mathbf{g}(\phi)$ are obtained as follow:

$$\begin{aligned} \frac{\partial \ell(\phi)}{\partial \phi_0} = & \sum_{i=1}^{N-N^*} (\sum_{j=2}^{d_i-1} -\mathbf{P}_{ij} + [1 - \mathbf{P}_{id_i}]) + \\ & \sum_{i=1}^{N^*} \sum_{j=2}^{n_i} -\mathbf{P}_{ij} \\ \frac{\partial \ell(\phi)}{\partial \phi_1} = & \sum_{i=1}^{N-N^*} (\sum_{j=2}^{d_i-1} -\mathbf{P}_{ij}\mathbf{y}_{ij} + [1 - \mathbf{P}_{id_i}]\mathbf{y}_{id_i}^*) + \\ & \sum_{i=1}^{N^*} \sum_{j=2}^{n_i} -\mathbf{P}_{ij}\mathbf{y}_{ij} \\ \frac{\partial \ell(\phi)}{\partial \phi_2} = & \\ & \sum_{i=1}^{N-N^*} (\sum_{j=2}^{d_i-1} -\mathbf{P}_{ij}\mathbf{y}_{i(j-1)} + [1 - \mathbf{P}_{id_i}]\mathbf{y}_{i(d_i-1)}) + \\ & \sum_{i=1}^{N^*} \sum_{j=2}^{n_i} -\mathbf{P}_{ij}\mathbf{y}_{i(j-1)}, \end{aligned}$$

where \mathbf{y}_{ij} is the m -vector of the responses at occasion j . The elements of the Hessian matrix are obtained as follow:

$$\begin{aligned} \frac{\partial^2 \ell(\phi)}{\partial \phi \partial \phi^T} = & \sum_{i=1}^{N-N^*} (\sum_{j=2}^{d_i-1} \frac{-\mathbf{P}_{ij}^2 (1-\mathbf{P}_{ij})}{2\mathbf{P}_{ij-1}} \dot{\mathbf{Y}}_{ij} \dot{\mathbf{Y}}_{ij}' + \\ & \frac{\mathbf{P}_{id_i} (1-\mathbf{P}_{id_i})^2}{2\mathbf{P}_{id_i-1}} \dot{\mathbf{Y}}_{id_i} \dot{\mathbf{Y}}_{id_i}') \\ & + \sum_{i=1}^{N-N^*} (\sum_{j=2}^{n_i} \mathbf{P}_{ij}^2 \dot{\mathbf{Y}}_{ij} \dot{\mathbf{Y}}_{ij}' - [1 - \mathbf{P}_{id_i}]^2 \dot{\mathbf{Y}}_{id_i} \dot{\mathbf{Y}}_{id_i}') \\ & + \sum_{i=1}^{N^*} \sum_{j=2}^{n_i} (\frac{-\mathbf{P}_{ij}^2 (1-\mathbf{P}_{ij})}{2\mathbf{P}_{ij-1}} \dot{\mathbf{Y}}_{ij} \dot{\mathbf{Y}}_{ij}' + \mathbf{P}_{ij}^2 \dot{\mathbf{Y}}_{id_i} \dot{\mathbf{Y}}_{id_i}'), \end{aligned}$$

where $\dot{\mathbf{Y}}_{ij} = [\mathbf{1} \ \mathbf{y}_{ij} \ \mathbf{y}_{i(j-1)}]'$ is the response matrix of order $(m \times 3)$.

It is worth noting that at an initial value ϕ_0 that might be chosen at which the Hessian matrix $\mathbf{H}(\phi)$ is not negative definite, the Levenberg-Marquardt modification of Newton's algorithm will be adopted. That approach approximates $\mathbf{H}(\phi)$ with another matrix \mathbf{H}_a , such that $\mathbf{H}_a = \mathbf{H} - \mu \mathbf{I}$, where the scalar μ is chosen large enough such that \mathbf{H}_a is negative definite (Chong and Zak, 2013).

The M2-Step (the normal step):

In this substep the MLEs of the parameters, $\theta^{(t)} = (\beta^{(t)}, \boldsymbol{\Psi}^{(t)}, \boldsymbol{\Sigma}^{(t)})$, for the multivariate normal is estimated numerically using a new proposed approach of Quazi-Newton type. The proposed Quazi-Newton algorithm begins with initial values of \mathbf{H}^* with $\mathbf{H}_0^* = \ddot{\mathbf{Q}}^{-1}(\theta_0, \theta_0)$, where $\ddot{\mathbf{Q}}(\theta_0, \theta_0)$ is computed as

$$\ddot{\mathbf{Q}}(\theta_0, \theta_0) = \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^{(t)}=\theta_0}, \quad (22)$$

where θ_0 is the initial values vector of the parameter θ and the function $\mathbf{Q}(\theta, \theta^{(t)})$ as in the EM algorithm. Generally, the algorithm proceeds as follows:

1. Initiate the algorithm with $\theta = \theta_0$, then compute the corresponding \mathbf{H}_0^* and \mathbf{g}_0 , where \mathbf{g}_0 is the gradient of the observed log-likelihood function.
2. Compute the step direction vector $\mathbf{d} = -\mathbf{H}^* \mathbf{g}$, then use the line search algorithm to specify α that maximizes $\ell(\theta + \alpha \mathbf{d})$. The Golden-Section algorithm is adopted here to specify the single parameter α that maximize $\ell(\theta + \alpha \mathbf{d})$ along the search direction \mathbf{d} from the point θ .
3. Set $\Delta \theta = \alpha \mathbf{d}$, and compute $\Delta \mathbf{g} = \mathbf{g}(\theta + \Delta \theta) - \mathbf{g}$.
4. Based on $\mathbf{H}^*, \Delta \theta$, and $\Delta \mathbf{g}$, $\Delta \mathbf{H}^*$ can be evaluated as $\Delta \mathbf{H}^* = (1 + \frac{\Delta \mathbf{g}' \mathbf{H}_t^* \Delta \mathbf{g}}{\Delta \mathbf{g}' \Delta \theta}) \frac{\Delta \theta \Delta \theta'}{\Delta \mathbf{g}' \Delta \theta} - \frac{\Delta \theta \Delta \mathbf{g}' \mathbf{H}_t^* + \mathbf{H}_t^* \Delta \mathbf{g} \Delta \theta'}{\Delta \mathbf{g}' \Delta \theta}$.
5. Replace θ by $\theta + \Delta \theta$, \mathbf{g} by $\mathbf{g} + \Delta \mathbf{g}$, and \mathbf{H}^* by $\mathbf{H}^* + \Delta \mathbf{H}^*$ and return to step (2) or stop based on a stopping criterion.

To implement the above algorithm, the quantities $\mathbf{Q}(\theta, \theta^{(t)})$, $\mathbf{g}(\theta)$, $\dot{\mathbf{Q}}(\theta_0, \theta^{(t)})$, and $\ddot{\mathbf{Q}}(\theta_0, \theta^{(t)})$ are needed. Deviations of such quantities are obtained as follows. First, the EM-quantity, $\mathbf{Q}(\theta, \theta^{(t)}) = E[\ell_c | \theta^{(t)}, \mathbf{Y}_i^0]$, can be easily deduced by taking the expectation of the log-likelihood of the complete-data given the observed responses \mathbf{Y}_i^0 and the current values of the parameter vector $\theta^{(t)}$. This function is given by

$$\begin{aligned} \mathbf{Q}(\theta, \theta^{(t)}) = & -\frac{mM}{2} \log(2\pi) + \frac{M}{2} \log |\boldsymbol{\Sigma}^{-1}| + \frac{N}{2} \log |\boldsymbol{\Psi}^{-1}| \\ & - \frac{1}{2} \sum_{i=1}^N \log |\mathbf{U}_i^{-1}| - \frac{1}{2} \sum_{i=1}^N \text{tr}(\mathbf{W}_i \mathbf{T}_i), \end{aligned} \quad (23)$$

where $\mathbf{T}_i = E[\text{vec}(\mathbf{Y}_i^c - \mathbf{X}_i \beta) \text{vec}(\mathbf{Y}_i^c - \mathbf{X}_i \beta)' | \theta^{(t)}, \mathbf{Y}_i^0]$ and \mathbf{Y}_i^c indicates the complete-data response matrix. To compute this expectation, set $\tilde{\mathbf{Y}}_i^c = \text{vec}(\mathbf{Y}_i^c - \mathbf{X}_i \beta)$, where $\tilde{\mathbf{Y}}_i^c \sim N_{n_i m}(\mathbf{0}_{n_i m}, \mathbf{V}_i = \mathbf{W}_i^{-1})$. Further the complete-data matrix, \mathbf{Y}_i^c , is partitioned into $[\mathbf{Y}_i^m, \mathbf{Y}_i^o]$, similarly, $\tilde{\mathbf{Y}}_i^c = [\tilde{\mathbf{Y}}_i^m, \tilde{\mathbf{Y}}_i^o]$. Based on axioms of conditional normal distribution, it can be easily shown that

$$E[\tilde{\mathbf{Y}}_i^c | \theta^{(t)}, \mathbf{Y}_i^o] = \mathbf{V}_{21} \mathbf{V}_{11}^{-1} \tilde{\mathbf{Y}}_i^o,$$

where \mathbf{V}_{11} is the square submatrix of \mathbf{V}_i corresponding to the observed elements, and \mathbf{V}_{21} is the rectangle sub-matrix of covariances between the missing and the observed elements. On the other hand, it also can be shown that

$$E[\tilde{\mathbf{Y}}_i^c (\tilde{\mathbf{Y}}_i^c)^T | \theta^{(t)}, \mathbf{Y}_i^o] = \begin{pmatrix} \tilde{\mathbf{Y}}_i^o (\tilde{\mathbf{Y}}_i^o)^T & \tilde{\mathbf{Y}}_i^o (\tilde{\mathbf{Y}}_i^o)^T \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \\ \mathbf{V}_{21} \mathbf{V}_{11}^{-1} \tilde{\mathbf{Y}}_i^o (\tilde{\mathbf{Y}}_i^o)^T & \mathbf{V}_{22} + \mathbf{V}_{21} \mathbf{V}_{11}^{-1} (\tilde{\mathbf{Y}}_i^o (\tilde{\mathbf{Y}}_i^o)^T - \mathbf{V}_{11}) \mathbf{V}_{11}^{-1} \mathbf{V}_{12} \end{pmatrix},$$

where \mathbf{V}_{22} is the square submatrix of covariances between the missing elements.

The gradient of the observed log-likelihood, $\mathbf{g}(\theta)$, is derived as

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \text{vec}(\beta)} &= -(\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \dot{\mathbf{X}}_i) \text{vec}(\beta - \tilde{\beta}), \\ \frac{\partial \ell(\theta)}{\partial \omega_{ks}} &= \frac{1}{2} \sum_{i=1}^N \text{tr}([\Psi - \mathbf{U}_i] \frac{\partial \psi^{-1}}{\partial \omega_{ks}} - \mathbf{r}_i' W_i \dot{\mathbf{Z}}_i \Psi \frac{\partial \psi^{-1}}{\partial \omega_{ks}} \Psi \dot{\mathbf{Z}}_i' W_i \mathbf{r}_i), \\ \frac{\partial \ell(\theta)}{\partial \sigma_{lt}} &= \frac{1}{2} \sum_{i=1}^N \text{tr}(n_i \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} - \mathbf{U}_i [\frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \otimes \mathbf{Z}_i' \mathbf{Z}_i] - \mathbf{r}_i' W_i [\Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \otimes \mathbf{I}_{n_i}] W_i \mathbf{r}_i), \end{aligned}$$

where, $\text{vec}(\tilde{\beta}) = (\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \dot{\mathbf{X}}_i)^{-1} (\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \text{vec}(\mathbf{Y}_i))$, $\dot{\mathbf{X}}_i = (\mathbf{I}_m \otimes \mathbf{X}_i)$, and $\dot{\mathbf{Z}}_i = (\mathbf{I}_m \otimes \mathbf{Z}_i)$. Further, $\frac{\partial \psi^{-1}}{\partial \omega_{ks}}$ and $\frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}}$ are replaced with $(2 - \delta_{ks})\xi_{ks}$ and $(2 - \delta_{lt})\xi_{lt}$, respectively, such that, the scalar δ_{ks} which is intensively known as *Kronecker delta*, equals 1 when $k = s$ and 0 otherwise, where $k \geq s = 1, 2, \dots, mq$, while $l \geq t = 1, 2, \dots, m$.

The second derivatives of the EM-Q function are needed. The computation of the first and the second derivative of the EM-Q function are

$$\begin{aligned} \frac{\partial \mathbf{Q}(\theta, \theta^{(t)})}{\partial \text{vec}(\beta)} &= -(\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \dot{\mathbf{X}}_i) \text{vec}(\beta - \tilde{\beta}), \\ \frac{\partial \mathbf{Q}(\theta, \theta^{(t)})}{\partial \omega_{ks}} &= \frac{1}{2} \sum_{i=1}^N \text{tr}([\Psi - \mathbf{U}_i] \frac{\partial \psi^{-1}}{\partial \omega_{ks}} - W_i \dot{\mathbf{Z}}_i \Psi \frac{\partial \psi^{-1}}{\partial \omega_{ks}} \Psi \dot{\mathbf{Z}}_i' W_i \mathbf{T}_i), \\ \frac{\partial \mathbf{Q}(\theta, \theta^{(t)})}{\partial \sigma_{lt}} &= \frac{1}{2} \sum_{i=1}^N \text{tr}(n_i \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} - \mathbf{U}_i [\frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \otimes \mathbf{Z}_i' \mathbf{Z}_i] - W_i [\Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \otimes \mathbf{I}_{n_i}] W_i \mathbf{T}_i), \end{aligned}$$

where $\text{vec}(\tilde{\beta}) = (\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \dot{\mathbf{X}}_i)^{-1} (\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \mathbf{T}_i)$ and $\mathbf{T}_i = E[\text{vec}(\mathbf{Y}_i^c) | \theta, \mathbf{Y}_i^o]$.

$$\begin{aligned} \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \text{vec}(\beta) \partial (\text{vec}(\beta))'} &= -\sum_{i=1}^N \dot{\mathbf{X}}_i' W_i \dot{\mathbf{X}}_i, \\ \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \omega_{ks} \partial (\text{vec}(\beta))'} &= \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \sigma_{lt} \partial (\text{vec}(\beta))'} = 0, \\ \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \omega_{krs} \partial \omega_{ks}} &= -\frac{1}{2} \sum_{i=1}^N \text{tr}([\psi \frac{\partial \psi^{-1}}{\partial \omega_{krs}} \psi \frac{\partial \psi^{-1}}{\partial \omega_{ks}}] - [\mathbf{U}_i \frac{\partial \psi^{-1}}{\partial \omega_{krs}} \mathbf{U}_i \frac{\partial \psi^{-1}}{\partial \omega_{ks}}] \\ &\quad + 2W_i \dot{\mathbf{Z}}_i \Psi \frac{\partial \psi^{-1}}{\partial \omega_{krs}} \Psi \dot{\mathbf{Z}}_i' W_i \dot{\mathbf{Z}}_i \Psi \frac{\partial \psi^{-1}}{\partial \omega_{ks}} \Psi \dot{\mathbf{Z}}_i' W_i \mathbf{T}_i \end{aligned}$$

$$\begin{aligned} &-2W_i \dot{\mathbf{Z}}_i \Psi \frac{\partial \psi^{-1}}{\partial \omega_{krs}} \Psi \frac{\partial \psi^{-1}}{\partial \omega_{ks}} \Psi \dot{\mathbf{Z}}_i' W_i \mathbf{T}_i) \\ \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \omega_{ks} \partial \sigma_{lt}} &= -\frac{1}{2} \sum_{i=1}^N \text{tr}([-\mathbf{U}_i \frac{\partial \psi^{-1}}{\partial \omega_{ks}} \mathbf{U}_i (\frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \otimes \mathbf{Z}_i' \mathbf{Z}_i)] \\ &\quad + 2W_i \dot{\mathbf{Z}}_i \Psi \frac{\partial \psi^{-1}}{\partial \omega_{ks}} \Psi \dot{\mathbf{Z}}_i' W_i (\Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \Sigma \otimes \mathbf{I}_{n_i}) W_i \mathbf{T}_i) \\ \frac{\partial^2 \mathbf{Q}(\theta, \theta^{(t)})}{\partial \sigma_{lts} \partial \sigma_{lt}} &= -\frac{1}{2} \sum_{i=1}^N \text{tr}(n_i \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lts}} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \\ &\quad - \mathbf{U}_i (\frac{\partial \Sigma^{-1}}{\partial \sigma_{lts}} \otimes \mathbf{Z}_i' \mathbf{Z}_i) \mathbf{U}_i (\frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \otimes \mathbf{Z}_i' \mathbf{Z}_i) \\ &\quad + 2W_i (\Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lts}} \Sigma \otimes \mathbf{I}_{n_i}) W_i (\Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \Sigma \otimes \mathbf{I}_{n_i}) W_i \mathbf{T}_i \\ &\quad - 2W_i (\Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lts}} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}} \Sigma \otimes \mathbf{I}_{n_i}) W_i \mathbf{T}_i), \end{aligned}$$

For a parameter θ , the SEM estimates is taken as the mean of the sequence $\{\theta^{(t)}\}$ as

$$\hat{\theta} = \frac{1}{T - T_0} \sum_{j=T_0+1}^T \theta^{(j)}, \quad (24)$$

where T is the total number of SEM iterations while T_0 is the length of burn-in period.

In the above derivations of the gradient vector and the derivatives of the EM-Q function, the quantities; $\frac{\partial \psi^{-1}}{\partial \omega_{ks}}$ and $\frac{\partial \Sigma^{-1}}{\partial \sigma_{lt}}$ are replaced with $(1 - 0.5\delta_{ks})[e_k e_s' + e_s e_k']$ and $(1 - 0.5\delta_{lt})[e_l e_t' + e_t e_l']$, respectively. The scalar δ_{ks} is *Kronecker delta* which equals 1 when $k = s$ and 0 otherwise, where $k \geq s = 1, 2, \dots, mq$. Similarly the scalar δ_{lt} is defined for $l \geq t = 1, 2, \dots, m$. The vector e_k is the basis vector of length mq with k^{th} element equals 1 and all other elements are zeros.

4. Simulation Study

The aim of this simulation is to evaluate the performance of the proposed approach. The evaluation criteria is the relative bias.

Simulation Setting and data generation

Two sample sizes were assumed; $N_1 = 25$, and $N_2 = 50$ to generate data. The number of outcomes for each subject is assumed to be three outcomes; $m = 3$. The number of time points is fixed at 10; $n_i = 10$. The number of replications is chosen as 200. It is assumed that subjects are allocated to one of three treatments A, B, and C. Hence, there are a matrix of dimension 3×3 of fixed-effects β of the three covariates across the three outcomes. The subject-specific design matrix is assumed to have two columns ($q = 2$); the random intercept and slope. Such that, the two subject-specific covariates $(z_{1i}, z_{2i}) = (1 \ 0)$, for odd subject, and $(z_{1i}, z_{2i}) = (1 \ 1)$, for even subjects. Also the random design matrix \mathbf{Z}_i is assumed to be time-invariant. Consequently, the individual level matrix of random-effect parameters is \mathbf{b}_i is of order (2×3) . It represents the parameters of random-effects across outcomes. This matrix is generated for each subject as $\text{vec}(\mathbf{b}_i) = (b_{01i} \ b_{02i} \ b_{03i} \ b_{11i} \ b_{12i} \ b_{13i})' \sim MVN_6(\mathbf{0}, \Psi)$,

where Ψ is of dimension 6×6 assumed to be unstructured matrix. The measurement errors associated with the three outcomes are assumed to be independent over the repeated measure having a common unstructured variance-covariance matrix Σ of dimension 3×3 . Hence, for each subject the matrix ε_i is generated from a matrix normal distribution of parameters $MVN_{(10 \times 3)}(\mathbf{0}, \Sigma_i = \Sigma \otimes \mathbf{I}_{10})$. The true value of the elements of the parameter matrices; β , Σ , and Ψ are assumed as

$$\beta = \begin{bmatrix} 1 & 2 & 3 \\ -3 & -2 & -1 \\ 2 & 3 & 4 \end{bmatrix}, \Sigma = \begin{bmatrix} 3.5 & 0.2 & -2.2 \\ 0.2 & 0.5 & 1.1 \\ -2.2 & 1.1 & 4.7 \end{bmatrix}, \text{ and}$$

$$\Psi = \begin{bmatrix} 7.2 & -0.9 & 4.8 & 0.2 & 1 & 2.2 \\ -0.9 & 5.3 & -0.1 & 0.1 & 0.4 & 1.6 \\ 4.8 & -0.1 & 6.3 & 2.4 & 0.1 & 1.7 \\ 0.2 & 0.1 & 2.4 & 5.1 & -2.5 & 0.4 \\ 1.0 & 0.4 & 0.1 & -2.5 & 3.6 & 0.9 \\ 2.2 & 1.6 & 1.7 & 0.4 & 0.9 & 3.4 \end{bmatrix}.$$

The response matrix \mathbf{Y}_i , for each subject, is generated from the model

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \varepsilon_i, i = 1, 2, \dots, N_s, s = 1, 2. \quad (25)$$

The missing data model is

$$\text{logit}(\mathbf{P}_{ij}) = \phi_0 + \phi_1 \mathbf{y}_{ij} + \phi_2 \mathbf{y}_{i(j-1)}, \quad (26)$$

where each of \mathbf{P}_{ij} , \mathbf{y}_{ij} , and $\mathbf{y}_{i(j-1)}$ is a 3-column vector, indicates the probabilities of dropout at the occasion j associated to the vector of responses, the vector of responses that were planned to be measured at occasion j , and the vector of responses at the previous occasion $(j-1)$, respectively. However, if the dropout occasion, for a subject i , is specified as d_i , then the dropout probability is computed as a probability of the dropout time D_i at all possible occasion as follow

$$Pr(D_i) = ([\prod_{j=2}^{d_i-1} (1 - \mathbf{P}_{ij})] \times \mathbf{P}_{id_i}) I(D_i)_{[d_i \leq 10]} \\ + ([\prod_{j=2}^{10} (1 - \mathbf{P}_{ij})]) I(D_i)_{[d_i = 11]}.$$

For each subject, the response matrix generated in the previous step is exposed to a non-random dropping process on the occasion level. Precisely, at occasion j , after the baseline ($j \geq 2$), the response vector \mathbf{y}_{ij} has been checked to be whether retained or dropped as follow:

1. The initial values for the parameter vector of logistic dropout model is assumed to be $\phi = [-0.4 \ 2.4 \ -2.7]$.
2. The probability vector P_{ij} is computed based on Eq. (26).
3. A random variable U is generated from uniform (0,1).
4. The response vector \mathbf{y}_{ij} is dropped if $\|\mathbf{P}_i\| \geq U$, otherwise it is retained.

Simulation Results

The proposed approach introduced has been applied to data. For each of the two maximization steps the stopping criterion that is suggested in Demidenko (2004) is used. Precisely, the iteration will be terminated if the norm value of the absolute difference of two successive values of the gradient vector at fixed parameter values is less than 0.0001.

To accelerate the algorithm the study adopted the Levenberg-Marquardt technique to modify each of the Hessian matrix of the dropout parameters and the second derivative matrix of the Q function of the model precision parameters, to be negative definite. The SEM algorithm is iterated for 500 iterations with a burn-in period of 250 iterations. The MLEs for the parameters were taken as the averages of the last 250 iterations. The performance of the proposed approach is evaluated via the relative bias (RB) of its obtained MLEs, which was computed as follow:

$$RB(\hat{\theta}) = \frac{|\hat{\theta} - \theta|}{|\theta|} \times 100.$$

The parameter estimates are displayed in Table (1) for $N=25$ and Table (2) for $N=50$.

Table (1). Parameter estimates for $N = 25$

Parameter	True Values	Estimates	Relative Bias (%)
β_{01}	1.0	1.3	30.7
β_{11}	-3.0	-2.4	21.0
β_{21}	2.0	2.1	3.2
β_{02}	2.0	2.3	13.4
β_{12}	-2.0	-1.1	44.1
β_{22}	3.0	3.0	1.0
β_{03}	3.0	3.6	21.0
β_{13}	-1.0	-0.6	42.1
β_{23}	4.0	3.9	1.9
ϕ_0	-0.4	-0.5	17.3
ϕ_1	2.5	2.5	0.9
ϕ_2	-2.7	-2.7	0.9

Table (2). Parameter estimates for $N = 50$

Parameter	True Values	Estimates	Relative Bias (%)
β_{01}	1.0	0.6	42.6
β_{11}	-3.0	-3.1	2.6
β_{21}	2.0	2.5	27.0
β_{02}	2.0	1.8	9.7
β_{12}	-2.0	-1.9	7.3
β_{22}	3.0	3.5	16.7
β_{03}	3.0	2.9	1.8
β_{13}	-1.0	-1.3	27.2
β_{23}	4.0	4.0	0.8
ϕ_0	-0.4	-0.4	1.9
ϕ_1	2.5	2.6	2.0
ϕ_2	-2.7	-2.8	2.0

The results show similar performance for the dropout parameter estimates based on the two settings of sample sizes. However, the estimates of the intercept parameter (β_{01}) of the first outcome show maximum bias in both tables (30.7%, and 42.6%). For sample size $N=25$, the estimate of the effect of the first treatment on the second and third outcomes show somewhat large bias (44.1%, and 42.1% respectively). Generally, the second table show less bias for most parameters (except for the estimate of the first

intercept). It almost does not exceed 27.0% approximately. Thus, the proposed approach show higher performance with large sample sizes.

5. Conclusions and Future Research

The study introduces a new estimating approach for modeling multiple outcomes longitudinal data under the multivariate linear mixed model (MLMM) in the presence of non-ignorable missingness. The missing pattern was assumed to be monotone, such that the m outcomes have the same pattern of dropout, while the covariates were assumed to be fully observed. To obtain valid estimates of the model parameters, the study incorporates the model of missing data into the estimation process via the logit selection model. The SEM algorithm is used as estimation approach. The S-Step of the algorithm was implemented via accept-reject principle for simulating the m -vector of outcomes at the dropout occasion from the specified distribution. The algorithm suggested two maximization steps; M1-Step to maximize the likelihood function of the dropout parameters of the multivariate logistic model, and M2-Step to maximize the likelihood function of the response model parameters. The study proposed a novel algorithm to implement the second maximization which is a new form of Quasi-Newton algorithm that incorporates EM-related quantity which is the Q function. The resulted proposed approach is a QN of EM-type.

The proposed approach is efficient in at least major two respects. It suggested using a new version of Quasi-Newton algorithm that embedded with some EM quantities. It takes the advantage of the EM algorithm in handling incomplete data by augmenting the missing and the observed observation to produce better estimates. However, it avoids complicated computation of the EM algorithm due to its iterating on the E-step which is infeasible with models with multi-parameters. So, it suggested computing the matrix of the second derivatives of the Q function with respect to the current guess of the precision parameters only once to initiate the basic Quasi-Newton algorithm.

Some possible extensions to the current study are recommended for further investigation. The current study could be extended assuming the intermittent pattern. Consequently, different likelihood function for the missing data mechanism parameters should be derived to fit such pattern. Moreover, it is very practically to violate the assumption that all outcomes have the same pattern of missingness. Violating such assumption will cause a considerable modification in the computation of the Q function. One further extension is to admit possible missing covariates. This will motivate incorporating missing data mechanism of the covariates into the estimation process.

It is very common to assume that the measurement errors within subjects are autocorrelated. It is possible to extend the variance structure of the errors from $\Sigma_i = \Sigma \otimes \mathbf{I}_{n_i}$ to be $\Sigma_i = \Sigma \otimes \mathbf{R}_i$, where the matrix \mathbf{R}_i can take a simple

autoregressive pattern. However, this will imply estimating more parameters.

The performance of the proposed approach is evaluated by a simulation study assuming three outcomes. The study is needed to be extended to check its applicability for a higher number of outcomes.

REFERENCES

- [1] Celeux, G. (1985) The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Computational statistics quarterly*, 2, 73–82.
- [2] Chong, E. K. & Zak, S. H. (2013) *An introduction to optimization* (Fourth Edition), Wiley.
- [3] Demidenko, E. (2004) *Mixed models: Theory and applications*, Wiley.
- [4] Diggle, P. J., Heagerty, P. J., Liang, K.-Y., Zeger, S. (2002) *Analysis of longitudinal data*, Oxford University Press.
- [5] Dempster, A. P., Liard, N. M., and Robin, D. B. (1977) Maximum Likelihood from Incomplete Data Via EM Algorithm (With Discussion), *Journal of Royal Statistical Society B*, 39, 1–38.
- [6] Diggle, P. & Kenward, M. G. (1994) Informative drop-out in longitudinal data analysis, *Applied Statistics*, 43, 49–93.
- [7] Enders, C. K. (2010) *Applied missing data analysis* (Second Edition), The Guilford Press, New York.
- [8] Gad, M. A. 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.
- [9] Graham, J. W. (2012) *Missing data: Analysis and design*, Springer. New York.
- [10] Jamshidian, M. & Jennrich, R. I. (1993) Conjugate gradient acceleration of the EM algorithm, *Journal of the American Statistical Association*, 88, 221–228.
- [11] Kim, K. & Tim, N. (2007) *Univariate and multivariate general linear models: Theory and applications with SAS* (2nd Edition), Chapman & Hall, New York.
- [12] Laird, N. M. & Ware, J. H. (1982) Random-effects models for longitudinal data, *Biometrics*, 38, 963–974.
- [13] Little, R. J. A. & Rubin, D. B. (2002) *Statistical analysis with missing data* (2nd Edition), Wiley.
- [14] Luwanda, A. G. & Mwambi, H. G. (2016) A nonlinear mixed-effects model for multivariate longitudinal data with dropout with application to HIV disease dynamics, *Journal of Agricultural, Biological, and Environmental Statistics*, 21(2), 277–294.
- [15] McLachlan, G. J. & Krishnan, T. (2008) *The EM algorithm and extensions* (2nd Edition), Wiley.
- [16] Reinsel, G. (1982) Multivariate repeated-measurement or growth curve models with multivariate random-effects

- covariance structure, *Journal of the American Statistical Association*, 77, 190–195.
- [17] Reinsel, G. (1984) Estimation and prediction in a multivariate random effects generalized linear model, *Journal of the American Statistical Association*, 79, 406–414.
- [18] Roy, J. & Lin, X. (2002) Analysis of multivariate longitudinal outcomes with non-ignorable dropouts and missing covariates: Changes in methadone treatment practices, *Journal of the American Statistical Association*, 97, 40–52.
- [19] Rubin, D. B. (1976) Inference and missing data, *Biometrika*, 63, 581–592.
- [20] Schafer, J. L. & Yucel, R. M. (2002) Computational strategies for multivariate linear mixed-effects models with missing values, *Journal of Computational and Graphical Statistics*, 11, 437–457.
- [21] Seber, G. A. F. (2008) *A matrix handbook for statisticians*, Wiley.
- [22] Shah, A., Laird, N. & Schoenfeld, D. (1997) A random-effects model for multiple characteristics with possibly missing data, *Journal of the American Statistical Association*, 92, 775–779.
- [23] Stubbendick, A. L. & Ibrahim, J. G. (2003) Maximum likelihood methods for non-ignorable missing responses and covariates in random effects models, *Biometrics*, 59, 1140–1150.
- [24] Yucel, R. M. (2015). R mlmmm package: Fitting multivariate linear mixed effects models with missing values. *Turkiye Klinikleri Journal of Biostatistics*, 7(1), 11-24.